

Redes neuronales profundas en búsquedas relacionadas con la producción de pares de quarks top en el CMS

Deep neural networks on searches related to top pair production at CMS

ANTONIO MÁRQUEZ GARCÍA

Grado en Física



Universidad de Oviedo

Supervisado por
FRANCISCO JAVIER CUEVAS MAESTRO,
SERGIO SÁNCHEZ CRUZ Y
PIETRO VISCHIA

Resumen

Este trabajo tiene como objetivo final el análisis del rendimiento y la construcción de dos algoritmos que clasifican leptones en procesos $t\bar{t}$ como *prompt* o *fake*.

El enfoque que se toma en el documento partirá de la descripción del contexto físico de los algoritmos (el Modelo Estándar de Física de partículas), así como del dispositivo experimental con el que se toman los datos (el detector *CMS* del *LHC*). Se revisará también la importancia en la clasificación de este tipo de leptones, en particular asociada al análisis de la producción del bosón de Higgs.

Se atenderán por otra parte a las características de los algoritmos, que emplea técnicas avanzadas de *Machine Learning* (en particular una *red neuronal*), y se describirá en detalle la optimización de estos comparándolos con un algoritmo más simple de *Machine Learning* ya empleado en el análisis (un árbol de decisión dopado o *boosted*).

Por último se incluye una discusión tras la construcción del clasificador en términos de su rendimiento, de sus posibles mejoras y de la implementación de redes neuronales en problemas más complejos.

Índice

2	Resumen	I
3	1. Introducción	1
4	2. El Modelo Estándar	3
5	2.1. Fermiones	3
6	2.2. Interacciones entre partículas según el modelo. Bosones	5
7	2.3. Límites del Modelo Estándar	8
8	3. El Gran Colisionador de Hadrones y su detector CMS	10
9	3.1. El Gran Colisionador de Hadrones	10
10	3.1.1. Descripción general del LHC	10
11	3.1.2. Diferencias con respecto al LEP	11
12	3.1.3. Inyección en el colisionador	12
13	3.1.4. Imanes del LHC	13
14	3.2. El detector CMS y sus componentes	14
15	3.2.1. Sistema de coordenadas	14
16	3.2.2. Solenoide magnético	15
17	3.2.3. Detector de trazas	15
18	3.2.4. Calorímetros	17
19	3.2.5. Cámaras de muones	18
20	3.2.6. Sistema de filtrado	19
21	4. La reconstrucción de sucesos	21
22	4.1. El rol de la simulación	21
23	4.2. Algoritmo de reconstrucción e identificación. Trazas y vértice primario	22
24	4.3. Técnica de <i>Lepton MVA</i>	22
25	4.4. Reconstrucción e identificación de leptones	23
26	4.5. Otras componentes de la reconstrucción	23
27	5. La detección del bosón de Higgs	25
28	5.1. El proceso ttH	25
29	5.2. Desintegraciones del bosón H	26
30	5.3. Desintegraciones del par $t\bar{t}$	28
31	5.4. Fondo en el análisis	29

32	5.5. Sección eficaz de los procesos.	30
33	5.6. Descripción de los datos empleados	34
34	6. El nuevo algoritmo de identificación	47
35	6.1. Introducción al <i>Machine Learning</i>	47
36	6.1.1. Arbol de decisión	48
37	6.1.2. Dopaje de algoritmos. Árbol de decisión dopado (<i>BDT</i>)	51
38	6.1.3. Red neuronal profunda (<i>DNN</i>)	53
39	6.1.4. Red neuronal propuesta. Curva ROC	56
40	6.2. Entrenamiento de la red	60
41	6.2.1. Consideraciones en el proceso experimental	60
42	6.2.2. Optimización de hiperparámetros	60
43	7. Resultados obtenidos y conclusiones	67
44	7.1. Resultados obtenidos	67
45	7.2. Conclusiones y perspectivas	68
46	Bibliografía.	72

47 Capítulo 1

48 Introducción

49 *Nobody ever figures out what life is all about, and it doesn't matter. Explore the world. Nearly everything is really interesting if you go into it deeply enough.*

Richard P. Feynman

50 La estructura fundamental de la materia ha sido el objeto de estudio de miles de científicos
51 hasta el momento. Ya en los años 200-100 a.C. en la Antigua Grecia se hablaba de *atomismo*,
52 una teoría filosófica que proponía un universo constituido por partículas indivisibles a las que
53 denominaba *átomos* (que se traduce como *indivisibles*).

54 La mejora en las teorías y técnicas de estudio que han habido desde entonces han llevado a
55 explicar los bloques de construcción más básicos del universo como *partículas fundamentales*,
56 las cuales están gobernadas por cuatro *fuerzas fundamentales*. Nuestro mejor entendimiento
57 de estas partículas junto con tres de estas fuerzas fundamentales se encuentra en una teoría
58 científica desarrollado en el siglo XX: el *Modelo Estándar*.

59 Este modelo ha tenido éxito explicando la gran mayoría de resultados experimentales en esta
60 área. De la misma manera, su gran capacidad de predicción se puede ejemplificar con el
61 descubrimiento del bosón de Higgs en 2012 en el *Gran Colisionador de Hadrones del CERN*.

62 No obstante, el Modelo Estándar no es una “teoría del todo” (no existe todavía dicha teoría).
63 Aún hay fenómenos en la física de partículas que no se ven explicados por el modelo; uno de
64 ellos es la fuerza gravitatoria, que no ha podido ser incluida en el modelo como una interacción
65 análoga a las demás. Actualmente se investigan maneras de completar este modelo, mediante
66 teorías conocidas como *Más Allá del Modelo Estándar*, que sean capaces de explicar todos
67 estos fenómenos.

68 El colisionador de partículas más grande del mundo, el Gran Colisionador de Hadrones, (*Large*
69 *Hadron Collider, LHC*), es capaz de lograr colisiones entre protones a energías del orden de 13
70 TeV. Sobre este acelerador se llevan a cabo distintos experimentos con el objetivo de probar
71 y refinar el Modelo Estándar, así como poner a prueba sus límites.

72 La información obtenida en estos experimentos es muy compleja, y de dimensiones muy
73 altas. Más aún: se está desarrollando una mejora del LHC: el *High Luminosity LHC*, HL-
74 LHC, que aumentará su luminosidad hasta 20 veces más, consiguiendo un mayor volumen y
75 complejidad en los datos. Nuevas técnicas de análisis multivariable de la información están
76 siendo investigadas con el objetivo de un manejo de los datos consistente.

77 En este trabajo se introducirá un algoritmo de análisis multivariable capaz de identificar el
78 origen de leptones en la producción de pares de quarks $t\bar{t}$, diferenciando los producidos por
79 la desintegración leptónica de un bosón W de aquellos con una distinta procedencia. Este
80 algoritmo se describe como una *red neuronal*, y es un tipo de herramienta prometedora en el
81 manejo de información de mayor complejidad, con un rendimiento comparable a las técnicas
82 actuales.

83 Para la presentación del algoritmo, primero se describirá con una mayor precisión las caracte-
84 rísticas y los límites del Modelo Estándar en el capítulo 2. Más tarde, en el capítulo 3 se
85 describirán las especificaciones del LHC y, en particular, del detector CMS, que permiten la
86 obtención de los datos con los que se trabaja.

87 La relevancia en el estudio de producción de pares $t\bar{t}$ nace de su relación con la producción
88 del importante bosón de Higgs. Esta relevancia será descrita en el capítulo 5, así como la
89 producción del aclamado bosón.

90 El contexto en el que trabaja el algoritmo viene descrito en el capítulo 4, donde además se
91 estudia el análisis que realiza: el *Leptón MVA*.

92 Es entonces cuando se describirá con precisión la técnica empleada por la red neuronal, así
93 como su proceso de construcción y entrenamiento, en el capítulo 6. Por último, el capítulo
94 7 expresará las conclusiones obtenidas tras la evaluación de la red sobre la muestra, entre
95 las cuales se analizará y comparará su rendimiento con técnicas anteriores empleadas en el
96 mismo ámbito.

97 Capítulo 2

98 El Modelo Estándar

99 El Modelo Estándar, o *Standard Model (SM)* en inglés, es una teoría cuántica y relativista
100 de campos desarrollada en la segunda mitad del siglo XX. Esta teoría describe todas las
101 partículas que se han podido observar hasta el momento: desde los fermiones, que son los
102 constituyentes de la materia que se considera, con base las observaciones experimentales
103 realizadas hasta la fecha, que son más fundamentales; hasta los bosones, que describen sus
104 interacciones.

105 2.1. Fermiones

106 En el Modelo Estándar, las partículas que constituyen la materia se caracterizan por su espín
107 semientero y se denominan *fermiones*. Estas verifican el principio de exclusión de Pauli¹ y
108 obedecen la estadística de Fermi-Dirac².

109 Con el objetivo de la descripción de cada partícula, se le asocia a cada una una serie de
110 números cuánticos, que se identifican con un tipo de carga y permiten estudiar la conserva-
111 ción de magnitudes físicas asociadas estos. Es el caso de la carga eléctrica en la interacción
112 electromagnética, o su análoga carga de color en la interacción fuerte.

113 Los fermiones son los constituyentes de la materia que no poseen carga de color son los
114 denominados *leptones*. Hay tres leptones diferentes que poseen carga eléctrica: el *electrón*,
115 el *muon* (μ) y el *tauón* (τ). Es frecuente denotar el valor de la carga eléctrica del electrón
116 como $-e$. Cada una de estas partículas tienen un neutrino asociado de carga eléctrica y
117 masa nula, formando así los tres leptones restantes. Los neutrinos son los únicos fermiones
118 que no reaccionan mediante interacciones electromagnéticas. En la figura 2.1 se representan
119 las partículas descritas organizadas en tres *generaciones* que separan a cada par de leptón
120 cargado y neutrino asociado

¹El principio de exclusión de Pauli estipula que dos o más partículas de espín semi-entero no pueden ocupar el mismo estado energético del mismo sistema simultáneamente.

²La estadística de Fermi-Dirac describe como un conjunto de partículas indistinguibles de espín semi-entero pertenecientes al mismo sistema se distribuyen en una colección de estados energéticos en equilibrio termodinámico.

121 De la misma manera, hay seis fermiones que poseen carga de color: los *quarks*. Esta magni-
 122 tud puede tomar tres valores, que convencionalmente se etiquetan como *rojo*, *verde* y *azul*³.
 123 Mediante la carga de color se pueden explicar las predicciones de la cromodinámica cuántica.
 124 Hay seis tipos de quark, divididos en tres generaciones de manera análoga al caso leptónico,
 125 que se describen en la Figura 2.1. Todos ellos poseen carga eléctrica, siendo la misma $\frac{2}{3}e$
 126 en el caso de los quarks *up* (*u*), *charm* (*c*) y *top* (*t*); y $-\frac{1}{3}$ en el caso de *down* (*d*), *strange* (*s*) y
 127 *bottom* (*b*).

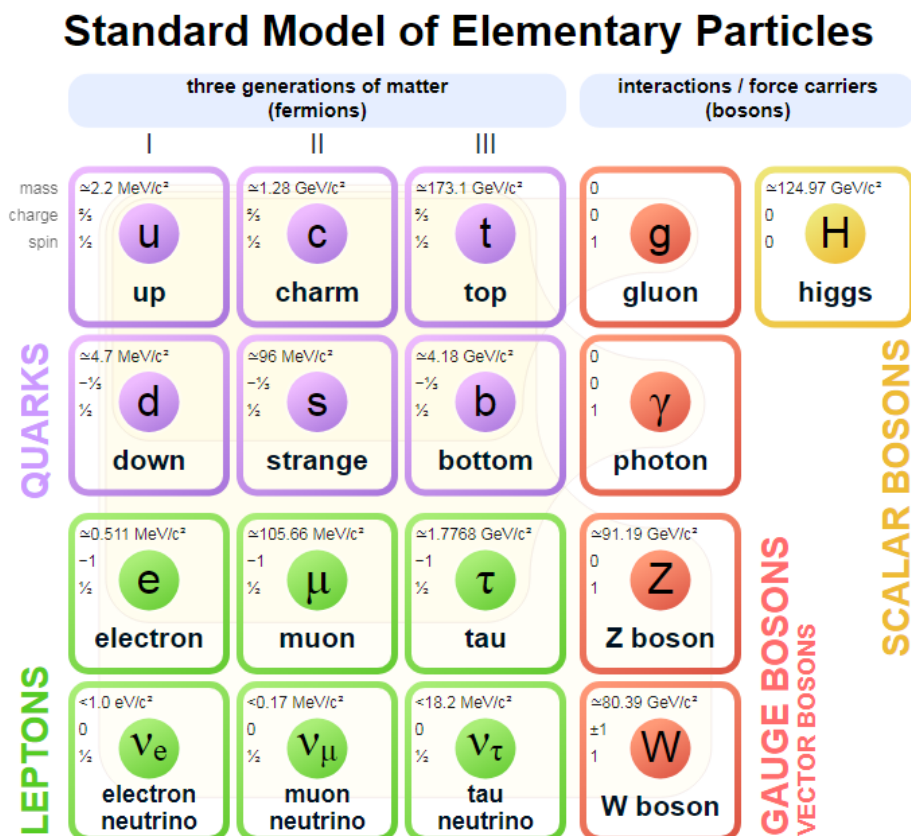


Figura 2.1: Tipos de partícula contemplados en el Modelo Estándar. Se representan los quarks (morado), leptones (verde) y bosones gauge (rojo y amarillo), diferenciando al importante bosón de Higgs. No se contemplan las antipartículas de los fermiones en el esquema. Tomado de [1].

128 Es notable la existencia de la *antipartícula* respectiva a cada tipo de fermión: una partícula
 129 idéntica a este en todos los aspectos excepto en su carga eléctrica, la cual es opuesta. De la
 130 misma manera, aparece un concepto análogo a las partículas con carga de color, en el cual
 131 se considera una carga opuesta a la que se denomina como *antirrojo*, *antiverde* o *antiazul*,
 132 respectivamente.

³No existe relación alguna entre estas etiquetas y los colores correspondientes del lenguaje común.

133 La interacción fuerte, al contrario que las otras interacciones, es proporcional a la inversa de
134 la distancia entre las partículas que interaccionan. Este hecho provoca que la interacción sea
135 más fuerte cuanto más se alejen las partículas, provocando que los quarks no puedan existir
136 libremente y se encuentren confinados en *hadrones*. Más aún, se han clasificado este tipo de
137 partículas en generaciones de masa creciente, basándose en sus números cuánticos. El estudio
138 de partículas de generaciones más altas, al ser más masivas, requiere experimentos de alta
139 energía.

140 Un ejemplo de hadrón que se encuentra entre los constituyentes de la materia ordinaria
141 es el del protón. El modelo de partones de Feynman explicaba como los hadrones están
142 compuestos por partes más pequeñas, a las que denominó *partones*. Años más tarde se verificó
143 que los hadrones estaban compuestos por quarks, y se puede observar como al interaccionar
144 con partículas de bajo momento se comportan como partículas libres cuyo momento es una
145 fracción x del momento del hadrón. Sin embargo, a medida que se aumenta el momento de
146 la partícula con la que interaccionan, se pudo observar como la descomposición del hadrón
147 es más compleja.

148 Experimentos con energía mucho más alta que las anteriores revelaron la existencia de *quarks*
149 *del mar*, que son constantemente originados por aniquilación de pares quark-antiquark en el
150 hadrón así como a través de gluones. La existencia de estos bosones en el hadrón fue validada
151 a través de la observación de que el momento de los partones no coincide con el momento
152 total del hadrón [2].

153 El protón se puede describir según lo descrito anteriormente como un hadrón con dos quarks
154 u de valencia y un quark d del mismo tipo.

155 2.2. Interacciones entre partículas según el modelo. Bosones

156 Como se introducía, el Modelo Estándar es una teoría cuántica de campos. En este tipo de
157 teorías, resulta crucial un tipo de invariancia sobre la energía de un sistema: la invariancia
158 *gauge* o *de calibrado*.

159 Una invariancia gauge en la descripción de un sistema físico provoca que se pueda describir
160 de manera equivalente por toda una clase de campos (escalares o vectoriales), relacionados
161 mediante *transformaciones de gauge*. Este tipo de transformaciones no tiene en cuenta sime-
162 trías espaciales que podrían surgir por un cambio de coordenadas, sino redundancias en la
163 descripción matemática del sistema.

164 Las transformaciones de gauge que caracterizan el SM son elementos el grupo $SU(3) \times SU(2) \times$
165 $U(1)$ ⁴⁵. Los *campos de gauge* son aquellos que surgen debido a los generadores del grupo
166 anterior; y al ser cuantizados, se denomina *bosón de gauge* a cada uno de sus cuantos.

167 Los bosones de gauge descritos por el Modelo Estándar son 12: el fotón, que media la inter-
168 acción electromagnética; tres tipos de bosones que median la interacción débil; y ocho tipos

⁴El grupo especial unitario de grado n , $SU(n)$, se representa por las matrices unitarias $n \times n$ de determinante 1 y tiene estructura de grupo de Lie con su operación producto.

⁵El grupo unitario de grado n , $U(n)$, se representa por las matrices unitarias $n \times n$ y tiene estructura de grupo con su operación producto.

169 de gluones que median la interacción fuerte. En la Tabla 2.1 se pueden encontrar los bosones
 170 intercambiados en cada uno de los tipos de interacción.

Interacción	Partícula mediadora
Electromagnética	Fotón (γ)
Débil	Bosón Z , Bosón W^\pm
Fuerte	Gluón (g)

Tabla 2.1: Partículas mediadoras en cada interacción según el Modelo Estándar. Hay ocho tipos distintos de gluones, caracterizados por ocho cargas distintas de color.

La fuerza gravitatoria no se tiene en cuenta en el Modelo Estándar dado que su intensidad es varios órdenes de magnitud inferior al resto de interacciones (por ejemplo, es 33 órdenes inferior a la interacción débil [3]) y se considera despreciable en las escalas de energía con la que se trabaja. En la Figura 2.2 se pone de manifiesto este hecho, mediante el concepto de *rango* de cada fuerza derivado del principio de incertidumbre de Heisenberg:

$$\left. \begin{array}{l} \Delta E \Delta t \sim \hbar \\ E = mc^2 \end{array} \right\} \implies mc^2 \sim \frac{\hbar}{\Delta t} \sim \frac{\hbar c}{r} \implies r \sim \frac{\hbar}{mc}$$

171 .

172 La fuerza electromagnética se ve mediada por el fotón, un bosón de espín 1, sin masa ni
 173 carga eléctrica. Similarmente, el gluón media la interacción fuerte (cuya carga asociada es la
 174 de color) y posee espín 1, masa nula y carga eléctrica neutra. Este último es el único que
 175 posee carga de color, lo que hace que interactúe consigo mismo y permite su clasificación en
 176 ocho tipos distintos de gluón atendiendo a su valor.

177 La interacción débil es mediada por los bosones W^\pm y Z . El primero (que puede tener carga
 178 eléctrica e o $-e$, siendo e el módulo de la carga del electrón) tiene una masa $m = 80.379$
 179 GeV [4] ⁶. El segundo posee carga eléctrica neutra y masa $m = 91.1876$ [4]. Ambos bosones
 180 poseen espín 1.

181 El *bosón de Higgs*, observado en 2012 [6], es un bosón de espín 0 asociado al *mecanismo de*
 182 *Higgs* mediante el cual los bosones que median la interacción débil adquieren su masa.

183 Un primer estudio del Modelo Estándar empleando teoría de campos sin considerar el me-
 184 canismo citado anteriormente, requiere que el bosón que describe la interacción débil sea de
 185 masa nula para preservar la invariancia de gauge, lo que contradice las medidas experimenta-
 186 les sobre la masa de los bosones W^\pm y Z . El mecanismo de Higgs, mediante el cual se genera
 187 masa para bosones y fermiones sin violar la invariancia gauge, resuelve esta discrepancia.

188 Dicho mecanismo introduce un potencial de energía V (el *potencial de Higgs*) con dependencia
 189 de dos campos escalares complejos derivados de la interacción débil, agrupados en un doblete
 190 ϕ . Su peculiar forma de “sombrero mejicano” observable en la Figura 2.3 describe un equilibrio
 191 inestable del potencial en el origen, así como infinitos equilibrios estables alrededor de este.
 192 La excitación del campo provoca una elección de alguno de los mínimos del potencial, y se
 193 describe mediante el bosón escalar que adopta el nombre de bosón de Higgs. Esta ruptura

⁶Se emplean unidades naturales, como es usual en el área de física de partículas.

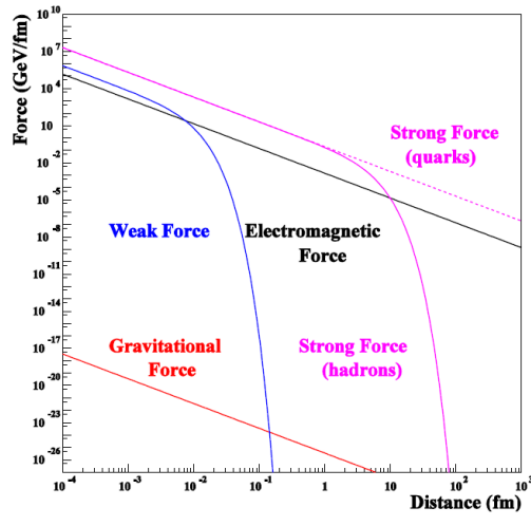


Figura 2.2: Rango de actuación de las fuerzas. Se comprueba que es inversamente proporcional a la masa. Tomado de [5].

194 en la simetría del potencial es el mecanismo mediante el cual las partículas mediadoras de la
 195 interacción débil adquieren una masa no nula.

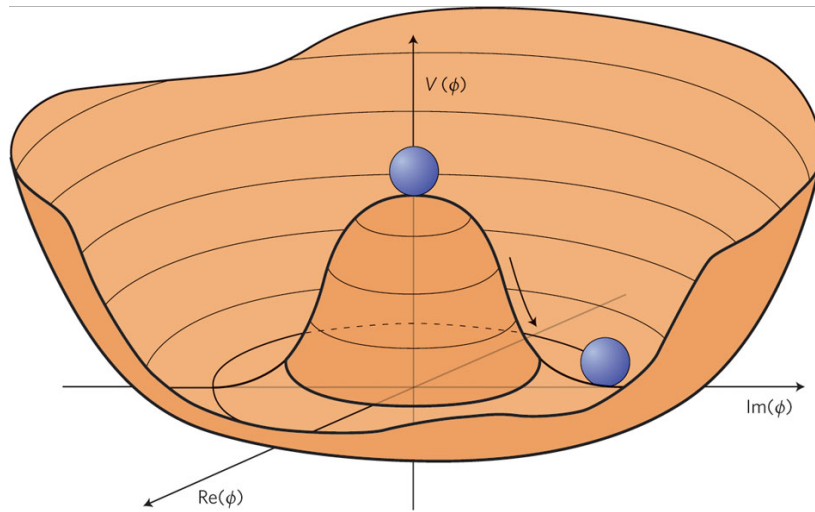


Figura 2.3: Forma del potencial de Higgs. Tomado de [7].

196 El campo de Higgs no solo permite explicar la generación de masa en los bosones W^\pm y Z .
 197 La *interacción de Yukawa* acopla un campo de Dirac, correspondiente con el del fermión, a
 198 un campo escalar, correspondiente con el bosón de Higgs. Esto se refleja en el lagrangiano del
 199 sistema con un término $y_F = \sqrt{\frac{m_F}{v}}$, donde m_F es la masa del fermión y $v \approx 246$ GeV. Así,
 200 el mayor valor de y_F se alcanza con el fermión de mayor masa, y considerando los fermiones
 201 explicados anteriormente se trata del quark t . Una explicación más detallada del mecanismo
 202 de Higgs y de los acoplamientos de Yukawa se puede encontrar en [8].

203 La masa del bosón de Higgs es de 125.18 GeV [4]. Hasta ahora, las propiedades predichas por
 204 el Modelo Estándar para dicho bosón concuerdan con las medidas experimentales. Ejemplo
 205 de esto son los acoplamientos de Yukawa a otros fermiones representados en la Figura 2.4.

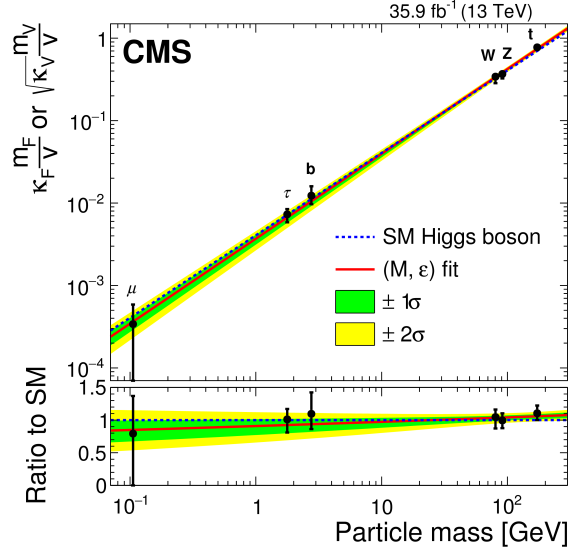


Figura 2.4: Medidas de los acoplamientos del bosón de Higgs a otras partículas del Modelo Estándar en función de sus masas. Tomado de [9].

206 2.3. Límites del Modelo Estándar

207 El Modelo Estándar describe de manera precisa las partículas conocidas en las escalas de
 208 energía con las que se trabaja, así como las interacciones entre ellas mediante tres fuerzas
 209 fundamentales. Sin embargo, pese a su éxito en las comprobaciones experimentales, este
 210 modelo no es capaz de explicar una serie de observaciones importantes.

211 El Modelo Estándar describe a los neutrinos como partículas sin masa: esto contradice el
 212 resultado de una gran cantidad de experimentos que prueban indirectamente la existencia
 213 de masa en los neutrinos, mediante la demostración de la existencia de **oscilaciones de**
 214 **neutrinos**.

215 Además, en el modelo no se proporciona una descripción de la **materia oscura**, cuya exis-
 216 tencia fue señalada en 1933 [10] a partir de la velocidad de dispersión de las galaxias miembro
 217 del cúmulo de Coma. La existencia de este tipo de materia era necesaria para que el siste-
 218 ma se mantuviese estable. Además, tampoco se proporciona por el modelo una explicación
 219 sobre la **energía oscura**, cuya existencia explica la aceleración en la expansión del universo
 220 (aceleración observada con ayuda del *HST*, *Hubble Space Telescope*, en 1998) [11].

221 La materia oscura debería sufrir únicamente interacciones débiles con la materia observable
 222 y sus radiaciones; sin embargo, a pesar de haberse practicado experimentos con el objetivo
 223 de observar directamente dicha interacción (como el *WIMP*, *Weakly Interacting Massive*
 224 *Particles* [12]), estos no han tenido éxito hasta ahora.

225 La energía oscura permite explicar la naturaleza expansiva acelerada del universo. Unas medi-
226 das proporcionadas por la NASA en 2016 [13] describen a aproximadamente un 26.8% de la
227 densidad de masa-energía del universo como materia oscura; cerca de un 68.3% como energía
228 oscura, y el porcentaje restante (alrededor de 4.9%) como la energía que el Modelo Estándar
229 describe.

230 Las grandes estructuras cosmológicas estén compuestas por materia en lugar de antimateria.
231 El mecanismo que momentos después del *Big Bang* produjo esta **asimetría** produciendo
232 ambos tipos de partícula de manera desequilibrada no se explica en este modelo⁷.

233 Más aún, la **gravedad** no ha podido ser incluida en el modelo como otra interacción gauge
234 mediante un bosón (el *gravitón*) debido a dificultades en la formulación matemática de una
235 teoría consistente de la gravitación cuántica. Además, se puede observar como la fuerza
236 gravitacional resulta mucho más débil que las otras interacciones, tales como la débil. Sin
237 embargo, el modelo no proporciona una explicación de este suceso.

238 El Modelo Estándar sufre la dependencia de **19 parámetros** que son medidos experimen-
239 talmente. Además de no permitir una predicción del valor de dichos parámetros, tampoco
240 explica el motivo por el que existen tres generaciones de partículas, ante la posibilidad de
241 que exista un número distinto de ellas.

242 Esta lista de problemas del modelo sugieren que el SM permite describir de manera precisa
243 las observaciones en los dispositivos experimentales debido a que sirve correctamente como
244 un límite de otro modelo más fundamental en bajas energías. Los modelos que tratan de
245 completar al Modelo Estándar son conocidos como modelos *Más Allá del Modelo Estándar*
246 (*BSM, Beyond Standard Model*).

247 Uno de los modelos más populares es el modelo de *Supersimetría, SUSY*. Si bien se han
248 planteado distintas versiones de dicho modelo, el principio común en todas ellas reside en
249 una nueva simetría entre bosones y fermiones. Esto predice una unificación de las interaccio-
250 nes gauge a una energía mayor y propone partículas que podrían explicar la materia oscura.
251 Además explica la diferencia de varios órdenes de magnitud entre entre las fuerzas débil y
252 gravitacional. No obstante, en la actualidad no ha sido posible encontrar pruebas experimen-
253 tales que sustenten dicho modelo en los experimentos del *CERN* descritos en el siguiente
254 capítulo.

⁷El SM permite explicar esta asimetría mediante violaciones de la *simetría CP* mediante interacción débil [14]. Existen fuentes de violación de esta última, pero sin embargo no son las suficientes para explicar la diferencia actual entre materia y antimateria.

255 Capítulo 3

256 El Gran Colisionador de Hadrones 257 y su detector CMS

258 En 1954 fue fundada la *Organización Europea para la Investigación Nuclear*, conocida por
259 las siglas de su antiguo nombre en francés *CERN*. Ubicada en Ginebra, en la frontera franco-
260 suiza, dispone del Gran Colisionador de Hadrones, dispositivo empleado para verificar la
261 validez y los límites del Modelo Estándar descrito anteriormente.

262 La construcción del algoritmo sobre el que se basa este trabajo tomará los datos obtenidos
263 por este dispositivo mediante uno de sus detectores: el *CMS*.

264 3.1. El Gran Colisionador de Hadrones

265 El acelerador de partículas más grande y potente del mundo, el LHC, ha permitido el des-
266 cubrimiento del bosón de Higgs en sus detectores ATLAS y CMS en 2012 [6]. Más aún, a
267 través del mismo se pretende encontrar pruebas relacionadas con teorías BSM que permitan
268 desmentirlas o darles validez.

269 3.1.1. Descripción general del LHC

270 El LHC, construido entre 1998 y 2008, se caracteriza por su forma circular de unos 26.7 km
271 de longitud ([15]). El experimento que se lleva a cabo en el dispositivo consiste en la colisión
272 de protones e iones pesados a muy alta energía (en la escala de los TeV). A cada colisión se
273 le llama *suceso*, y es característico en el acelerador su alta luminosidad (que mide la cantidad
274 de sucesos por unidad de tiempo y por unidad de área) [16]. Dicha propiedad le permite
275 una recolección de datos más exhaustiva que otros aceleradores y una medida precisa de los
276 parámetros clave del SM.

277 Principalmente se producen colisiones protón-protón ($p-p$) con una energía de centro de masas
278 cuando se puso por primera vez en funcionamiento de $\sqrt{s} = 7$ TeV ([15]). A lo largo de los
279 años se ha logrado aumentar la dicha energía hasta $\sqrt{s} = 13$ TeV.

280 Los sucesos se producen en cuatro lugares bien localizados, donde se pueden encontrar los

281 experimentos principales: El *CMS* (*Compact Muon Solenoid*), el *ATLAS* (*A Thoroidal LHC*
282 *Apparatos*), el *ALICE* (*A Large Ion Collider Experiment*), y el *LHCb* (*LHC-beauty*). En la
283 Figura 3.1 se representa la localización de los experimentos sobre el LHC.

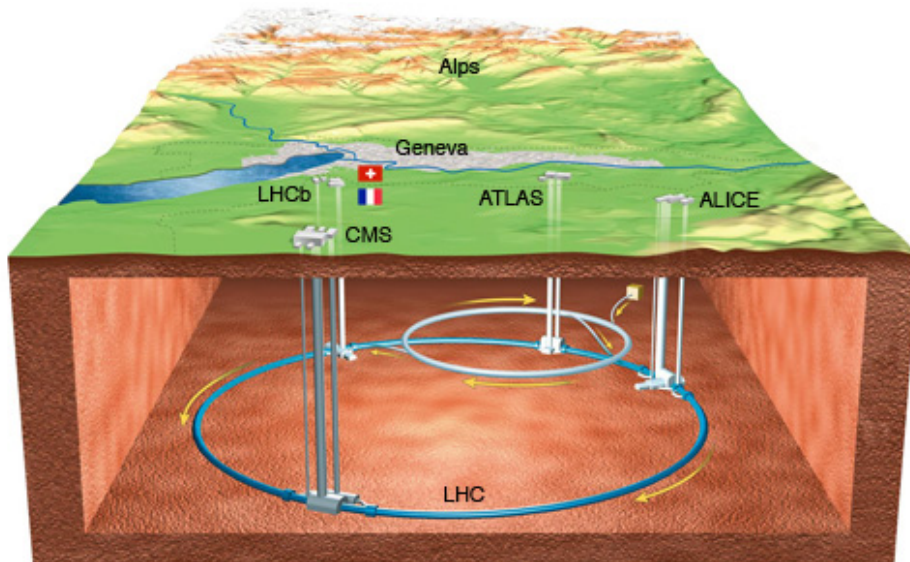


Figura 3.1: Esquema del LHC, donde se evidencian los puntos de interacción en los cuales se hacen colisionar los protones asociados a los cuatro grande detectores ATLAS, CMS, ALICE y LHCb.

284 3.1.2. Diferencias con respecto al LEP

285 El túnel empleado por el LHC pertenecía previamente al *LEP* (*Large Electron-Positron colli-*
286 *der*), cuya construcción se empezó en 1985 y terminó tres años después ([18]). El experimento
287 que se llevaba a cabo en él era la colisión entre electrones y positrones a altas energías.

288 Durante 11 años de investigación, los experimentos realizados en el LEP han supuesto un
289 avance en el estudio de la interacción electromagnética y débil, así como en el estudio sobre
290 el número de generaciones de fermiones.

291 Sin embargo, cuando una partícula cargada eléctricamente es acelerada radialmente, produce
292 una tipo de radiación (radiación de sincrotrón) que es proporcional a $1/m^4$, siendo m la masa
293 de la partícula acelerada. El hecho de que $m_p/m_e \approx 1836$ supone una ventaja clave del LHC
294 con respecto del anterior LEP en la producción de sucesos.

295 Las grandes energías a las que se llegan en el LHC permiten que en una colisión entre protones,
296 choquen componentes pertenecientes al *sea quarks* o a los gluones que lo componen, con
297 una fracción del momento del protón lo suficientemente grande como para producir sucesos
298 interesantes. Ejemplo de esto es la producción del bosón de Higgs mediante la fusión de
299 gluones.

300 **3.1.3. Inyección en el colisionador**

301 El LHC recibe protones previamente acelerados a la máxima energía posible en una serie de
 302 aceleradores construidos previamente, representados en la Figura 3.2.

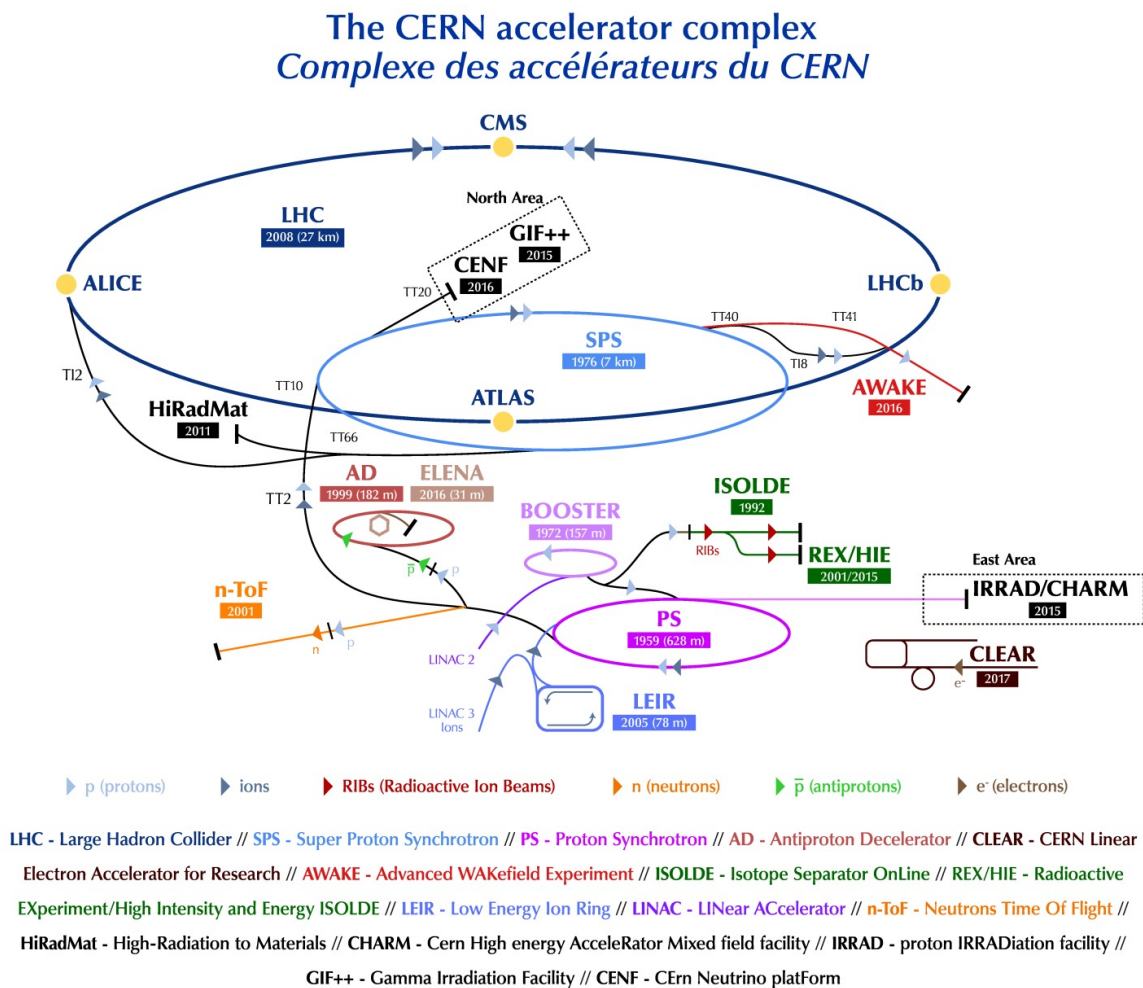


Figura 3.2: Diagrama que representa los distintos aceleradores previos al LHC. Se indica mediante flechas las direcciones de aceleración de los protones. Tomado de [19].

303 Para ello, se extraen protones de gas de hidrógeno mediante un campo eléctrico y se con-
 304 centran en “paquetes” (*bunches*) que se introducen en el acelerador lineal (*Linac*) para su
 305 aceleración hasta 50 MeV [20]. Posteriormente se inyectan en el acelerador protón-sincrotrón
 306 (*PSB*), donde alcanzan los 1.4 GeV ([21]). Los protones pasan después por el protón-sincrotrón
 307 (*PS*), el súper protón-sincrotrón (*SPS*) y finalmente son acelerados mediante imanes hasta el
 308 LHC. Así, alcanzan las energías respectivas en cada lugar de 26 GeV, 450 GeV ([22]) y, tras
 309 algo más de 25 minutos siendo acelerados con ayuda de los imanes ([23]), alcanzan su energía
 310 final.

311 Los haces estables de protones están compuestos por 2808 bunches de alrededor de 300 billones
312 de hadrones con una separación entre ellos de unos 25 ns a estas energías (en las que alcanzan
313 velocidades muy próximas a la de la luz) [23].

314 3.1.4. Imanes del LHC

315 El LHC se puede dividir en ocho octantes, en los cuales se puede encontrar un arco de unos
316 2.45 km de longitud, como se muestra en la Figura 3.3. Cada uno de estos arcos contiene 154
317 imanes dipolares (1232 en total) cuyo objetivo es mantener la trayectoria de los protones que
318 son acelerados hasta el lugar de su colisión [23].

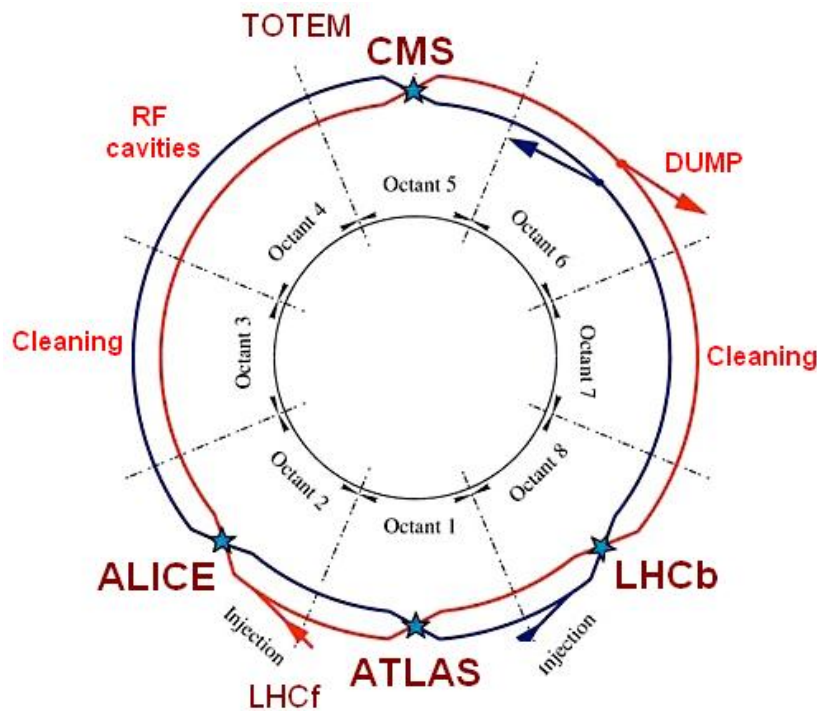


Figura 3.3: Distribución por octantes del LHC. En el esquema se representan los distintos experimentos que se practican.

319 Estos imanes forman parte de los 9593 de imanes con tecnología niobio-titanio ($NbTi$) con
320 los que cuenta el acelerador. Gracias a ellos, se logran campos magnéticos de hasta 8.3 T [15].

321 Los 392 cuadrupolos magnéticos ([24]) concentran los haces tanto vertical como horizontal-
322 mente, e imanes de más órdenes multipolares permiten refinar la geometría de los campos en
323 los extremos de los dipolos.

324 Para un campo magnético de estas magnitudes, es necesario mantener a los imanes en un
325 estado superconductor constante mediante un enfriamiento de su temperatura de hasta 1.9
326 K ([24]).

3.2. El detector CMS y sus componentes

El *Solenoides de Muones Compacto* (*Compact Muon Solenoid, CMS*) es uno de los cuatro experimentos principales ubicados en el LHC. Tiene forma cilíndrica, con 21 m de longitud, diámetro de 15 m y peso de 14000 toneladas ([25]). El detector ATLAS tiene 46 m de longitud, 25 m de ancho y alto, y pesa 7000 toneladas ([26]): es por este motivo que al CMS se le califica como “compacto”.

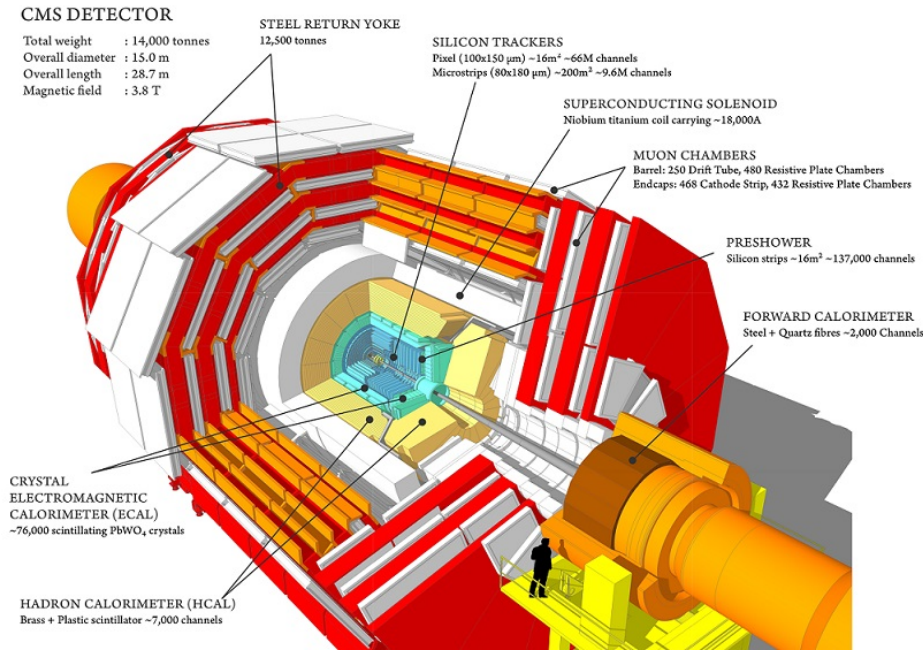


Figura 3.4: Esquema de los componentes del detector CMS. Las partículas aceleradas por el LHC son transportadas por el eje central del detector hasta su punto de colisión en la parte central del dispositivo. Tomado de [27].

El estudio practicado con ayuda de este detector sobre los sucesos producidos tiene un carácter general en contraposición a otros detectores, como el LHCb (especializado en la física del quark b) o el ALICE (especializado en colisiones de iones pesados). Además, es característico el alto campo magnético al que se llega en este detector (de 4 T, [25]), que se traduce en una gran precisión al medir el momento de las partículas cargadas (con ayuda de su detector de trazas de silicio, o *silicon tracker*).

3.2.1. Sistema de coordenadas

Para el estudio de sucesos producidos en el CMS, se considera un espacio tridimensional cuyo origen se sitúa en el lugar de colisión, y el eje Z del mismo tiene la dirección longitudinal del detector señalando al macizo del Jura, dirección que siguen las partículas aceleradas antes de la colisión. Así, los ejes X e Y describen el plano transversal a cada choque en el momento que se produce.

Este sistema permite describir variables como el momento transversal $p_T = \sqrt{p_x^2 + p_y^2}$, así

346 como los ángulos polar (θ , medido en el plano $Y - Z$ desde el eje Y) y azimutal (ϕ , medido
 347 en el plano $X - Y$ desde el eje X). Mediante dichos ángulos se puede definir la pseudorapidez
 348 $\eta = -\ln \tan(\theta)$, que se aproxima a infinito cuando θ se aproxima a 0, o describir todas las
 349 direcciones que pasan por el centro de coordenadas mediante el par (η, ϕ) .

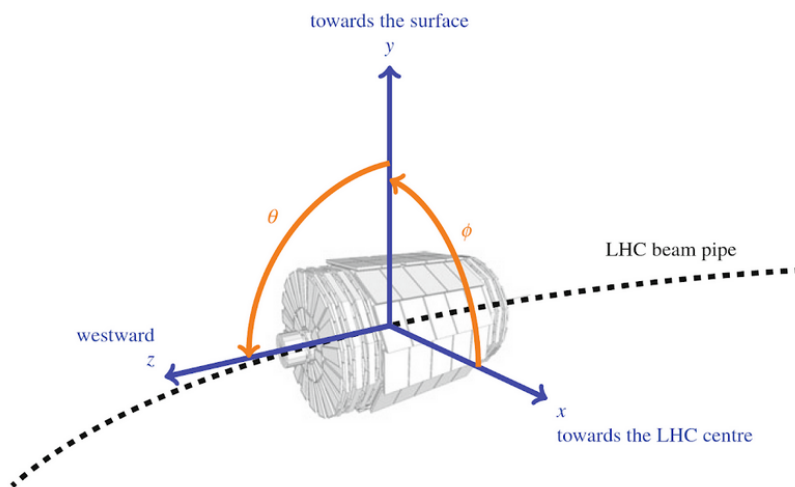


Figura 3.5: Descripción de los ángulos θ y ϕ en el CMS. Tomado de [8].

350 3.2.2. Solenoide magnético

351 Un componente clave en el experimento CMS es el solenoide superconductor, de forma cilíndrica,
 352 que genera un campo magnético paralelo a las partículas aceleradas de 4 T. Gracias a
 353 este componente de 28 m^2 de superficie transversa, 12.5 m de largo, y 12500t de peso [28],
 354 las partículas cargadas en el plano transversal ven su trayectoria curvada (consecuencia de la
 355 *ley de Lorentz*).

356 En el interior del solenoide se localizan el detector de trazas y los calorímetros del CMS (lo
 357 que evita que las medidas de estos detectores se vean comprometidas por la interacción de
 358 las partículas con el propio material del solenoide). Gracias al alto campo magnético que
 359 se genera, se puede determinar sin ambigüedad la carga y el momento transversal de las
 360 partículas producidas en los sucesos.

361 3.2.3. Detector de trazas

362 La medida del momento transversal de las partículas cargadas producto de cada colisión
 363 medida por el CMS se lleva a cabo a partir del detector de trazas, o *tracker*. El dispositivo
 364 se encuentra en la parte más cercana al punto de interacción.

365 Las partículas cargadas depositan energía por ionización en semiconductores de silicio, crean-
 366 do pares electrón-hueco (*hits*) que son detectados mediante electrodos que producen una se-
 367 ñal. La reconstrucción de la trayectoria de la partícula mediante estos *hits* resulta sumamente
 368 precisa (cada medida tiene una precisión de $10 \mu\text{m}$ [30]).

369 La parte interna del detector se compone por tecnología de píxeles de silicio y está sometida



Figura 3.6: Solenoide magnético en el momento de su ensamblaje. Tomado de [29].

370 a un alto flujo de partículas. La forman tres capas cilíndricas de detectores a 4 cm, 7 cm
 371 y 11 cm del lugar donde se produce el suceso ([31]). La detección de los primeros puntos
 372 de las trayectoria que se forman tras la colisión (conocidos como *seeds*) supone una parte
 373 importante en la reconstrucción de estas.

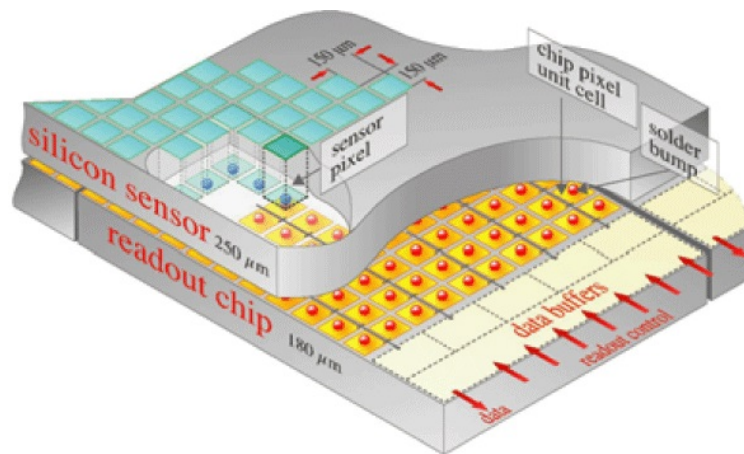


Figura 3.7: Pixel de silicio del detector de trazas. Tomado de [31].

374 La parte externa del detector de trazas la componen capas de láminas de silicio y su límite
 375 está a los 130 cm del eje Z [32]. Dichas láminas funcionan de manera similar a la parte interna
 376 del detector.

377 **3.2.4. Calorímetros**

378 Los calorímetros, situados a continuación del detector de trazas, tienen como propósito medir
 379 de manera precisa la energía de las partículas. Se dispone de un calorímetro electromagné-
 380 tico, que detecta la energía de electrones y fotones, así como de un calorímetro hadrónico
 381 especializado en los jets de hadrones.

382 El *calorímetro electromagnético (ECAL)* está formado por unos 80000 cristales de $PBWO_4$
 383 que actúan a modo de centelleador ante partículas cargadas de pseudorapidez $|\eta| < 3$ [33]. Se
 384 elige este material por su resistencia a la radiación, rápida emisión de luz y buena sensibilidad.

385 Una de las funciones principales del *ECAL* es la medida de masas invariantes en procesos
 386 difotónicos, ya que pueden ser producto del proceso $H \rightarrow \gamma\gamma$.

387 El *calorímetro hadrónico (HCAL)* mide la energía de las partículas que sufren principalmente
 388 de interacción fuerte tras atravesare el *ECAL*. Es capaz de detectar partículas con pseudora-
 389 pidez $|\eta| < 5.2$ a lo largo de sus cuatro capas (apreciables en la Figura 3.8) ([34]).

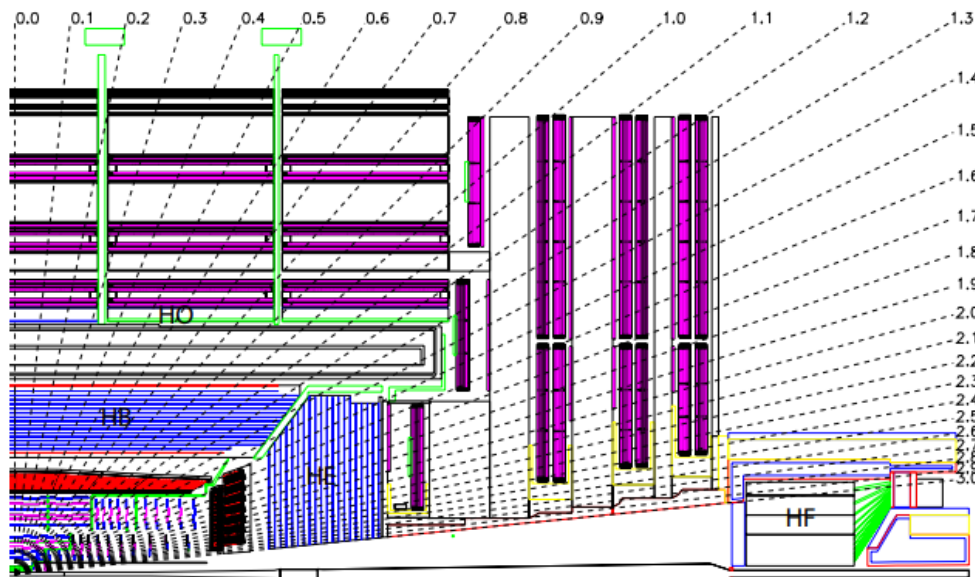


Figura 3.8: Esquema de los componentes del calorímetro hadrónico donde se representa tam-
 bién la pseudorapidez de cada dirección. Tomado de [34].

390 Dos de las capas que lo componen, el “barril” (*Hadron Barrel, HB*, $|\eta| < 1.4$) y las “tapas”
 391 (*Endcaps, HE*, $1.4 < |\eta| < 3$), están situadas dentro del solenoide. Las partículas de alta
 392 energía que no llegan a ser frenadas por estas capas son absorbidas por los imanes antes de
 393 llegar al *Outer (HO)* situado justo a continuación. Una última capa, el calorímetro “delantero”
 394 (*Forward Calorimeter, HF*, $2.9 < |\eta| < 5.2$) detecta las productos con mayor pseudorapidez
 395 y está expuesto a los flujos de partículas más altos ([34]).

396 **3.2.5. Cámaras de muones**

397 Uno de los objetivos principales del detector CMS era la identificación de muones, como se
 398 deduce del nombre del experimento. Situada tras el *HO*, la *cámara de muones* mide con
 399 precisión y robustez estos leptones, los cuales han perdido poca energía (en comparación con
 400 los electrones) al pasar por el detector de trazas. Más aún, se ven separados del resto de
 401 productos que se han generado también en el mismo suceso mediante el resto de componen-
 402 tes del detector¹. (los neutrinos no son apantallados). Al estar situado tras el solenoide, los
 403 muones que llegan a esta región sufren un campo magnético en sentido contrario a las partí-
 404 culas cargadas en el interior del detector, provocando que su trayectoria se curve en sentido
 405 contrario.

406 Este componente del CMS se puede dividir en su zona de “barril” (en radios del cilindro
 407 de valores entre 4 y 7.5 m) y su zona de “tapas de barril” (situadas entre 5.5 y 11 m en
 408 la dirección longitudinal) y consigue medir partículas de hasta $|\eta| < 2.4$. Se compone
 409 de tres tipos de detector: *los tubos de deriva* (*Drift Tubes, DT*, $|\eta| < 1.2$, las *cámaras de tiras*
 410 *catódicas* (*Cathode Strip Chambers, CSC*, $1 < |\eta| < 2.4$) y las *cámaras de placas resistivas*
 411 (*Resistive Plate Chambers, RPC*, $|\eta| < 1.6$) [35].

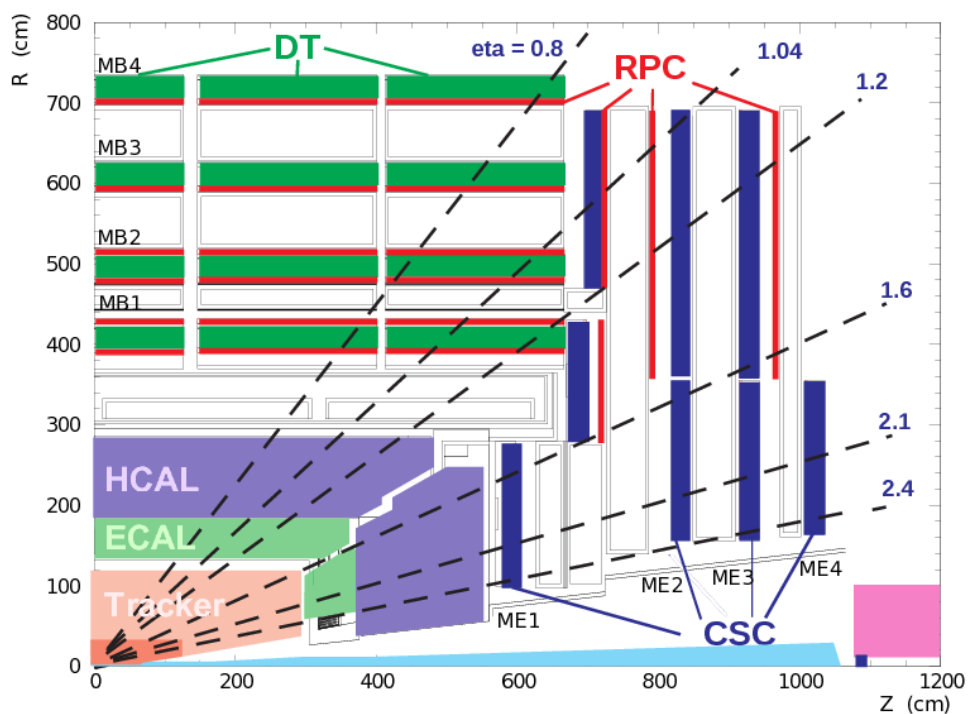


Figura 3.9: Diagrama de los componentes de la cámara de muones. Tomado de [36].

412 Los *DT* son tubos de longitud 2.4 m y sección transversa $1.3 \times 4.2 \text{ cm}^2$, compuestos por
 413 células de gas (*Drift Cells*) ([34]). El contenido de las células es ionizado con el paso de un
 414 muon, de forma que se genera un electrón y se transmite a la parte central del compartimento

¹Los neutrinos no son apantallados al apenas interactuar con la materia. Sin embargo, precisamente por este motivo, no son detectados tampoco en la cámara de muones.

415 donde se localiza el cable que transmite la señal (posteriormente amplificada). Se localizan
416 en una región donde el el flujo de partículas es menor.

417 Las *CSC* están situadas en las tapas del sistema de medida de muones. El campo magnético
418 en esta región es mayor y menos uniforme y por lo tanto se opta por usar un sistema de
419 medida distinto al de los *DT*. Este sistema es análogo, compuesto también por recipientes
420 rellenos de gas, pero cuando este es ionizado el electrón se expone a una serie de cables
421 actuando como ánodos y tiras de cobre ortogonales a ellos actuando como cátodos. Los iones
422 positivos serán por tanto transportados también hacia la salida. Los *CSC* se localizan en una
423 región donde hay una mayor cantidad de flujo de partículas.

424 Por último, *RPCs* son localizados en el barril y las tapas del sistema y proporcionan una
425 respuesta más rápida que la de los dos componentes anteriores, aunque con una resolución
426 espacial más gruesa (del orden de 1 cm, en comparación con las resoluciones entre 40 y 150 μm
427 de los *DT* y *CSC* [37]). Su función es la de filtrar de manera rápida las señales detectadas
428 para determinar si se conservan los datos recogidos. Esta función se desempeña mediante
429 compartimentos que contienen un gas ionizante, de manera análoga al resto de detectores,
430 situado entre dos placas cargadas transparentes a los electrones producidos en la ionización
431 del gas (este es recogido por una tira externa metálica).

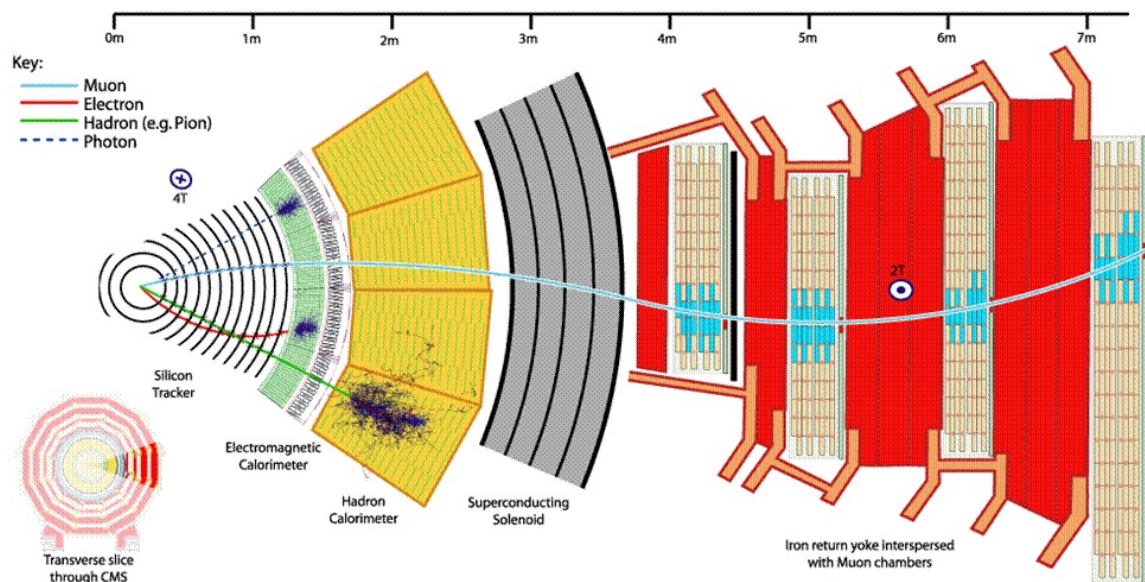


Figura 3.10: Resumen de las partículas detectadas en cada región del CMS. Tomado de [38].

432 3.2.6. Sistema de filtrado

433 Uno de los efectos que tiene la alta luminosidad del acelerador es la dificultad de analizar
434 todos los sucesos detectados en el CMS. Debido a que tan sólo una fracción de todas estas
435 observaciones tienen un interés físico, surge la necesidad de un sistema capaz de clasificar de
436 manera rápida cada una de las colisiones para decidir si sus datos deben ser almacenados o
437 no.

438 El sistema que lleva a cabo esta tarea en el detector es el *sistema de filtrado (trigger system)*.
439 En el CMS se puede dividir en dos niveles, $L1$ y $L2^2$, en cada uno de los cuales se reduce la
440 cantidad de información a una fracción que se considera interesante.

441 En el primer nivel de filtrado, $L1$, se reduce la cantidad de observaciones a unos 100 kHz. Es
442 un proceso muy rápido, que unos $4 \mu\text{s}$ después de cada colisión decide si el suceso podría tener
443 interés o no. Este sistema primeramente toma información de los calorímetros *ECAL*, *HCAL*
444 y *HF*, así como de los *DT*, *RPC* y *CSC* de la cámara de muones, reconstruyendo objetos
445 físicos a través de ella (tales como leptones o fotones). A continuación, los datos procesados
446 son evaluados en el *global trigger (GT)* y se realiza el filtrado sobre ellos [39].

447 En el siguiente nivel del filtrado, el $L2$, se recrea el suceso en su totalidad y determinan los
448 sucesos de interés a través de una gran cantidad de ordenadores, conocidos en su conjunto
449 como la *event filter farm (EVF)*. A finales de 2012, la *EVF* se componía por unos 13000
450 procesadores capaces de reducir la cantidad de sucesos a unos 400 Hz que son posteriormente
451 almacenados [39].

²El $L2$ también se conoce como *High Level Trigger, HLT*.

452 Capítulo 4

453 La reconstrucción de sucesos

454 En los detectores del LHC se recogen una gran cantidad de datos por cada suceso producido.
455 A pesar del filtrado producido para una disminución de datos de trabajo (introducido en
456 la sección 3.2), siguen habiendo muchos sucesos que, dependiendo del proceso físico que se
457 quiera estudiar, sería preciso incluir o rechazar en el análisis. Por ello se vuelven necesarias
458 técnicas computacionales para el manejo de todos ellos, ya sea en el análisis de las trazas
459 como en la identificación de los objetos producidos en cada suceso.

460 4.1. El rol de la simulación

461 Primeramente, la información obtenida por cada componente del detector (objetos bajo nivel,
462 o *low-level features*) es tratada, con el objetivo de la reconstrucción de los sucesos producidos
463 en cada colisión¹. El proceso de reconstrucción se realiza tras la toma de datos, y con un
464 enfoque en un proceso físico de interés, por lo que se realiza una selección de los datos
465 obtenidos en el experimento que sean de interés para el proceso. Los sucesos en los que se ha
466 producido el fenómeno físico de estudio se consideran *señal*, o *signal*; mientras que el resto
467 se considera *fondo*, o *background*.

468 Para determinar reglas de selección de los sucesos al clasificarlos como señal o fondo, se
469 emplean técnicas de simulación (a través de procesos de Monte Carlo) sobre los sucesos que
470 se reconstruyen. El objetivo es la observación del paso de cada suceso simulado por el detector
471 para contrastarlo con la información experimental.

472 La simulación del proceso y sus interacciones posteriores suele basarse en herramientas de
473 programas como *PYTHIA* [41]. Esta se combina con programas como *Geant4* [42] para la
474 simulación de su paso por el detector CMS.

475 La cantidad de procesos simulados resulta muy elevada (mayor que la cantidad de procesos
476 detectados) y la proporción de cada proceso simulado tiene en cuenta la cantidad total de
477 sucesos simulados así como la sección eficaz del proceso.

¹Tradicionalmente, el empleo de la información obtenida en los componentes de bajo nivel ha sido complicado debido a la alta dimensionalidad de los datos. Actualmente, se investigan formas de combinar esta información con la de niveles mayores mediante el empleo de técnicas de *Deep Learning* [40].

4.2. Algoritmo de reconstrucción e identificación. Trazas y vértice primario

Para reconstruir e identificar todas las partículas estables que se han producido en un suceso (electrones, muones, fotones, hadrones cargados y neutros) se maneja la información tomada de todos los detectores del CMS a través del algoritmo *Particle Flow* ([43]).

Mediante la información sobre la dirección, energía y tipo de las partículas, se procede a formar los jets de cada una de estas para determinar la energía transversa faltante (lo cual proporciona información sobre la energía y dirección de neutrinos u otras partículas invisibles). Con esta energía se puede, entre otras cosas, identificar leptones τ , cuantificar el aislamiento de un leptón respecto otras partículas o identificar jets formados por un quark b .

Mediante los *hits* detectados en el detector de trazas (sección 3.2.3) y con la ayuda de la señal en la cámara de muones (sección 3.2.5) se reconstruyen las trayectorias de las partículas involucradas en la colisión.

Otro punto de interés en el estudio es la localización del *vértice primario* de la colisión, es decir, el lugar donde se sitúa el choque entre los protones que se produce en el experimento. Para su situación, se atiende al cruce de las trayectorias detectadas con la zona central del detector.

Debido a que se pueden registrar varias colisiones simultáneamente en un mismo estudio (*pile-up*), el vértice primario será el que de lugar a la reconstrucción de objetos físicos con un mayor momento transverso.

4.3. Técnica de *Lepton MVA*

Con el objetivo de la identificación de electrones o muones en el análisis, se introduce la técnica de *Lepton Multivariate Analysis* (o *Lepton MVA*). Se trata de un análisis en múltiples dimensiones, en un contexto de *machine learning* (en particular, de un *BDT*; ver secciones 6.1 y 6.1.1), que permite distinguir leptones (electrones o muones) de sucesos de señal de los de sucesos de fondo. En particular, se puede practicar el análisis para distinguir leptones producidos en un proceso $t\bar{t}H$ de los producidos en $t\bar{t}$.

El algoritmo que se propone en el trabajo efectúa este tipo de análisis, en 13 dimensiones que se corresponden cada una con una medida sobre el leptón que se estudia. Dichas medidas, o variables, estudian las siguientes características del leptón:

- El momento transverso, p_T .
- La pseudorapidez, η .
- El número de trazas cargadas encontradas dentro del cono considerado.
- El mini-aislamiento del leptón respecto a partículas cargadas del jet.
- El mini-aislamiento del leptón respecto a partículas neutras del jet.
- El cociente del momento transverso del leptón y el del jet, p_T^{ratio} .

- 514 ▪ El momento transverso relativo entre el leptón y el jet, p_T^{rel} .
- 515 ▪ Algoritmos de b -tagging de target hadrónico.
- 516 ▪ Algoritmos de b -tagging de target leptónico.
- 517 ▪ El parámetro de impacto en el plano transverso, d_{xy} .
- 518 ▪ El parámetro de impacto en el plano longitudinal, d_z .
- 519 ▪ La significancia del parámetro de impacto, SIP.
- 520 ▪ El valor del clasificador bayesiano ingenuo para compatibilidad de trazas de muones
521 (no aplicable en electrones).
- 522 ▪ El valor del bosque aleatorio identificador de electrones (no aplicable en muones).

523 En esta lista, las variables dependen en general de la cinemática del leptón, estudiada a través
524 de los depósitos de energía que deja a lo largo de su trayectoria en su paso por el detector y
525 mediante una comparación con el propio momento del leptón. Se estudiará cada una de ellas
526 en más detalle en la sección 5.6, donde se describen sus valores en la muestra que toma el
527 algoritmo.

528 4.4. Reconstrucción e identificación de leptones

529 Para la reconstrucción de la trayectoria de muones, se emplean las señales del detector de
530 trazas junto con la de las cámaras de muones. Su posterior identificación emplea el algoritmo
531 *leptón MVA* descrito en la sección anterior, entrenado con la señal y el fondo que caracterice
532 el análisis que se está considerando.

533 De manera análoga, la reconstrucción de la trayectoria de electrones tiene en cuenta la señal
534 del detector de trazas esta vez junto con la obtenida en el calorímetro electromagnético. En
535 este proceso se está teniendo en cuenta que la señal en el *ECAL* puede proceder de otras
536 partículas con carga eléctrica, así como que la radiación de frenado es mayor que la de los
537 muones debido a su mayor masa. La posterior identificación del leptón se vale, de nuevo, de
538 algoritmos de tipo *leptón MVA* o *Particle-Flow*.

539 El estudio de leptones τ deja de ser análogo a los anteriores, debido a que su masa le permite
540 desintegrarse quarks y producir jets hadrónicos no contemplados en el análisis anterior. Su
541 otro modo principal de desintegración produce un electrón o un muon.

542 El modo de desintegración leptónico del leptón τ resulta difícil de identificar, ya que a pesar
543 de poder reconstruirse el leptón en el que decae, resulta complicado deducir que proviene
544 del τ . Por otra parte, hay algoritmos especializados en la reconstrucción de τ que decaen en
545 hadrones (como el algoritmo *hadron-plus-strips* [44]).

546 4.5. Otras componentes de la reconstrucción

547 Para la reconstrucción de los jets hadrónicos producidos en las colisiones, se emplea el algo-
548 ritmo *Particle-Flow* junto con otros algoritmos que colaboran a este proceso (por ejemplo, el

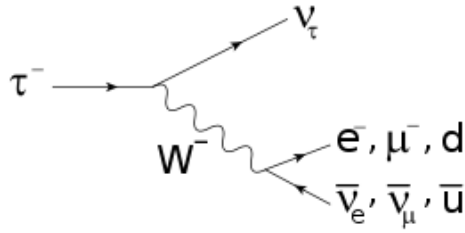


Figura 4.1: Algunos modos de desintegración del leptón τ . La desintegración en quarks \bar{u} y d produce jets hadrónicos.

549 algoritmo *anti- k_t jet clustering* [45]).

550 En la identificación de los jets, donde se emplean técnicas de análisis multivariable, una de las
 551 características que se estudian es la formación de vértices secundarios ciertamente alejados
 552 del primario, señal de que se ha podido producir una desintegración de un hadrón compuesto
 553 por un quark b (ver Figura 5.22). Con el objetivo de identificar dichos vértices, se pueden
 554 emplear algoritmos como *Combined Secondary Vertex (CSV)*, [46]).

555 El estudio de las partículas que no interaccionan con el detector, las cuales son principalmente
 556 los neutrinos de los procesos², se practica atendiendo a la *energía transversa faltante* del
 557 suceso.

558 El principio de conservación del momento en la colisión entre partículas permite comparar el
 559 momento (transverso) esperado tras la colisión (a través del momento de las partículas antes
 560 de esta³) frente al momento total medido (suma de los momentos de cada una de las partículas
 561 cuya trayectoria fue reconstruida). Una diferencia entre el valor medido y el esperado, que
 562 se puede expresar por un vector, denota la existencia de partículas en los productos cuyo
 563 momento no ha sido medido. El momento transverso faltante es por tanto un vector que
 564 al sumarse a los vectores momento de los objetos construidos restaura la conservación del
 565 momento inicial.

²Los neutrinos pueden interaccionar de manera muy débil con la materia (como se introducía en secciones anteriores, tienen masa, aunque sea significativamente pequeña). De la misma manera, otras partículas no descritas por el Modelo Estándar pueden no interaccionar con el detector.

³Se asume que el momento transverso inicial sea 0 debido a que la colisión se produce en el eje longitudinal.

566 Capítulo 5

567 La detección del bosón de Higgs

568 Las propiedades del bosón de Higgs pueden ser estudiadas mediante procesos que producen
569 quarks t . En este capítulo se introduce la razón de este motivo, así como la descripción de
570 los procesos $t\bar{t}$ y ttH , muy importantes en este contexto.

571 5.1. El proceso ttH

572 Como se introdujo en la Sección 2.2, la fuerza de la interacción de Yukawa relacionada con el
573 mecanismo de Higgs resulta mayor para fermiones con más masa, siendo el quark top el más
574 masivo de todos los observados. Es por ello que la medida del acoplamiento entre H y t cobra
575 importancia para la comprobación experimental de las predicciones del Modelo Estándar.

576 Una forma de medir el acoplamiento entre el bosón de Higgs con otras partículas es mediante
577 los canales de desintegración del bosón. Una mayor interacción entre las partículas viene
578 acompañada de una mayor probabilidad de desintegración en el canal dado. Se ha podido
579 observar la desintegración del bosón en pares $\gamma\gamma$, ZZ , W^+W^- , $\tau^-\tau^+$ y $b\bar{b}$; sin embargo, la
580 ley de conservación de la energía impide que un bosón de Higgs se desintegre en un par $t\bar{t}$ al
581 el valor de su masa observado en 2012 [6] menor que el de las partículas producidas.

582 La medida de la interacción entre el bosón H y el par $t\bar{t}$ debe por lo tanto tomar un enfoque
583 distinto. Al igual que la probabilidad de desintegración viene relacionada con el acoplamiento,
584 la probabilidad de la producción del bosón de Higgs también. Las colisiones que muestran la
585 presencia de las partículas t , \bar{t} y H serán clave para obtener información sobre el acoplamiento
586 de Yukawa entre H y t . A esta producción de bosones de Higgs se la conoce como "producción
587 asociada al bosón de Higgs con dos quarks top", o producción ttH .

588 Sólo hay dos posibles diagramas de Feynman que involucran el acoplamiento entre el quark
589 t y el H . Estos son la creación de dos pares $t\bar{t}$ en la que un t de un par se fusiona con el \bar{t} del
590 otro, dando lugar al bosón H ; o la creación de un solo par $t\bar{t}$ en la que se radia un bosón H de
591 uno de los quarks. En ambos casos, el bosón de Higgs fue producido mediante los quarks top,
592 por lo que la observación de este estado final supone una medida directa del acoplamiento
593 de H a t . Los diagramas de Feynman asociados a los dos distintos casos se describen en la
594 Figura 5.1.

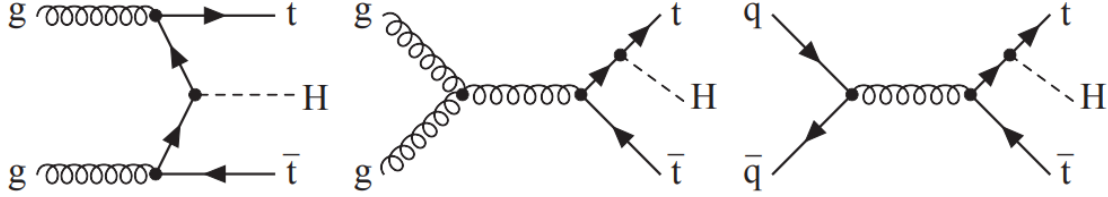


Figura 5.1: Ejemplos de procesos $pp \rightarrow t\bar{t}H$ producidos en el LHC. El diagrama de la izquierda se corresponde con una producción del bosón H mediante la fusión de un par $t\bar{t}$, mientras que los otros dos suponen la radiación del bosón de un quark t . Tomada de [47].

595 El estudio de la producción ttH no es trivial, debido a la gran cantidad de canales de de-
 596 sintegración que tienen las partículas en cuestión, así como a la rareza de que se produzcan
 597 y la similitud que comparten con otros sucesos de fondo más frecuentes atendiendo a sus
 598 productos finales.

599 5.2. Desintegraciones del bosón H

600 Para poder estudiar los sucesos en los que ha habido producción ttH , es necesario buscar
 601 indicios entre los productos de la colisión de que, en efecto, intervinieron un bosón H y un
 602 par $t\bar{t}$. Por ello son de especial interés los modos de desintegración de estas partículas.

603 El decaimiento del bosón de Higgs en un fermión y su antipartícula es uno de los principales
 604 canales de desintegración observados. Este canal se puede representar mediante el diagrama
 605 de la Figura 5.2, y como se indicaba anteriormente, no considerara entre los estados finales
 606 de esta desintegración dos quarks t para no violar la conservación de energía.

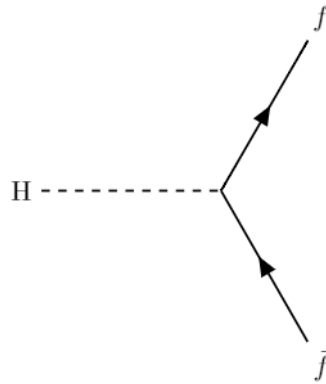


Figura 5.2: Desintegración de H en un fermión f y su antipartícula \bar{f} .

607 Otro canal de desintegración del bosón de Higgs es aquel en el que se desintegra en dos
 608 bosones, produciéndose dos posibles casos: $H \rightarrow W^+W^-$ y $H \rightarrow ZZ$. Los diagramas que
 609 representan este decaimiento se encuentran en la Figura 5.3.

610 Por último, se puede considerar la desintegración del bosón H en dos gluones, $H \rightarrow gg$. Este

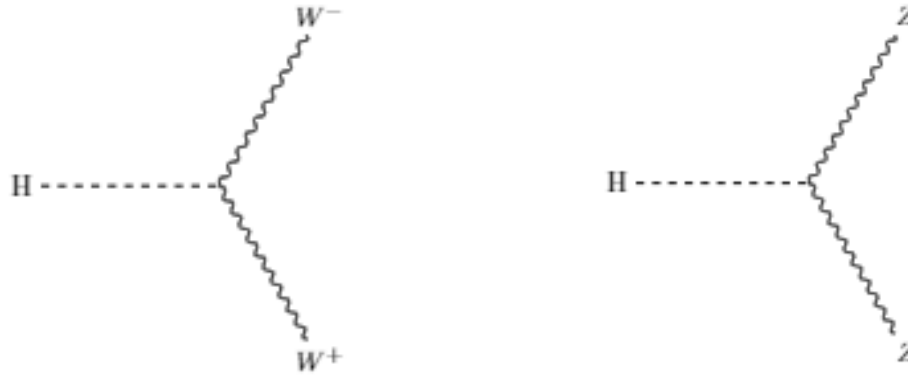


Figura 5.3: Desintegración de H en dos bosones W^+W^- (izquierda) o ZZ (derecha).

611 proceso se obtiene de ciclos de quarks pesados (como se muestra en el diagrama de la Figura
 612 5.4), y por lo tanto la probabilidad de desintegración en este canal dependerá de la masa de
 613 los quarks del ciclo.

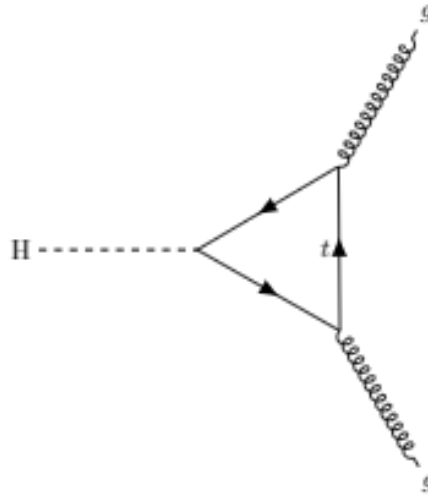


Figura 5.4: Ejemplo de desintegración de H en gluones g . El ciclo lo define un quark pesado, en este caso t .

614 La dependencia de la fracción de desintegración¹ con la masa que se le asigna al bosón de
 615 Higgs no resulta trivial, y puede observarse en la Figura 5.5.

¹*Branching Ratio, BR*. Se trata del cociente de la amplitud parcial de desintegración asociada a un canal en particular entre la amplitud parcial de desintegración total ($BR = \Gamma_{canal}/\Gamma_{total}$).

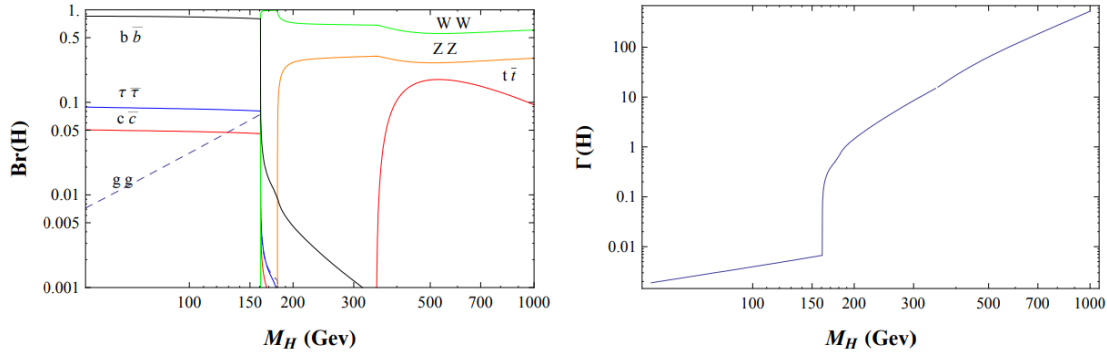


Figura 5.5: Fracciones de desintegración del bosón H (izquierda) y amplitud parcial de desintegración total (derecha) en función de la masa del bosón. Tomado de [48].

616 5.3. Desintegraciones del par $t\bar{t}$

617 Siguiendo con el análisis de la producción de $t\bar{t}H$ en el experimento CMS , es preciso centrarse
 618 también en los productos asociados a la producción de un par de quarks $t\bar{t}$. En particular,
 619 se puede estudiar como se desintegra un único quark t , ya que el comportamiento de su
 620 antipartícula será análogo.

621 La desintegración del quark t tiene como producto un bosón W y un quark, que usualmente
 622 es el quark bottom b . El quark top es el único lo suficientemente pesado para producir un
 623 bosón W real, que media la interacción asociada a esta desintegración.

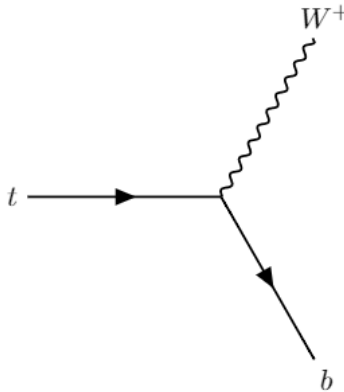


Figura 5.6: Desintegración de t en un bosón W^+ y un quark b .

624 El quark en el que se desintegra el t puede ser d , s ó b . La proporción de cada uno de
 625 estos canales de desintegración es proporcional a $|V_{tq}|^2$, donde V_{tq} es el elemento de matriz
 626 correspondiente de la *matriz de Cabibbo-Kobayashi-Maskawa* ($q = d, s, b$). Definiendo R como
 627 la razón entre la fracción de desintegración de t en W, b y la fracción de desintegración de t
 628 en W con cualquier otro quark, se puede apreciar como el canal usual de desintegración es

629 $t \rightarrow Wb$:

$$R \equiv \frac{BR(t \rightarrow Wb)}{BR(t \rightarrow Wq)} = \frac{|V_{tb}|^2}{|V_{tb}|^2 + |V_{ts}|^2 + |V_{td}|^2} \approx 0.998.$$

630 El valor de este cálculo (con valores tomados de [4]) se ha comprobado de manera experimental
631 en diversos detectores (en particular, se obtuvo una medida de $R = 1.01 \pm 0.03$ en el CMS
632 en 2014 consistente con lo predicho [49]).

633 A su vez, el bosón W sufre un proceso de desintegración cuando es producido, que se tiene en
634 cuenta en el análisis posterior para la reconstrucción del suceso. Este bosón tiene la capacidad
635 de desintegrarse en un leptón con su misma carga (se denotará por l) y su neutrino asociado
636 (se denotará por ν_l); así como en dos quarks (se denotarán por q y \bar{q}'). Es preciso recordar
637 que el quark t posee una masa mayor que la del bosón y no se encontrará por tanto entre los
638 productos de su desintegración.

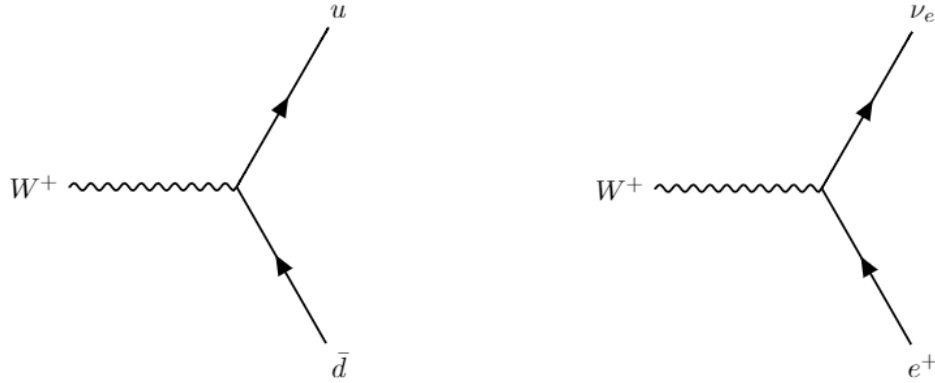


Figura 5.7: Ejemplos de desintegración de W^+ en quarks (izquierda, u y \bar{d}) o leptones (derecha, e^+ y ν_e).

639 La desintegración de un par $t\bar{t}$ se reconstruye, por tanto, a partir de los jets originados por
640 dos quarks b y los productos de la desintegración de dos bosones W de carga opuesta.

641 5.4. Fondo en el análisis

642 A menudo, la selección de sucesos que contienen quarks t se basa en la identificación de un
643 leptón cargado (aislado) producido por la desintegración de un bosón W , al que se refiere
644 como leptón *real* o *prompt*. Es posible que otras partículas sean confundidas con este leptón
645 real, las cuales suponen el fondo (*background*) leptónico *non-prompt* o *false* (*fake*).

646 Algunos sucesos que colaboran en dicho fondo asociado a electrones, son aquellos en los
647 que se producen desintegraciones semileptónicas de quarks b o c , así como las conversiones
648 de fotones altamente energéticos (procesos de Drell-Yan). En estos sucesos se produce un
649 electrón *non-prompt* que puede confundirse con el producto de una desintegración del bosón

650 W . También hay una contribución de leptones *fake* por parte de jets con una gran energía
 651 electromagnética que pueden ser identificados falsamente como electrones.

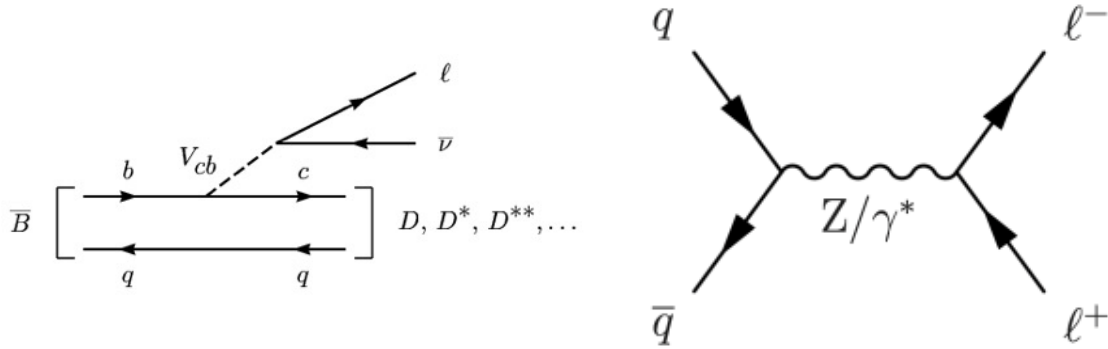


Figura 5.8: Ejemplos de procesos que producen fondo en la identificación de electrones *prompt*. Desintegración semileptónica de un hadrón B (izquierda, tomado de [50]) y producción de un electrón y un positrón mediante un fotón (derecha, tomado de [51]).

652 De la misma manera, el fondo asociado a muones considera las desintegraciones semileptónicas
 653 de quarks b o c , así como otras desintegraciones de hadrones cargados que pueden originar
 654 muones *non-prompt*. Estos muones se encuentran menos aislados que los *prompt*, así como
 655 poseen un mayor parámetro de impacto.

656 Más aún: los calorímetros hadrónicos detectan la energía de los hadrones que absorben al
 657 interactuar con ellos, y produciendo lo que se conoce como una “cascada hadrónica”. En
 658 ocasiones, hadrones con un alto momento pueden producir muones secundarios en esta casca-
 659 da que no son absorbidos por el calorímetro y contribuyendo al fondo de muones *non-prompt*
 660 (se conocen como muones *punchthrough*). Estos leptones son producidos generalmente en la
 661 desintegración de mesones π y K en el *HCAL*.

662 Cuando se estudian desintegraciones de la producción ttH , se atienden a los casos en los que
 663 se ha producido un leptón *prompt*, dos leptones *prompt* o ninguno. El fondo correspondiente
 664 a cada uno de estos canales de estudio no tienen por que tener la misma fuente.

665 Debido a que un buen análisis de los sucesos ttH depende en gran medida del estudio sobre
 666 los leptones finales, la identificación de los leptones *prompt* frente a los *fake* o *non-prompt*
 667 adquiere importancia en el análisis. Métodos computacionales de análisis multivariable son
 668 empleados en este ámbito para realizar la selección.

669 5.5. Sección eficaz de los procesos.

670 Una de las dificultades en el estudio de la producción ttH está relacionada con el bajo
 671 número de sucesos en los que se puede observar. El bosón de Higgs ya es una partícula difícil
 672 de observar en el CMS de por sí; pero como se puede observar en la Figura 5.9, los sucesos
 673 en los que se observa la producción asociada ttH suponen aproximadamente una centésima
 674 parte de los sucesos en los que se produce un H .

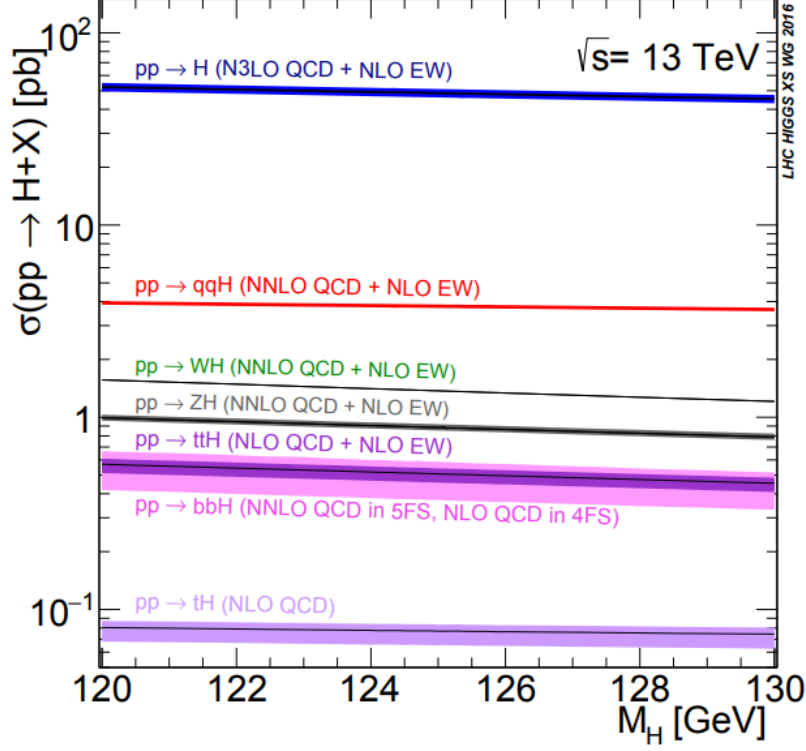


Figura 5.9: Sección eficaz del bosón de Higgs en energía $\sqrt{s} = 13$ TeV, en función de la masa del bosón. Tomado de [52].

675 Este tipo de sucesos poco frecuentes tienen un fondo en su análisis cuya sección eficaz es
676 comparable a la de la señal, poniendo de manifiesto una vez más la importancia en la sepa-
677 ración de estas dos clases de observaciones. En la figura 5.10 se ilustra esta afirmación en el
678 estudio del número de sucesos esperados en una desintegración que produce dos leptones con
679 el mismo signo en su carga eléctrica. Se pueden observar en ella un gran número de sucesos
680 $t\bar{t}$, así como una gran cantidad de observaciones en las que no son *prompt* ambos leptones
681 (se proucen leptones *fake*). Por otra parte, los sucesos en los que se espera que se observe la
682 producción ttH resultan bastante menos frecuentes.

683 En particular, el estudio sobre la producción $t\bar{t}$ para la reducción de sucesos de fondo, tam-
684 poco resulta trivial. Los sucesos de fondo en este análisis se pueden observar en las Figuras
685 5.11 y 5.12, que resultan análogas a la mencionada anteriormente. En particular son de es-
686 pecial relevancia los sucesos calificados como “Non W/Z” porque suponen el fondo *fake* en la
687 identificación de leptones para desintegraciones en este canal.

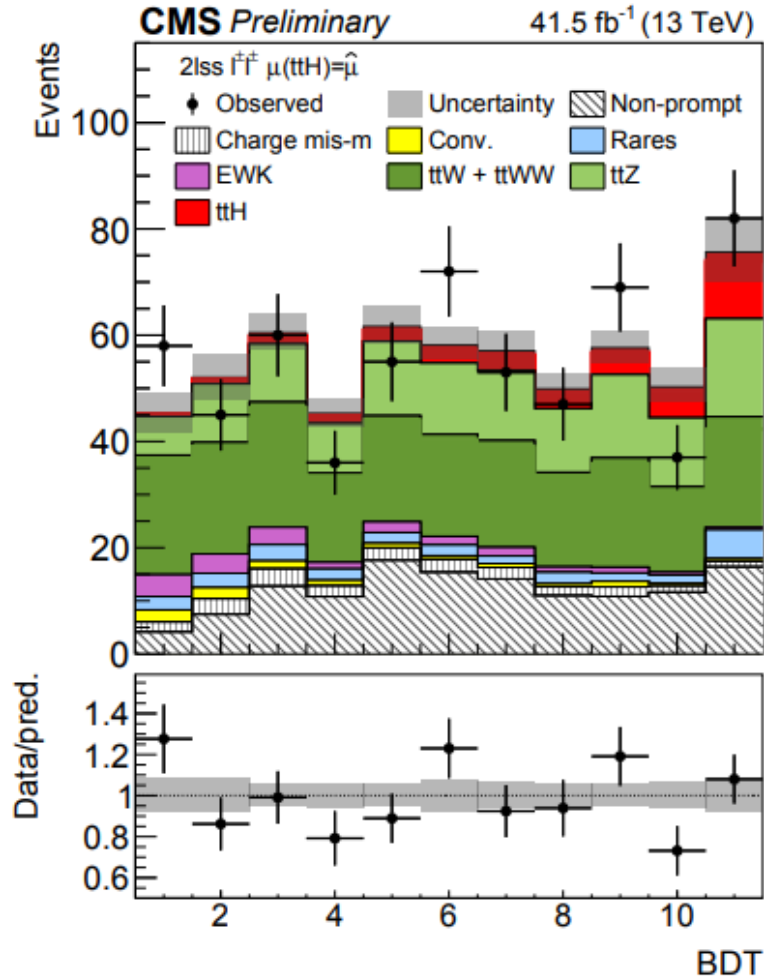


Figura 5.10: Cantidad de sucesos que se espera que produzcan dos leptones con el mismo signo en su carga eléctrica, clasificados según el proceso que los produce. En el eje X se representan distintos tipos de clasificaciones según los productos del suceso. El color rojo representa los sucesos en los que se observa producción ttH ; y el color blanco rayado los sucesos que producen dos leptones pero alguno no es *prompt*. También se incluyen observaciones reales representadas mediante cruces, y en las figuras inferiores se muestra un cociente entre el número de sucesos esperados y el número de sucesos observados. Tomado de [53].

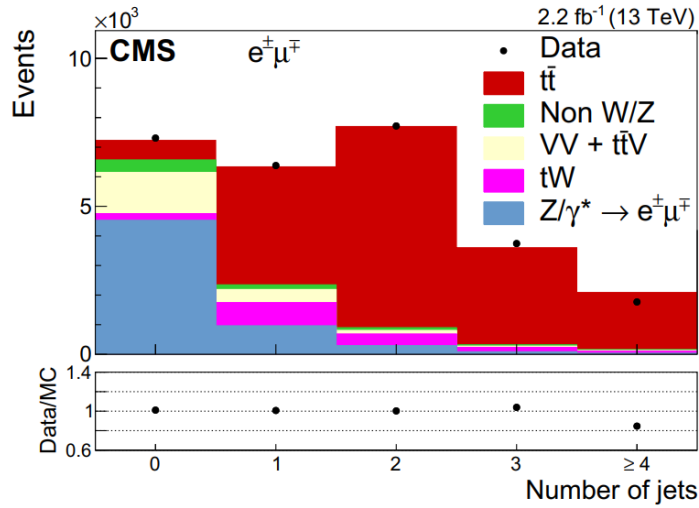


Figura 5.11: Multiplicidad de los jets producidos en sucesos que producen un leptón *prompt*. Los histogramas coloreados suponen el número de sucesos esperados para cada valor, donde la señal de sucesos $t\bar{t}$ se representa de color rojo. Las observaciones realizadas en el estudio se representan por puntos y la figura inferior representa el cociente entre el número de sucesos observados y el número esperado. Tomado de [54].

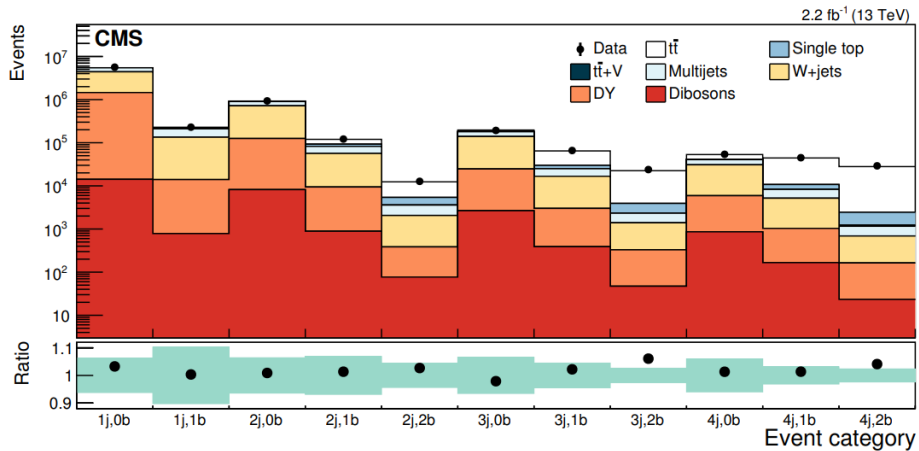


Figura 5.12: Multiplicidad de jets y b -jets producidos en los sucesos que producen dos leptones *prompt*. Los histogramas coloreados suponen el número de sucesos esperados para cada valor, donde la señal de sucesos $t\bar{t}$ se representa de color blanco. Las observaciones realizadas en el estudio se representan por puntos y la figura inferior representa el cociente entre el número de sucesos observados y el número esperado. Tomado de [55].

- 705 1. *Leptón Prompt*. Se trata de un electrón o muon producido por la desintegración de un
706 bosón W (producido por el par $t\bar{t}$) de manera directa.
- 707 2. *Tau Prompt*. El bosón W puede haberse desintegrado también en un leptón τ , por lo
708 que este canal de desintegración forma parte de la señal de producción. En este caso,
709 el leptón se detectará a través de un producto leptónico de su desintegración.
- 710 3. *Light Fake*. Leptón originado en un jet hadrónico ligero². Es fondo del análisis de
711 leptones producidos por el bosón W .
- 712 4. *Heavy Fake*. Otra componente del fondo en la que el origen del leptón analizado es un
713 jet hadrónico producido por un quark b .

714 Las primeras características del leptón que se estudian son su *momento transverso*, p_T , y su
715 *pseudorapidez*, η , en el momento inmediato que se produce el choque (en el vértice primario).
716 Leptones que tienen una alta pseudorapidez pueden ser rechazados en el estudio al no tener
717 una reconstrucción igual de fiable que los demás, así como se puede exigir un momento
718 transverso mínimo que permita librarse de leptones que muy probablemente no provengan
719 de la señal.

720 El estudio sobre el momento transverso se puede observar en la Figura 5.14; y en él, una
721 ligera tendencia de objetos categorizados como *heavy fake* a tomar momentos menores. La
722 pseudorapidez de los leptones se ve descrita en la Figura 5.15, donde se puede apreciar una
723 menor transversalidad en el caso de sucesos que producen un *light fake*. Se espera que esta
724 diferencia ayude a la red en la clasificación de este tipo de productos.

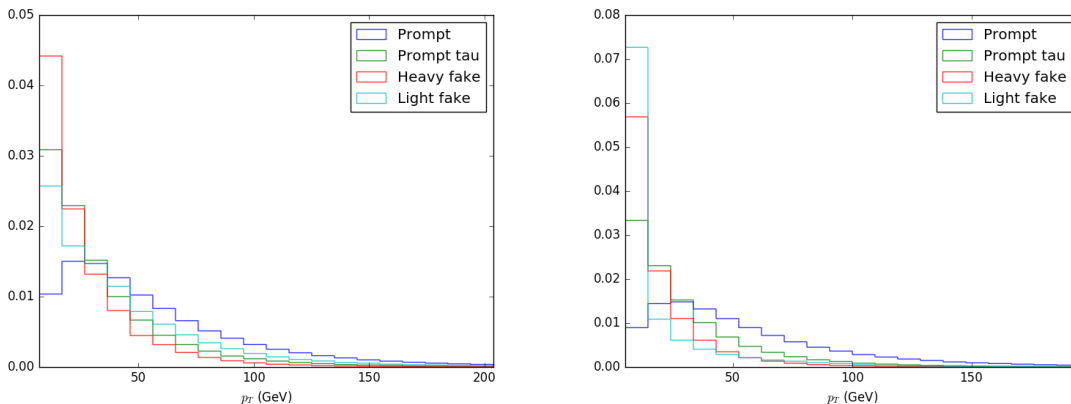


Figura 5.14: Momento transverso de la muestra (red especializada en electrones a la izquierda, y en muones a la derecha). Se representan en distintos colores las distintas categorías de los sucesos.

725 Con el objetivo de clasificar a los leptones producidos en el vértice primario, separándolos
726 de los producidos debido a la hadronización de los quarks debida a la interacción fuerte, se
727 estudia el *aislamiento* del objeto en cuestión.

²Ligero en este contexto indica que no es producido por la desintegración de un quark b .

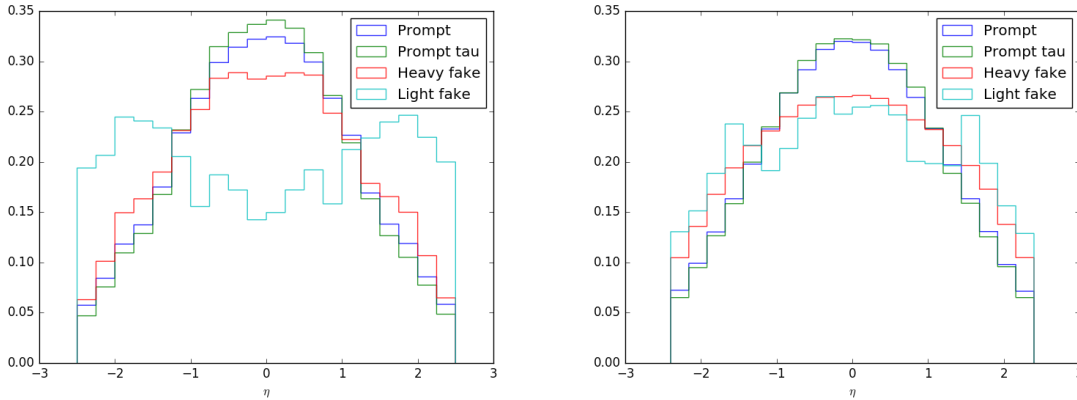


Figura 5.15: Pseudorapidez de la muestra (red especializada en electrones a la izquierda, y en muones a la derecha). Se representan en distintos colores las distintas categorías de los sucesos. Se observa un comportamiento claramente distinto en objetos categorizados como *light fake*.

728 Para poder estudiar esta característica, es necesario centrarse en la dirección del objeto de
729 estudio (fijado por (η, ϕ) de ahora en adelante; ver sección 3.2.1). La distancia respecto de
730 otro objeto distinto (de dirección (η_i, ϕ_i)) se puede definir por $\Delta R \equiv \sqrt{(\eta - \eta_i)^2 + (\phi - \phi_i)^2}$.
731 Así, si se consideran las partículas alrededor de la estudiada cuya distancia sea mayor que
732 una cierta constante $((\Delta R)_{max})$, se está teniendo en cuenta un cono cuyo eje coincide con el
733 del leptón del análisis.

734 Hay varias características relacionadas con el estudio del cono alrededor del leptón que se
735 analizan en el análisis *Lepton MVA*. Por ejemplo, el *número de partículas cargadas* en el
736 mismo cono es algo a tener en cuenta ya que se espera que sea un número reducido para los
737 leptones más aislados. Un estudio sobre este número (Figura 5.17) muestra, en el caso de los
738 sucesos donde podría haberse producido un muon, como los sucesos con un mayor número
739 de trazas posiblemente no describan la producción de un *light jet*.

740 El valor del aislamiento de un leptón se puede definir como sigue:

$$I \equiv \sum_{i: \Delta R < (\Delta R)_{max}} p_T(i),$$

741 donde $p_T(i)$ define el momento transversal correspondiente a cada una de las partículas conte-
742 nidas en el cono $\Delta R < (\Delta R)_{max}$, exceptuando el objeto estudiado. Se puede observar como
743 es de esperarse valores menores para objetos de estudio aislados.

744 Dado que I depende del momento transversal del leptón analizado, es interesante un estudio
745 relativo que no dependa del mismo mediante la variable *aislamiento relativo*, $I_{rel} = \frac{I}{p_T}$. Esta
746 variable tomará valores próximos a 0 en leptones más aislados, y valores mayores a medida
747 que el cono admite más partículas.

748 Sin embargo, esta característica no es óptima para cantidades de $(\Delta R)_{max}$ arbitrarias en los

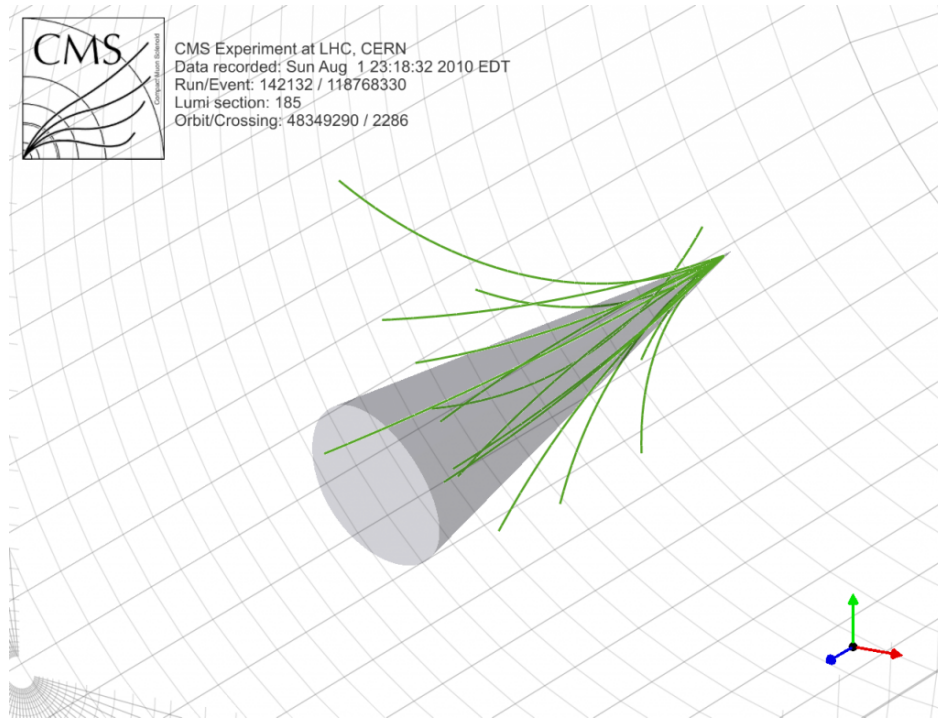


Figura 5.16: Jet de partículas originado en el vértice primario (líneas verdes) sobre el que se considera un cono centrado en una de las trazas. Tomado de [57].

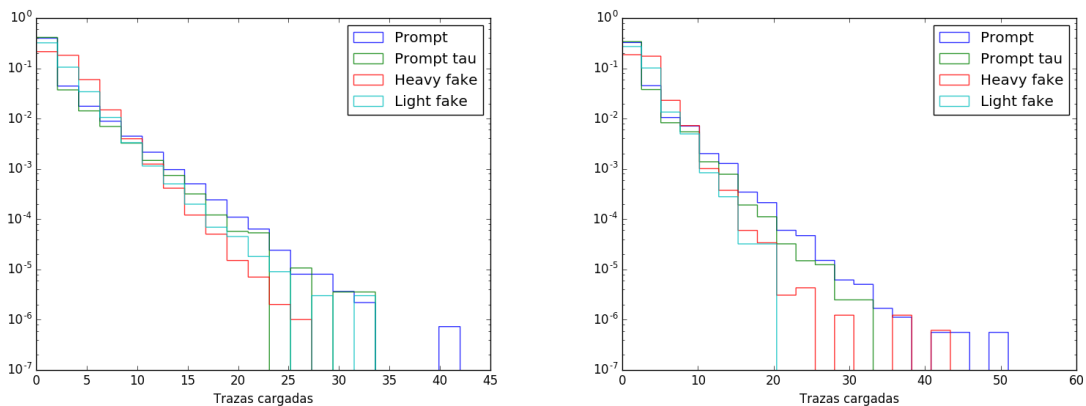


Figura 5.17: Número de trazas cargadas en el jet (red especializada en electrones a la izquierda, y en muones a la derecha). Se representan en distintos colores las distintas categorías de los sucesos.

749 casos en que el suceso produce partículas con un gran *boost* de Lorentz, como es el caso de
 750 las desintegraciones de quarks t en colisiones de energía el orden de varios TeV [58]. En este
 751 caso se puede compensar considerando tamaños del cono inversamente proporcionales al p_T
 752 del objeto, $(\Delta R)_{max} \propto \frac{1}{p_T}$. Esta definición modificada se conoce como *mini-aislamiento* del
 753 objeto.

754 En particular, se pueden considerar de manera independiente los valores de mini-aislamiento
755 del leptón respecto de partículas cargadas en el jet y respecto de partículas de carga eléctrica
756 neutra. En la muestra de la que se dispone, se disponen estos valores en la Figura 5.18; la
757 cual es especialmente interesante porque muestra comportamientos esencialmente distintos
758 entre sucesos de señal y de fondo. En ambos casos se observa una mayor cantidad de sucesos
759 de fondo para mayores valores de la variable (sucesos menos aislados), lo cual es de esperar
760 debido a que los leptones *fake* son producidos dentro de jets y el algoritmo tiene sensibilidad
761 a otros componentes del mismo jet a través de esta característica.

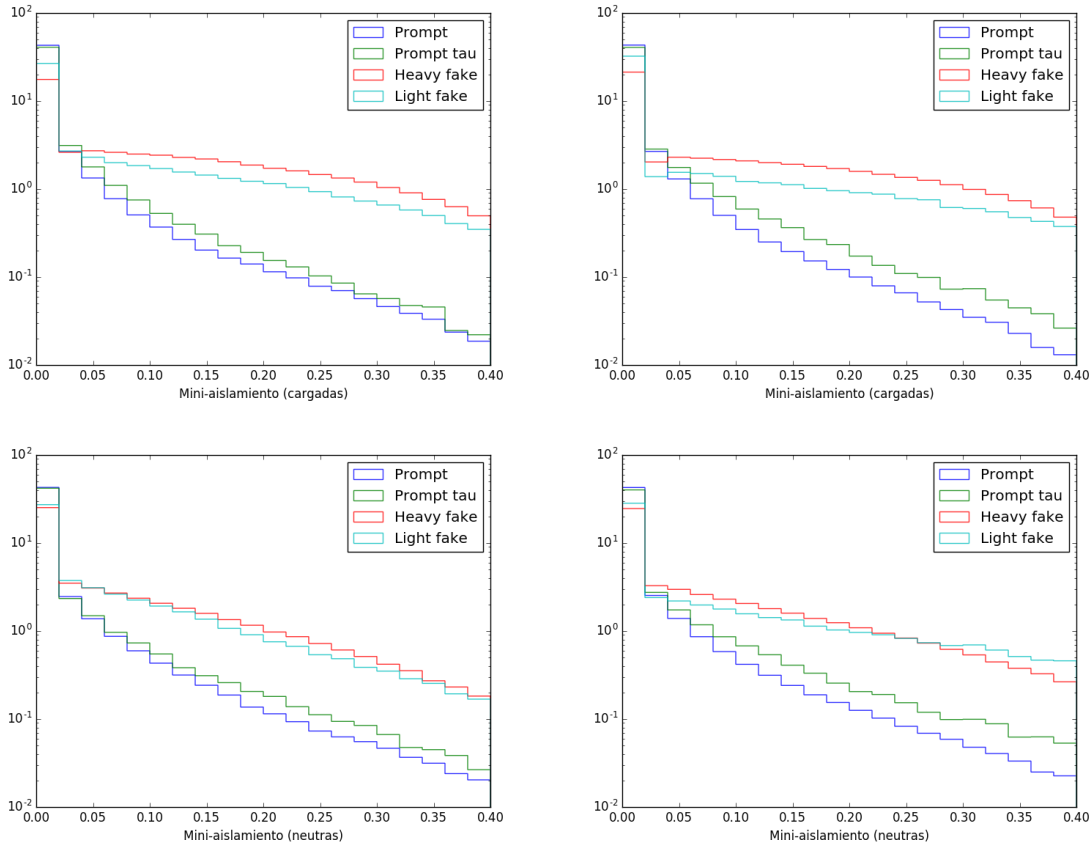


Figura 5.18: Mini-aislamiento del objeto estudiado (red especializada en electrones a la izquierda, y en muones a la derecha). Las gráficas superiores se corresponden con el mini-aislamiento respecto a partículas cargadas del jet, y las inferiores el respectivo a las partículas neutras. Se observa un distinto comportamiento entre sucesos de señal y de fondo. Se representan en distintos colores las distintas categorías de los sucesos.

762 Atendiendo también al análisis en particular de leptones con objetivo de identificar produc-
763 ciones ttH , se pueden construir variables que relacionan el objeto analizado con el jet más
764 cercano que se identifique como aquel que se forma debido a la desintegración de quarks b pre-
765 sentes en estos sucesos. Si denotamos por $p_T(jet)$ al momento transversal del jet en cuestión
766 (a menudo referido como b -jet), son ejemplos de estas variables las siguientes:

$$p_T^{ratio} \equiv \frac{p_T}{p_T(jet)}, \quad p_T^{rel} \equiv \frac{(\vec{p}_T(jet) - \vec{p}_T) \cdot \vec{p}_T}{\|\vec{p}_T(jet) - \vec{p}_T\|}.$$

767 La variable p_T^{ratio} evaluada sobre la muestra de la que se dispone (Figura 5.19 muestra un pico
 768 en valores próximos a 1 por sucesos de señal (este valor representa que el momento transversal
 769 del objeto estudiado tiene un valor muy similar al momento transversal del jet). La distinción
 770 entre señal y fondo sigue siendo clara en el estudio de p_T^{rel} (Figura 5.20).

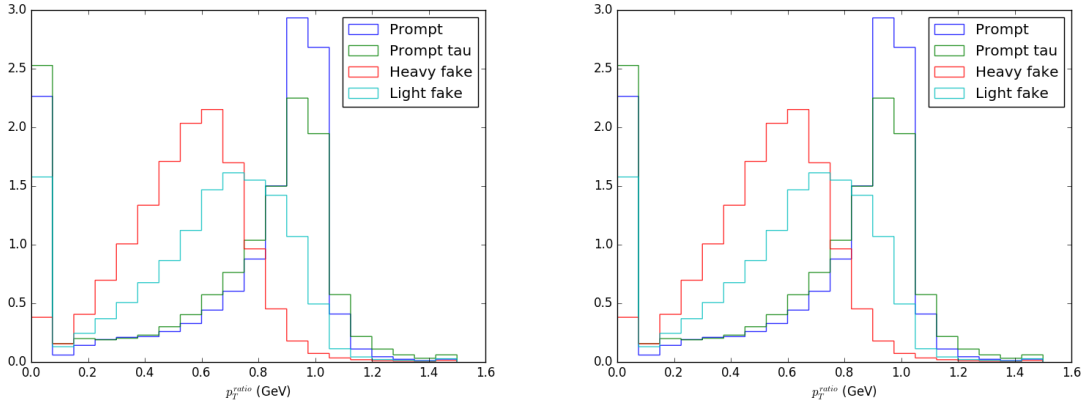


Figura 5.19: Cociente del momento transversal del objeto y el del jet (p_T^{ratio}) (red especializada en electrones a la izquierda, y en muones a la derecha). Se observa un pico alrededor del valor 1.0 en sucesos de señal. Se representan en distintos colores las distintas categorías de los sucesos.

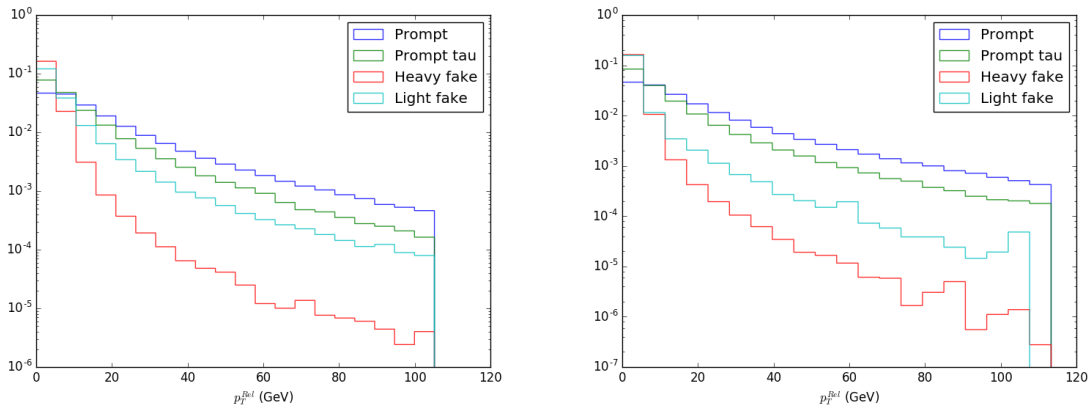


Figura 5.20: Momento transversal relativo entre el objeto y el jet (p_T^{rel}) (red especializada en electrones a la izquierda, y en muones a la derecha). Se representan en distintos colores las distintas categorías de los sucesos.

771 Por otra parte, para la identificación del origen de un jet con el de una desintegración de un
 772 quark b , es necesario el empleo de *algoritmos de b-tagging*. El valor del discriminante en cues-
 773 tión también puede ser de interés en el análisis multivariable. Más aún, en la desintegración

774 del quark interviene un bosón W que puede desintegrarse en leptones o en hadrones. Resulta
775 de interés la consideración tanto del valor del discriminante enfocado a desintegraciones que
776 dan lugar a leptones (*target leptónico*) como al enfocado a las que dan lugar a hadrones (*target*
777 *hadrónico*).

778 En la Figura 5.21 se representan algoritmos de *b-tagging* aplicados a los objetos estudiados.
779 En particular, el algoritmo de *target hadrónico* toma valores próximos a 1 en leptones con
780 una mayor probabilidad de que provengan de la desintegración en hadrones de un quark b ; así
781 como el de *target leptónico* toma valores próximos a 1 en leptones con una mayor probabilidad
782 de provengan de la desintegración en leptones de un quark b . En este último, dado que está
783 enfocado a una desintegración del quark b en su modo leptónico, si no se detecta la presencia
784 de ningún leptón se asigna el valor -100 a la observación. En ambas variables se puede
785 observar una mayor proporción de sucesos de fondo para valores próximos a 1 (los leptones
786 *prompt* provienen de la desintegración del bosón W , no de un quark b , y deberían tener un
787 valor menor en el discriminante).

788 También es de importancia la consideración del *parámetro de impacto* del objeto de estudio.
789 Este se define como la distancia a la que se produce el leptón respecto del vértice primario
790 de la colisión. Se espera que en leptones *prompt* la distancia sea próxima a 0 debido a que se
791 producen en el mismo vértice, de la misma manera que un leptón originado en un b -jet tendrá
792 valores de parámetro de impacto mayores (se hadroniza antes de desintegrarse, al contrario
793 que los quarks t , y tienen una vida media mayor en la que se produce un cierto desplaza-
794 miento apreciable). En particular, se consideran en el análisis la componente longitudinal del
795 parámetro, d_z , de manera independiente a la transversa, d_{xy} .

796 En la Figura 5.23 se observa como los electrones y muones *prompt* tienen una tendencia
797 a parámetros de impacto menores y, como era de esperar, en los productos *heavy fake* se
798 observa un parámetro de impacto mayor. No hay que olvidar que estos últimos nacen de
799 la desintegración de un quark b , lo cual da lugar usualmente a la creación de un vértice
800 secundario en el suceso (sección 4.5).

801 La variable que relaciona el parámetro de impacto con la incertidumbre de esta también se
802 tiene en cuenta en análisis de tipo MVA, y se define como $SIP \equiv \frac{d}{\Delta d}$ (donde SIP denota
803 *Significance of Impact Parameter*, y Δd denota la incertidumbre sobre la medida del pará-
804 metro de impacto d). Una mayor significancia supone una mayor precisión en la medida de
805 los parámetros d_{xy} , d_z ; lo cual resulta más preciso en sucesos que generan *heavy fake* (Figura
806 5.24).

807 Hasta el momento las variables descritas para el análisis eran independientes de si el leptón
808 estudiado es un electrón o un muon. Sin embargo, el MVA puede especializarse a uno de estos
809 dos casos, y considerar variables específicas de ellos que no compartan.

810 Esto es el caso del valor del discriminante *muon segment compatibility*, que expresa la com-
811 patibilidad entre la traza del muon en el detector de trazas con la reconstruida en la cámara
812 de muones, basado en un clasificador probabilístico llamado “clasificador bayesiano ingenuo”.
813 De la misma manera, se consiera el valor de otro clasificador probabilístico para electrones
814 basado en un “bosque aleatorio”, que a su vez es otro análisis multivariable independiente.

815 El valor del clasificador de electrones sobre la muestra se ve representado en la Figura 5.25,

816 donde valores próximos a 1 denotan mayor seguridad de que el objeto sea un electrón, y
817 valores próximos a -1 mayor seguridad en lo contrario. Se observa así una mayor cantidad
818 de sucesos de señal en el valor 1, y una gran proporción de sucesos *light fake* con valores
819 próximos a -1 .

820 Por otra parte, los valores del clasificador que analiza la compatibilidad de las trazas en
821 la cámara de muones se representan en la Figura 5.26. Mayores valores de este algoritmo
822 denotan una mayor probabilidad de que la traza del muon está correctamente construida. Se
823 vuelve a notar en este caso como los sucesos *light fake* contribuyen en gran medida a valores
824 menores de la variable.

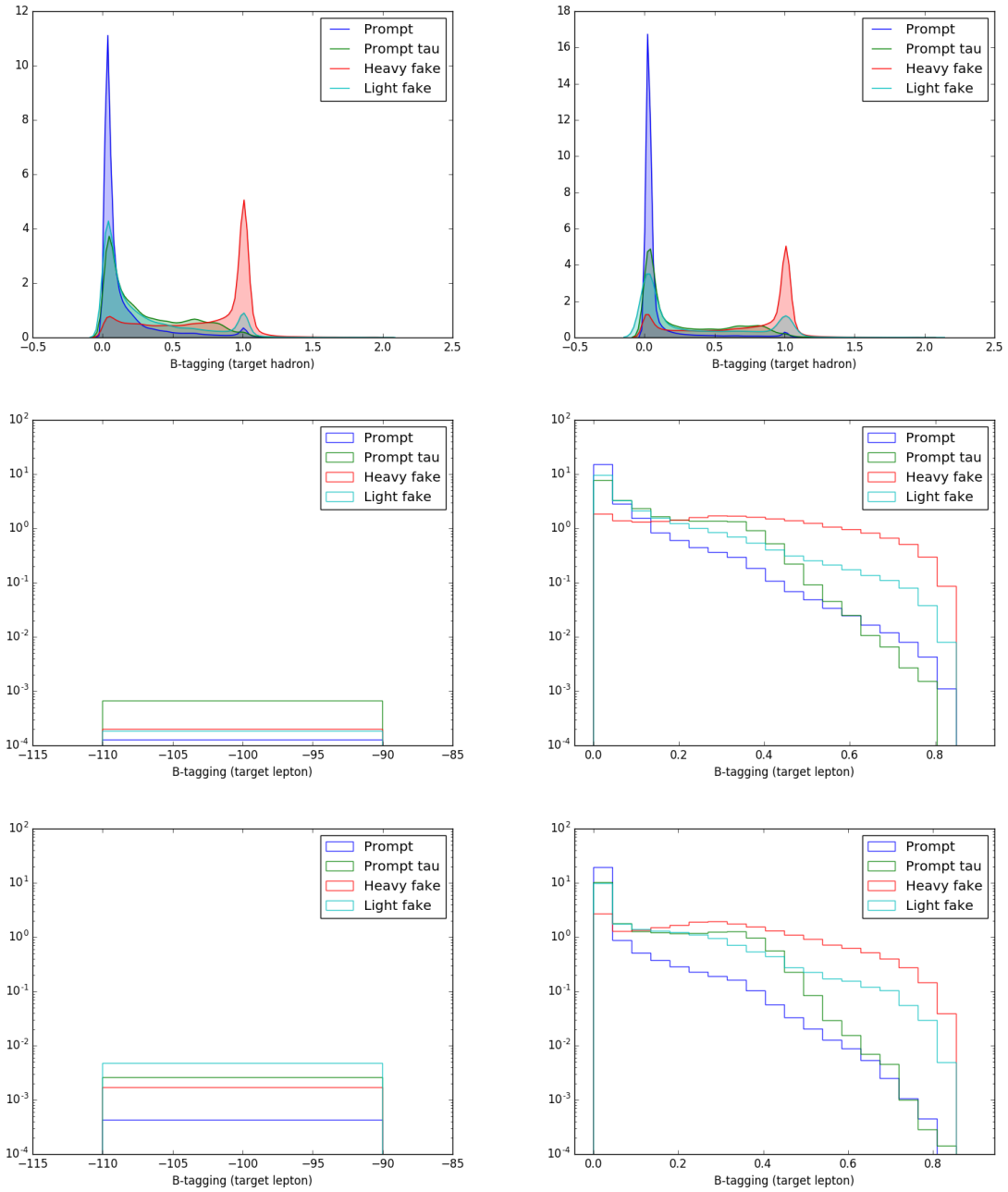


Figura 5.21: Valor de los algoritmos de B-tagging. Se representa en la parte superior la densidad de valores del algoritmo de target leptónico (red especializada en electrones a la izquierda, y en muones a la derecha); y en la parte central (red para electrones) e inferior (red para muones) el histograma de valores del algoritmo de target hadrónico. Este último se divide en dos partes, para representar la cantidad de sucesos clasificados con un valor -100 correspondiente con la falta de leptones detectados en el jet. Se representan en distintos colores las distintas categorías de los sucesos.

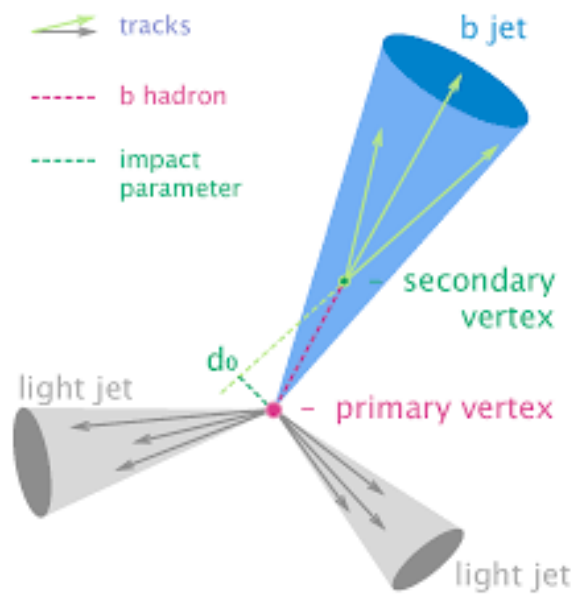


Figura 5.22: Representación del parámetro de impacto (d_0 en la imagen) de una partícula originada en un b -jet. En la misma se observa como los vértices originados en el vértice primario tienen un parámetro de impacto teórico nulo. Tomado de [59].

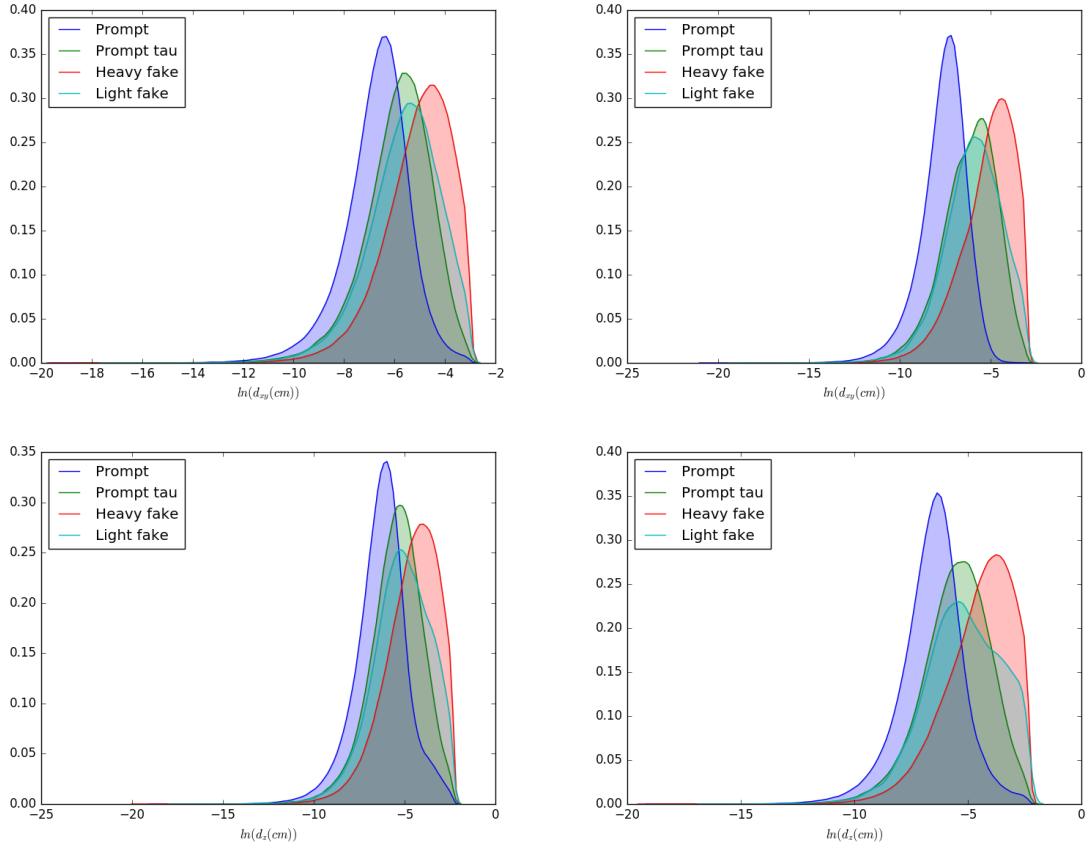


Figura 5.23: Densidad de los valores de los parámetros de impacto observados para la partícula (red especializada en electrones a la izquierda, y en muones a la derecha). Se representa el parámetro de impacto transverso d_{xy} en las gráficas superiores, y el longitudinal d_z en las inferiores.

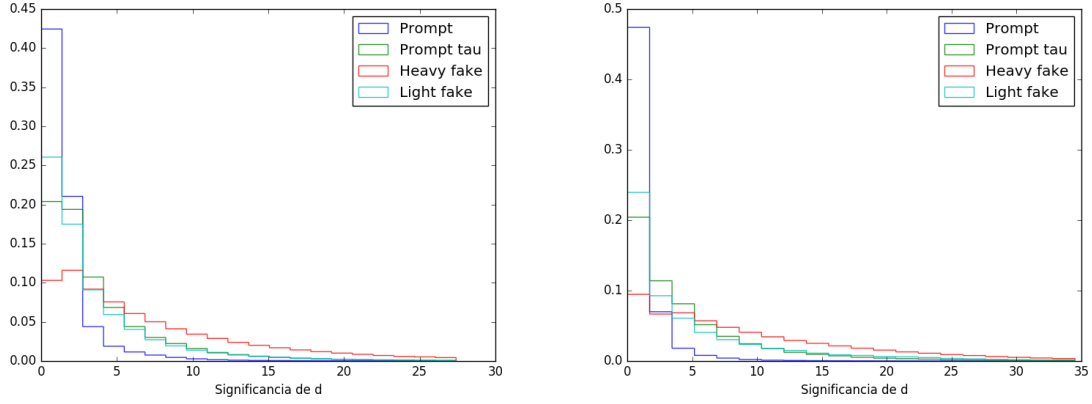


Figura 5.24: Valor de la significancia en el parámetro de impacto, SIP (red especializada en electrones a la izquierda, y en muones a la derecha). La relación entre d , d_{xy} y d_z es $d = \sqrt{d_{xy}^2 + d_z^2}$.

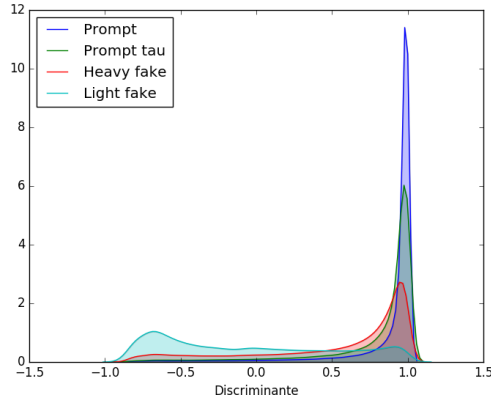


Figura 5.25: Densidad de valores del discriminante externo al algoritmo que tiene en cuenta la red especializada en electrones. Basado en una técnica de bosque aleatorio.

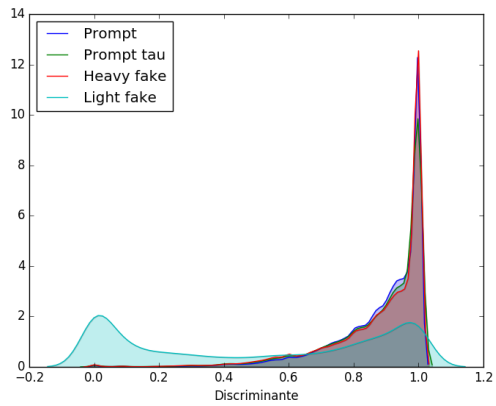


Figura 5.26: Densidad de valores del discriminante externo al algoritmo que tiene en cuenta la red especializada en muones. Basado en un clasificador bayesiano ingenuo.

825 Capítulo 6

826 El nuevo algoritmo de identificación

827 El empleo de algoritmos computacionales en el análisis de sucesos resulta natural para el
828 manejo de todos los datos que proporciona el CMS, como se introdujo en capítulos anteriores.

829 En particular, los algoritmos para el análisis multidimensional *Leptón MVA* emplean una
830 técnica conocida como *aprendizaje automático* o *machine learning*, que se introduce en este
831 capítulo.

832 Será propuesto un algoritmo alternativo también basado en *machine learning*, cuyo objetivo
833 es la identificación de electrones y muones *prompt* productos de la producción de tipo $t\bar{t}$ en
834 el CMS, con una técnica distinta al tradicional *árbol de decisión dopado* (*Boosted Decision*
835 *Tree, BDT*).

836 6.1. Introducción al *Machine Learning*

837 El aprendizaje automático, o *Machine Learning* (ML) es un método de análisis computacio-
838 nal de datos mediante sistemas que aprenden y evolucionan a partir de los datos que son
839 introducidos.

840 En este contexto, “aprender” significa ser capaz de detectar patrones complejos en una gran
841 cantidad de datos y dimensiones, y modificar la salida del algoritmo en función de ellos. El
842 adjetivo “automático” denota aquí la capacidad de los sistemas de mejorar sin intervención
843 humana con el tiempo.

844 Un problema típico en el área de ML se puede describir como la búsqueda de una determinada
845 función $f : X \rightarrow Y$ (la *función objetivo*), que toma como argumentos los datos observados
846 contenidos en X , y cuya salida es un número o etiqueta contenido en el espacio Y . De entre
847 todas las funciones de esta forma, se trata de encontrar aquella que optimiza una determinada
848 métrica asociada al problema, $L(y, f(x))$ (donde $y \in Y$), denominada *función pérdida* (*loss*
849 *function*).

850 Un algoritmo ideal encontraría la función objetivo f que optimiza L para todo valor (x, y) . A
851 medida que se aumenta la dimensión de X (número de variables observadas en cada uno de

852 los múltiples datos)¹, esta tarea se vuelve más cara computacionalmente. En el *aprendizaje*
853 *supervisado* (*supervised learning*), se dispone de cierta muestra $\{x_i, y_i\}_{i=1}^N$ ya evaluada o
854 clasificada, a la que se denomina *muestra de entrenamiento* (y al proceso que optimiza la
855 función objetivo mediante esta muestra se le conoce por *entrenamiento*). Más aún, es frecuente
856 que la función f pertenezca a una familia de funciones $f_\phi(x)$ parametrizada por una serie de
857 parámetros ϕ , de forma que el espacio de búsqueda de la función objetivo se reduzca.

858 La función obtenida por aprendizaje supervisado no sólo debe minimizar el valor de L en la
859 muestra de entrenamiento: debe ser generalizable a observaciones posteriores. Dado que esta
860 función supone un modelo físico, es imprescindible que su capacidad de predicción no se vea
861 reducida a la muestra ya clasificada previamente.

862 Cuando el modelo obtenido por ML se ajusta muy bien a la muestra de entrenamiento, pero
863 falla en otros datos, se dice que ha sido *sobreentrenado* (se ha producido *overfitting*). Las
864 técnicas que evitan este fenómeno se conocen por técnicas de *regularización*.

865 Para ilustrar este efecto, se propone la Figura 6.1 como ejemplo, donde un modelo físico,
866 en principio desconocido, se representa mediante una línea negra separando sucesos azules y
867 rojos. Para la aproximación de dicha línea, se realizan una serie de medidas, representadas con
868 puntos, etiquetadas con colores previamente a su introducción en el algoritmo como muestra
869 de prueba. En la figura se representa con una línea verde un modelo en el que se separan
870 ambas categorías, sin ningún tipo de error, pero que aproxima pobremente el modelo real (en
871 futuras mediciones, la capacidad predictiva del modelo será muy poco eficiente).

872 El fenómeno descrito en la figura es el de *overfitting* y ocurre cuando el algoritmo sigue
873 de manera demasiado estrecha las fluctuaciones estadísticas de las mediciones. El fenómeno
874 opuesto se conoce por *underfitting*, y se produce cuando un modelo trata de aproximar un
875 modelo físico complejo con una función muy poco adecuada (por simpleza de la función, o
876 por un incorrecto entrenamiento). Siguiendo el ejemplo, se produciría *underfitting* al tratar
877 de aproximar la línea negra mediante una línea recta.

878 Es frecuente, en la evaluación de los algoritmos entrenados, el empleo de una segunda muestra
879 previamente evaluada o etiquetada (la *muestra de prueba*). El objetivo es comparar el valor
880 obtenido por el modelo con el proporcionado por la muestra, para determinar como de bien
881 se ajusta f a datos fuera de la muestra de entrenamiento.

882 En esta sección se van a describir algunos tipos de algoritmos de aprendizaje supervisado
883 importantes empleados en la técnica de análisis *Lepton MVA*.

884 6.1.1. Arbol de decisión

885 Los métodos de aprendizaje supervisado basados en árboles de decisión (*tree-based methods*)
886 tienen por principio la partición del espacio de variables introducidas en el algoritmo en cajas²

¹La maldición de la dimensión (*curse of dimensionality*) es el nombre que reciben una serie de fenómenos en análisis de múltiples dimensiones. Cuando las dimensiones de una variable estudiada aumentan, los datos tienen una mayor dispersión, disminuyendo la significancia estadística. Para obtener un resultado confiable estadísticamente, el número de datos necesarios aumenta de manera exponencial con el número de dimensiones consideradas.

²En este contexto, “caja” se refiere a la generalización de un rectángulo a dimensiones mayores (es decir, un producto cartesiano de intervalos).

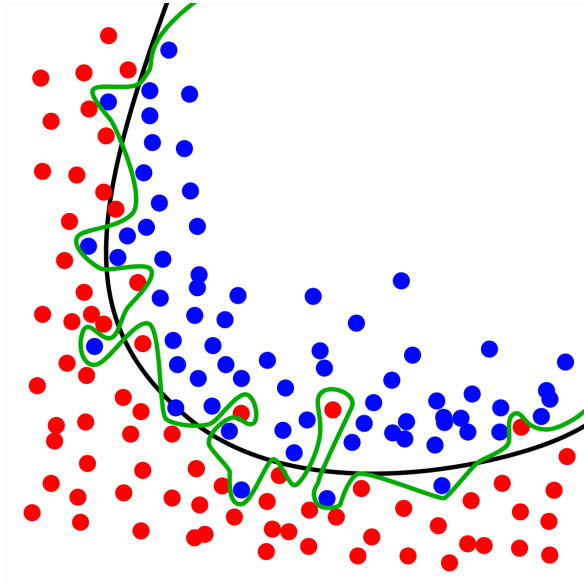


Figura 6.1: Ejemplo de modelo donde se ha producido *overfitting*. La línea negra representa un modelo físico real, que separa sucesos azules de rojos, y los puntos representan medidas experimentales de dichos sucesos. La línea verde es una mala aproximación del modelo mediante ML, en el sentido de poca capacidad predictiva. Tomado de [60].

887 (cuya dimensión se determina por el número de atributos estudiadas). Una ventaja clave de
 888 estos algoritmos nace de la facilidad de interpretarlos una vez construidos.

889 Para simplificar la descripción de estos métodos, se va a considerar que se estudian dos caracte-
 890 rísticas X_1 y X_2 de un suceso (como podrían ser el momento transverso y la pseudorapidez
 891 en la identificación de leptones). En la Figura 6.3 (esquema de la izquierda) se disponen los
 892 datos de dichas características sobre dos ejes. El algoritmo de árbol de decisión va a tener
 893 una respuesta distinta en función de los valores estudiados en cada una de estas variables,
 894 $(x_1, x_2) \in X_1 \times X_2$.

895 En el ejemplo de la figura, se estudia primeramente si la variable X_1 es mayor o menor
 896 que un cierto t_1 , produciéndose la primera división del espacio. Posteriormente, en la región
 897 $X_1 \leq t_1$ se estudia cuando X_2 es mayor o menor que t_2 , produciéndose una segunda división.
 898 La reiteración de estas divisiones produce la partición del espacio de entrada del algoritmo
 899 en cajas, denotadas en la figura por $\{R_m\}_{m \in \mathbb{N}}$. A este método particular de partición se le
 900 conoce como *CART*.

Un algoritmo con una cantidad arbitraria M de cajas va a producir una respuesta (una función objetivo f) del tipo

$$f(x_1, x_2) = \sum_{m=1}^M c_m I\{(x_1, x_2) \in R_m\}.$$

901 donde $I\{(x_1, x_2) \in R_m\}$ es una función que tiene valor 1 si $(x_1, x_2) \in R_m$, y 0 en caso
 902 contrario; y los valores c_m son constantes determinadas por el algoritmo una vez entrenado.

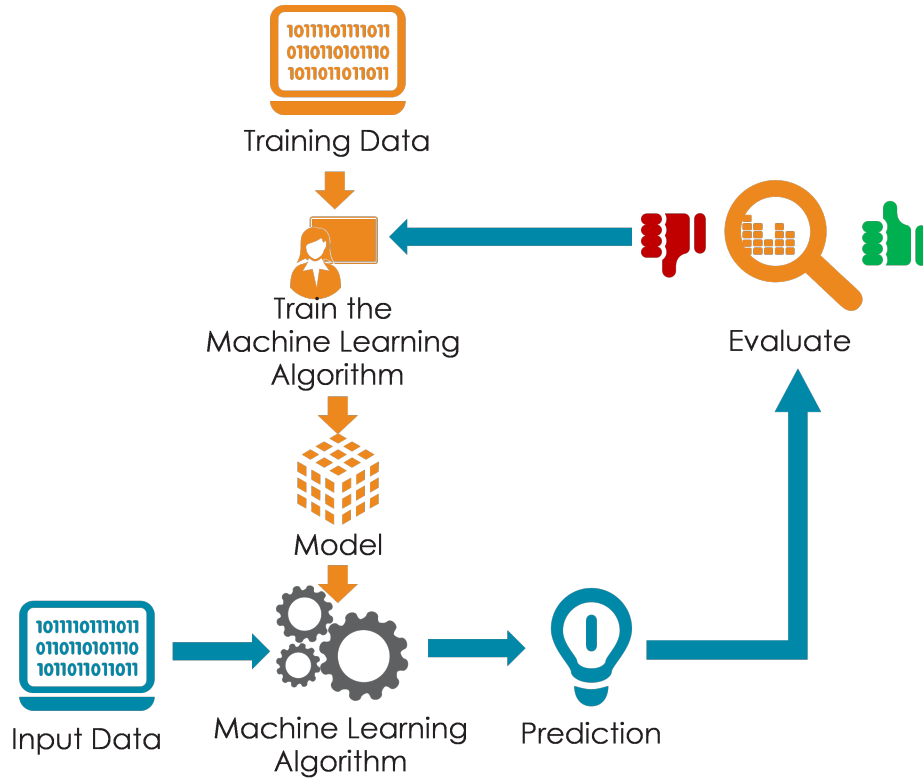


Figura 6.2: Diagrama que muestra el proceso de entrenamiento en algoritmos de ML supervisados. En la sección de “evaluación” (evaluate) se emplea la muestra de prueba. Tomado de [61].

903 Este tipo de algoritmos puede ser empleado en problemas de clasificación, donde la salida
 904 de la función objetivo tendría dimensión P (siendo P el número de categorías de entre las
 905 cuales se quiere clasificar el suceso), y el valor de dicha función en cada componente re-
 906 presentaría la probabilidad de pertenecer a la categoría correspondiente. Así, $f(x_1, x_2) =$
 907 $(f^1(x_1, x_2), \dots, f^P(x_1, x_2))$ sería de la forma

$$f^\alpha(x_1, x_2) = \sum_{m=1}^M c_m^\alpha I\{(x_1, x_2) \in R_m\}.$$

908 donde $\sum_{\alpha} f^\alpha(x_1, x_2) = 1$ para que las probabilidades estén bien definidas, y donde $f^\alpha(x_1, x_2)$
 909 representa la probabilidad del suceso a pertenecer a la categoría α .

910 La generalización a más de dos dimensiones resulta inmediata, considerando características
 911 X_1, X_2, \dots, X_N , y por tanto cajas de dimensión N . En la sección 4.3 referente al *Leptón*
 912 *MVA* se describían hasta 13 variables en el análisis de electrones o de muones.

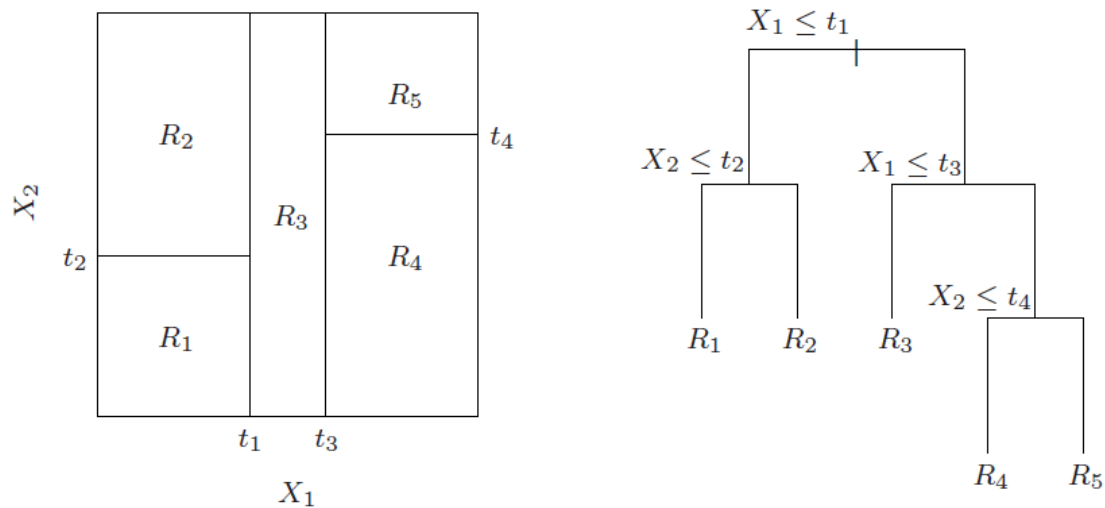


Figura 6.3: Representación de la partición del espacio de variables mediante el método *CART* (caso particular en el que $m = 1, \dots, 5$). El árbol de decisión toma valores distintos para cada región R_i de las variables de entrada. Tomado de [62].

913 6.1.2. Dopaje de algoritmos. Árbol de decisión dopado (*BDT*)

914 El dopaje de un algoritmo de aprendizaje supervisado (*boosting*) consiste en un método apli-
 915 cado a técnicas conocidas que permiten aumentar su rendimiento. La base de este proceso
 916 consiste en la evaluación de pequeñas modificaciones del algoritmo sin dopar, y la asignación
 917 de una mayor importancia a aquellas que consigan un menor error en su ejecución. El algo-
 918 ritmo dopado final consiste en la combinación de cada uno de los anteriores ponderados por
 919 su capacidad predictiva.

Para ilustrar este concepto, se va a introducir el algoritmo de dopaje más popular empleado en tareas de clasificación: *AdaBoost* [63]. Considérese por tanto un problema de clasificación en dos clases distintas, $Y = \{-1, 1\}$. Sea además un vector de medidas $X = (x_1, \dots, x_N)$ introducidas en un clasificador G (a menudo referido como *clasificador débil*) para predecir la clase de cada uno de los componentes del vector. La proporción de errores asociada a esta clasificación se puede expresar como:

$$err = \frac{1}{N} \sum_{i=1}^N I(y_i \neq G(x_i))$$

920 Considérense ahora M clasificadores débiles, G_1, \dots, G_M . Cada uno de ellos emplea la misma
 921 técnica de clasificación, pero las componentes del algoritmo que los definen son distintas.
 922 De esta manera, cada G_i pueden asignar una predicción distinta para el mismo vector X .
 923 Estudiando el error en cada uno de ellos, se puede asignar un valor α_i que toma un valor
 924 mayor en los clasificadores débiles con menor proporción de error.

925 Un dopaje de la técnica de predicción débil actúa como un “voto ponderado” de cada una de

926 las repeticiones de la misma, y clasifica cada medida empleando la mayoría de cada votación.
 927 Es decir, dado que $Y = \{-1, 1\}$, se puede definir un dopaje G de las técnicas G_1, \dots, G_M de
 928 la forma

$$G(x) = \text{sign} \left(\sum_{m=1}^M \alpha_m G_m(x) \right)$$

929 donde $x \in X$, $\text{sign}(x) = \frac{x}{|x|}$.

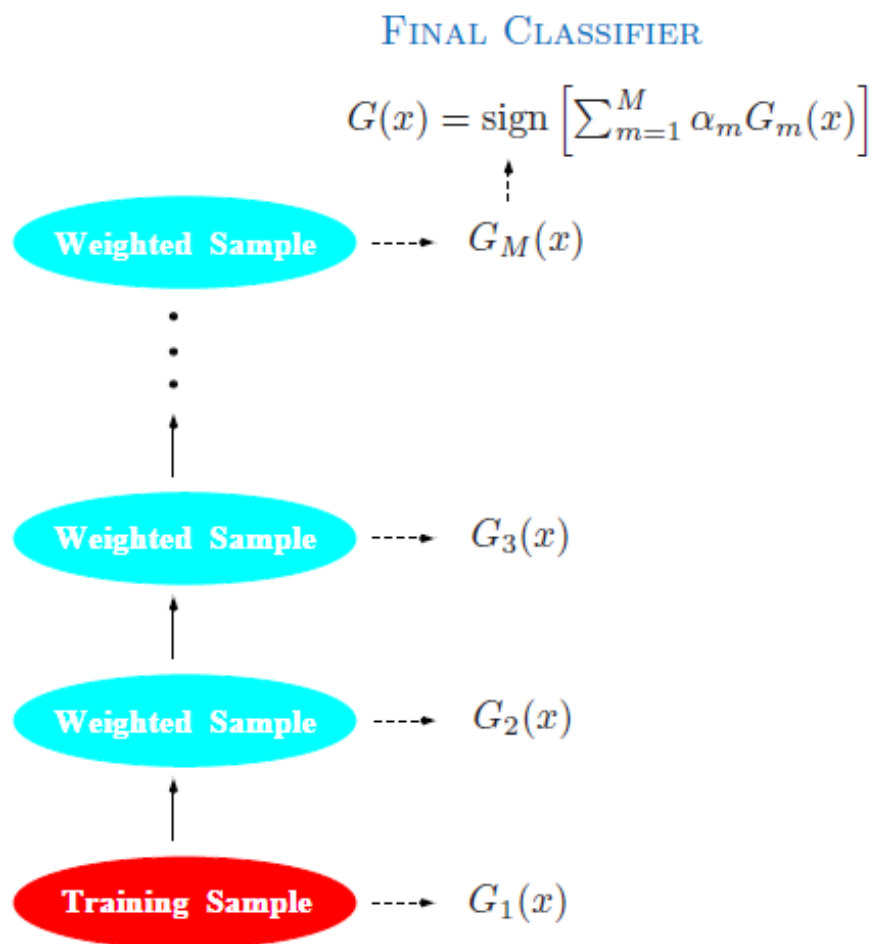


Figura 6.4: Representación del dopaje de un algoritmo mediante AdaBoost. En particular, en esta representación cada modificación del algoritmo G_i tiene en cuenta G_{i-1} . Tomado de [62].

930 Un empleo en particular de las técnicas de dopaje ve su utilidad en la clasificación *Leptón*
 931 *MVA*, al ser aplicadas a árboles de decisión. Al algoritmo se le conoce como *Árbol de decisión*
 932 *dopado* (*Boosted Decision Tree, BDT*), y también es aplicado en otras clasificaciones del
 933 experimento *CMS*.

934 **6.1.3. Red neuronal profunda (*DNN*)**

935 Las redes neuronales son un tipo de método de aprendizaje desarrollado en ámbitos de la
 936 estadística y la inteligencia artificial. Se basa en la construcción de características, que son
 937 combinaciones lineales de las variables estudiadas, para modelizar la función objetivo como
 938 una combinación no necesariamente lineal de dichas características.

939 Considérese un problema general de aprendizaje supervisado. En este, se estudian vectores
 940 X con p componentes, y la salida se representa por Y . La función objetivo modelizada por
 941 un modelo de tipo *PPR* (*Projection Pursuit Regression*) será de la forma:

$$f(X) = \sum_{i=1}^M g_m(\omega_m^T X),$$

942 donde $M \in \mathbb{N}$; y $\omega_m^T X \equiv V_m$ son las características introducidas al principio de la sección. A
 943 los vectores ω_m se les denomina *pesos* del algoritmo, y las funciones g_m son estimadas junto
 944 con los pesos en el entrenamiento de este.

945 Esta construcción resulta muy general, dado que se pueden generar una gran cantidad de
 946 modelos mediante estas combinaciones no lineales. En particular, la elección de una constante
 947 M lo suficientemente grande permite una aproximación del método lo suficientemente precisa
 948 a un modelo cualquiera [64]. Es por ello que a este tipo de construcciones se les conoce como
 949 *aproximadores universales*.

La función de pérdida que permite optimizar el modelo *PPR*, para ciertos datos de entrena-
 miento $(x_i, y_i)_{i=1}^N$, dependerá de la elección de g_m y ω_m ($m = 1, \dots, M$), y se puede expresar
 de la forma

$$\sum_{i=1}^N \left[y_i - \sum_{m=1}^M g_m(\omega_m^T x_i) \right]^2.$$

Una *red neuronal de una sola capa* puede ser vista como un tipo de modelo *PPR*, el cual es
 visto de la forma:

$$\left. \begin{aligned} Z_m &= \sigma(\alpha_{0m} + \alpha_m^T X), & m &= 1, \dots, M \\ T_k &= \beta_{0k} + \beta_k^T Z, & k &= 1, \dots, K \\ f_k(X) &= g_k(T), & k &= 1, \dots, K \end{aligned} \right\} g_m(\omega_m^T X) = \beta_m \sigma(\alpha_{0m} + \alpha_m^T X),$$

950 donde $\beta_m \sigma(\alpha_{0m} + \alpha_m^T X) = \beta_m \sigma(\alpha_{0m} + \|\alpha_m\|(\omega_m^T X))$ y donde las expresiones empleadas son:

- 951 ■ $Y = (Y_1, \dots, Y_k)$ es el argumennto de salida de dimensión K (por ejemplo, K categorías
 952 de entre las que se quiere clasificar un objeto, con un vector Y asociado que determina
 953 la probabilidad de pertenecer a cada una de ellas).
- 954 ■ $\sigma(v)$ es la función activación, elegida previamente a la construcción de la red, y usual-
 955 mente es la función sigmoide $\text{sigmoid}(v) = \frac{1}{1+e^{-v}}$ en tareas de clasificación. También es
 956 empleada la función *ReLU* (*Rectified Linear Unit*), definida por $\text{ReLU}(v) = \max(0, v)$.
- 957 ■ α_m es un vector de pesos que denota una combinación lineal de los elementos introdu-
 958 cidos en la función, X , y α_0 representa un sesgo que generaliza la combinación lineal al
 959 añadir una constante.

- 960 ▪ β_k es un vector análogo a α_m que actúa sobre Z , y de la misma manera β_0 es una
- 961 constante (el sesgo).
- 962 ▪ $Z = (Z_1, \dots, Z_M)$ son las unidades *ocultas* (llamadas así porque no son directamente
- 963 observadas en la ejecución del programa, a menudo referidas como *neuronas*). Se pueden
- 964 interpretar como una extensión de los datos introducidos X con la que se trabajará en
- 965 la red como si fuesen los datos iniciales.
- 966 ▪ $g_k(T)$ es la función de salida de la red (de la componente k), donde su argumento
- 967 $T = (T_1, \dots, T_K)$ puede interpretarse de manera análoga al vector Z como un vector de
- 968 salida previo al que se obtiene por la red al transformarse mediante g_k . En clasificación
- 969 a menudo se emplea la función *softmax*, $g_k(T) = \frac{e^{T_k}}{\sum_{l=1}^K e^{T_l}}$.

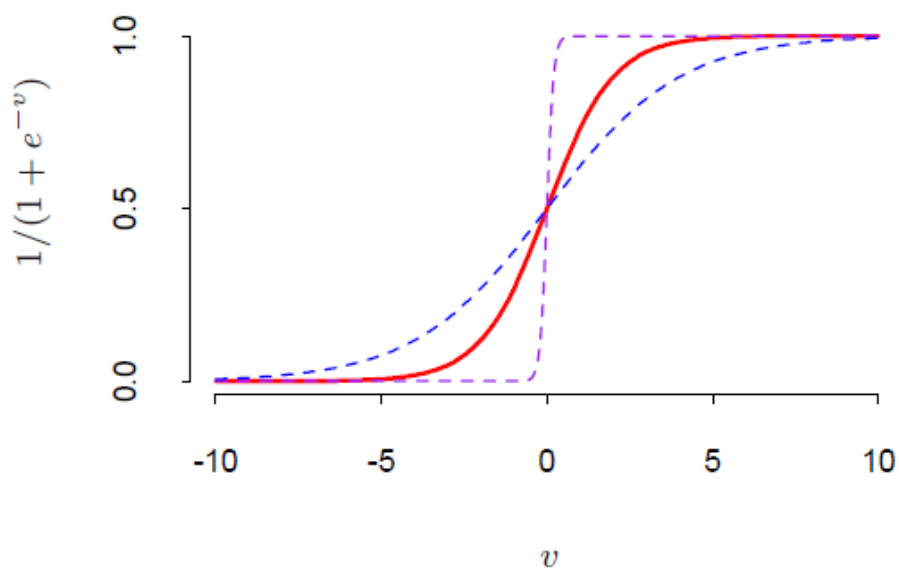


Figura 6.5: Representación de la curva sigmoide (de rojo), a menudo tomada como función de activación $\sigma(v)$. Es común también un reescalado de la misma en las redes neuronales, asociado al problema, y mediante el parámetro de escala s (en la notación empleada, $s = \|\alpha_m\|$). Las curvas discontinuas representan la función $\text{sigmoid}(sv)$ con $s = 1/2$ (azul) y con $s = 10$ (violeta). Tomado de [62].

970 La transformación no lineal σ introducida por las redes neuronales debe ser implementada
 971 para lograr aproximar modelos que no son lineales (sin ella, la función final f se podría
 972 expresar como combinación lineal de los elementos de entrada X). Por otra parte, tiene
 973 una menor complejidad que una transformación arbitraria g de un *PPR* general. Con la
 974 introducción de la función de activación en redes neuronales se logran modelos capaces de
 975 emplear una alta cantidad de características V_m sin perder la propiedad del modelo de ser
 976 un aproximador universal.

977 Una característica muy apreciada en las funciones de activación, es su diferenciabilidad. Esto
 978 se debe a que muchas técnicas que se centran en la optimización de los parámetros del

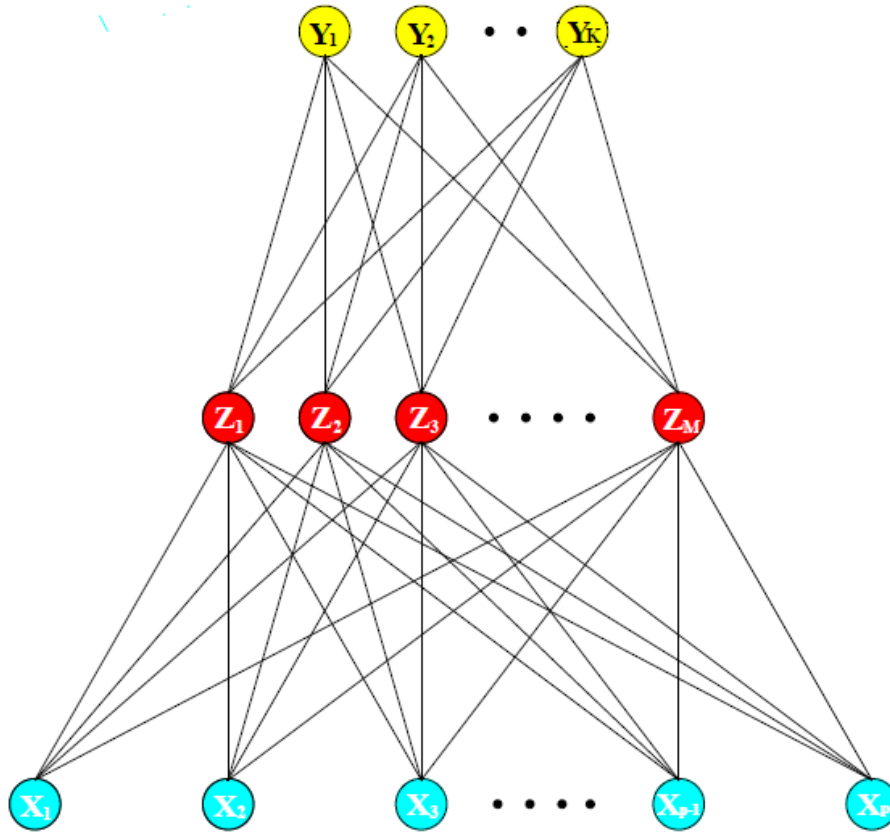


Figura 6.6: Diagrama de una red neuronal de una sola capa. Se puede observar como el modelo debe su nombre al comparar cada Z_m con las neuronas de un cerebro, y a las conexiones que las conectan con las sinapsis. Tomado de [62].

979 algoritmo necesitan calcular el valor de la derivada en cada Z_m (en particular, el método
 980 aplicado en la sección 6.2.2 requiere esta propiedad). Por este motivo se emplean funciones
 981 como la función sigmoide o ReLU, que no son lineales y resultan fáciles de derivar.

982 En particular, la función ReLU (o funciones derivadas de esta) ha probado ser muy eficiente
 983 en redes neuronales modernas, entre otros motivos porque no se produce el *problema del*
 984 *gradiente que se desvanece*³ al contrario que con la función sigmoide.

985 Las funciones g_k no dejan de ser un tipo de función de activación, en las cuales se emplea
 986 una notación distinta ya que suelen tener expresiones diferentes a las de σ . En el caso de la
 987 función *softmax*, su popularidad en el ámbito de los problemas de clasificación reside en su
 988 no linealidad, su diferenciabilidad, y su propiedad $\sum_k^K g_k = \text{Id}_K$ que permite modelizar el
 989 vector de salida como un vector de probabilidades sobre cada clase.

³El problema del gradiente que se desvanece (*vanishing gradient problem*) se deriva de técnicas de optimiza-
 ción basadas en derivadas parciales de la función, las cuales toman valores cada vez más pequeños al evaluarse
 en Z_m en redes neuronales con muchas capas, suponiendo un problema en la velocidad de convergencia del
 valor de los pesos hacia su valor óptimo.

990 Se puede generalizar la construcción de redes neuronales de una capa a redes de un número
991 arbitrario de capas. Para ello, las unidades ocultas en una capa α , Z^α , tomarán como entrada
992 una combinación no lineal de los valores de la capa anterior, $Z^{\alpha-1}$, y su salida será empleada
993 en la capa siguiente.

994 Una red neuronal con una gran cantidad de capas se conoce como *Red Neuronal Profunda*
995 (*Deep Neural Network, DNN*). El estudio de estas redes pasó inadvertido debido a que,
996 como se comentaba anteriormente, un *PPR* es capaz de aproximar cualquier modelo con una
997 precisión tan grande como se desee aumentando el número M . En 2012, sin embargo, se
998 comenzó a plantear la posibilidad de construir algoritmos más eficientes mediante el uso de
999 más capas que permitían tomar M más pequeños [65] (dicha área es conocida como *Deep*
1000 *Learning*).

1001 **6.1.4. Red neuronal propuesta. Curva ROC**

1002 En este trabajo se introducen dos redes neuronales cuyo objetivo es el de realizar un aná-
1003 lisis multivariable sobre muones y electrones (respectivamente) en la producción $t\bar{t}$ para su
1004 identificación.

1005 Se aplicará cada algoritmo en sucesos que podrían haber producido un par $t\bar{t}$, el cual se
1006 desintegra en un leptón y el resto de productos son hadrónicos (*Single Lepton Channel*).

1007 Los algoritmos toman como entrada las 13 variables listadas en la sección 4.3 que describe
1008 el *Lepton MVA*, teniendo en cuenta que la última de ellas es distinta en el caso de muones
1009 (clasificador bayesiano ingenuo) al de electrones (bosque aleatorio). Así, en cada suceso, dichas
1010 variables tendrán un valor, y una descripción general de la muestra se incluye en la sección
1011 siguiente.

1012 La salida de las redes será un vector de dimensión cuatro, en el cual se listan las probabilidades
1013 de pertenecer a las categorías *Leptón Prompt*, *Tau Prompt*, *Light Fake* o *Heavy Fake*, que
1014 vienen descritas en la sección 5.6.

1015 La red neuronal tiene 4 capas distintas. El número de unidades ocultas en la primera capa es
1016 30, en la siguiente es 10, en la tercera 10, y en la última 4. Nótese que la última capa es la
1017 que se podía interpretar como vector de salida previo al que proporciona el algoritmo, como
1018 se introducía en la sección anterior, y siguiendo esta notación se corresponde con el vector T
1019 de igual dimensión que el vector final.

1020 La función de activación en cada capa es la función *ReLU*, a excepción de la capa final, en la
1021 que se toma la función *softmax* como es usual.

1022 La función pérdida que minimizará el algoritmo es la de *entropía cruzada categórica* (*cath-*
1023 *egorical cross-entropy*). Se trata de la función de pérdida usual en problemas de clasificación,
1024 debido a que está relacionada directamente con la información de Fisher de la muestra y se
1025 ha comprobado experimentalmente que, en general, tiene un mayor rendimiento en el con-
1026 texto del aprendizaje supervisado. Esta función actúa en cada elemento de la muestra de
1027 entrenamiento de la forma:

$$L(y, f(x)) = - \sum_{c=1}^4 y_c \log(f(x)_c),$$

1028 donde:

- 1029 ■ La entrada de la red es x (de dimensión 13), y la salida es $f(x)$ (de dimensión 4). El
1030 valor real de la observación con el que se entrena la red es y , un vector de 4 dimensiones
1031 de valor 1 en la categoría a la que pertenece, y valor 0 en el resto.
- 1032 ■ Los elementos y_c y $f(x)_c$ hacen referencia a cada uno de los componentes del vector
1033 y y $f(x)$. En la suma, solo será no nulo el componente asociado a la clase real de la
1034 observación ($y_c = 1$).
- 1035 ■ Una función de pérdida ideal tendría valor 0 ($f(x)_c = y_c$), y cuanto peor se ajusta el
1036 modelo, mayor será L .
- 1037 ■ Esta expresión se considera en una sólo observación. Para tener en cuenta una muestra
1038 de tamaño N , la función de pérdida final es una suma de valores de L aplicada a cada
1039 elemento de la muestra.

1040 En particular, en esta red se establecen unos pesos para cada categoría en la función de
1041 pérdida, para compensar que los datos introducidos contienen una mayor cantidad de sucesos
1042 de señal que de fondo. Esto se modeliza de la forma $L(y, f(x)) = - \sum_{c=1}^4 w_c y_c \log(f(x)_c)$, con
1043 $w_c = \frac{N_{sig} + N_{bkg}}{N_c}$, y

- 1044 ■ $N_{sig} = N_1 = N_2$: número de sucesos de señal.
- 1045 ■ $N_{bkg} = N_3 = N_4$: Número de sucesos de fondo.

Otro concepto que ayuda a evaluar la calidad del algoritmo tras su entrenamiento, es el valor de su función de precisión (*Accuracy Function*), que puede ser descrito por

$$AF = \frac{\text{Número de predicciones correctas}}{\text{Número total de predicciones}}.$$

1046 Un modelo fiable, por tanto, maximizará dicha función (acercándola al valor 1).

1047 El concepto final introducido en esta sección que también ayuda a comprobar la calidad del
1048 programa, es el de la *curva ROC (Receiver Operating Characteristic)*. Con ayuda de esta
1049 curva se puede analizar el rendimiento de un algoritmo de clasificación, no solo el de una
1050 red neuronal. En el contexto en el que se aplica la nueva red construida, ya se empleaban
1051 otros análisis multivariable con anterioridad, y mediante la curva *ROC* se puede comparar el
1052 rendimiento del popular análisis *BDT* aplicado con la misma muestra al de la red neuronal.

1053 Para comprobar de una curva *ROC* es necesario primero definir una serie de conceptos:

- 1054 ■ La *sensibilidad* de un algoritmo (o *proporción de verdaderos positivos, TPR*) denota
1055 la cantidad $TPR = \frac{TP}{TP+FN}$, donde *TP* es la cantidad de *verdaderos positivos* (ob-
1056 servaciones de señal clasificadas correctamente) y *FN* la cantidad de *falsos negativos*
1057 (observaciones de señal clasificadas como fondo, a menudo denotado por *error tipo II*).

1058 ▪ La *proporción de falsos positivos* (FPR) denota la cantidad $FPR = \frac{FP}{TN+FP}$, donde
1059 TN es la cantidad de *verdaderos negativos* (observaciones de fondo clasificadas correc-
1060 tamente) y FP es la cantidad de *falsos positivos* (observaciones de fondo clasificadas
1061 como señal, a menudo denotado *error tipo I*).

1062 La curva ROC establece la dependencia entre TPR y FPR . Se trata de una curva de pro-
1063 babilidad, y su integral (el *área debajo de la curva*, o AUC) denota la probabilidad de que
1064 el modelo realice una predicción correcta. Un modelo ideal tendrá un valor de $AUC = 1$
1065 (siempre predice correctamente la clase del objeto), mientras que el peor algoritmo posible
1066 tiene $AUC = 0.5$. Un algoritmo con AUC cercano a 0, a pesar de errar la gran mayoría de
1067 las veces, da más información que en el caso de $AUC \sim 0.5$, dado que basta con cambiar la
1068 etiqueta de señal y fondo para obtener un clasificador fiable.

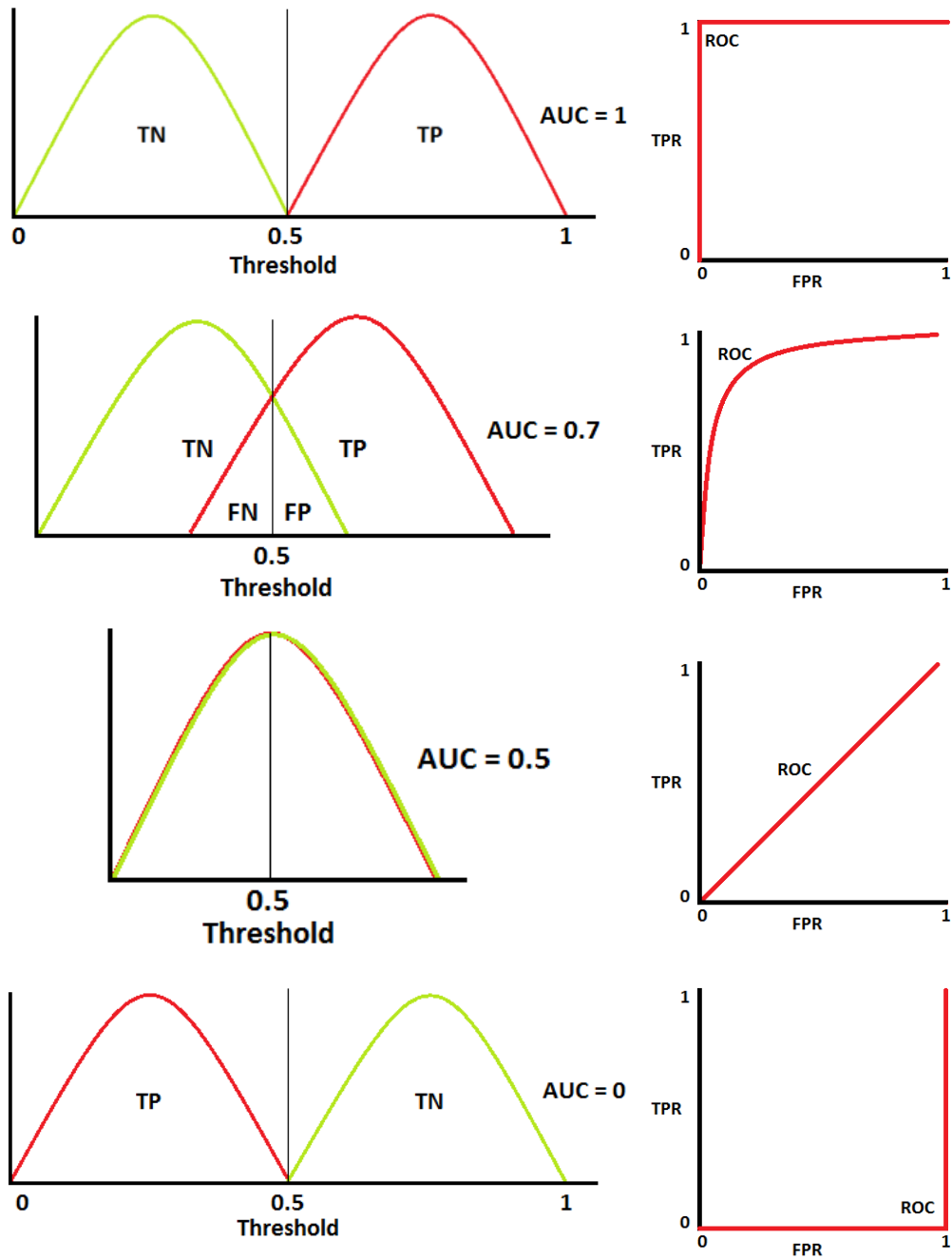


Figura 6.7: Ejemplo de algoritmos de clasificación y sus curvas *ROC*. Se representa, de arriba a abajo, el caso ideal en la clasificación ($AUC = 1$), un caso realista $AUC = 0.7$, un algoritmo que falla por completo al clasificar ($AUC = 0.5$), y un caso en el que el algoritmo se equivoca en toda clasificación ($AUC = 0$). Obsérvese como el caso superior y el inferior proporciona la misma información. Tomado de [66].

1069 **6.2. Entrenamiento de la red**

1070 **6.2.1. Consideraciones en el proceso experimental**

1071 Para una red neuronal óptima como la que se propone en la identificación de electrones, o
1072 su análoga en la identificación de muones, la elección de sus *hiperparámetros*⁴ característicos
1073 no puede ser arbitraria. En esta sección se describe el procedimiento de elección de cada
1074 hiperparámetro, así como se explica la influencia de cada uno de ellos sobre la red. Esta
1075 elección se realiza teniendo en cuenta las directrices de Leslie N. Smith en su publicación *A*
1076 *disciplined approach to neural network hyper-parameters* [67].

1077 En el estudio citado, a menudo se estudia el valor de la función de pérdida de la red con
1078 los distintos parámetros, para localizar el valor óptimo de ellos. Por otra parte, se emplea la
1079 función de precisión para la comparación de su rendimiento con el de otras redes neuronales.
1080 En el estudio particular de esta red se atenderá a dichas funciones, además del valor *AUC* de
1081 cada configuración del algoritmo.

1082 En todo momento hay algunas características que se mantienen (tales como el número de
1083 capas de la red o las unidades ocultas por capa), que son las introducidas en la sección
1084 anterior. Cada optimización de parámetros de la red es completamente dependiente de estas
1085 características, y no se espera que los mismos valores funcionen de manera óptima en un
1086 algoritmo de arquitectura distinta.

1087 En particular, se mantiene además el número de etapas (denominadas *épocas*) de entrena-
1088 miento de la red, fijado en 10 etapas. El entrenamiento de una red es un proceso cíclico que
1089 cambia los valores de los pesos y las funciones que relacionan las unidades ocultas conti-
1090 nuamente. Cuando todas las muestras han sido evaluadas, y la red ha sido reestructurada
1091 teniendo en cuenta cada una de ellas, se dice que se ha completado una época del entrena-
1092 miento; el siguiente paso considera las muestras desde el principio, comenzando la segunda
1093 época.

1094 Por otra parte, el valor óptimo de los hiperparámetros es más fácilmente perceptible con
1095 un aumento de tamaño muestral, pero este supone un mayor tiempo de entrenamiento del
1096 algoritmo. Dado que en el estudio de los valores óptimos se están realizando una gran cantidad
1097 de entrenamientos, que podrían tardar el orden de días en completarse cada uno de ellos, se
1098 decide tomar un 1 % de los datos, tomado completamente al azar. Esto no desplaza el valor
1099 óptimo, y los datos son suficientes para una correcta determinación del mismo, en un tiempo
1100 manejable.

1101 **6.2.2. Optimización de hiperparámetros**

1102 Hay distintos métodos para optimizar la función objetivo buscada en un problema de ML.
1103 La red propuesta emplea el método iterativo que se conoce como *Descenso del Gradiente*
1104 *Estocástico* (*Stochastic Gradient Descent, SGD*). Dicho método es aplicable en la optimización
1105 de funciones que son diferenciables con respecto de cada uno de sus parámetros, y resulta un
1106 método eficiente en este contexto dado que el coste computacional de las derivadas parciales

⁴Se denominan hiperparámetros de la red a aquellos que la caracterizan y son determinados a priori de su entrenamiento.

1107 de la función con respecto a cada parámetro es el mismo que el de la evaluación de la propia
1108 función.

1109 En el contexto de una red neuronal, la función objetivo final se describía como una com-
1110 binación de características construidas a partir de combinaciones lineales de las variables
1111 consideradas. Un método *SGD* puede resultar más eficiente en la evaluación de la derivada
1112 en cada una de las características.

1113 En particular, la red estudiada emplea un tipo de *SGD* conocido como *Adam*. El algoritmo
1114 de optimización que recibe este nombre se describe a continuación:

$$\begin{aligned} t &\leftarrow t + 1 \\ g_t &\leftarrow \Delta_{\theta} f_t(\theta_{t-1}) \text{ (Método de entrenamiento, empleando el gradiente)} \\ m_t &\leftarrow \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t \text{ (Actualiza el primer vector de momento)} \\ v_t &\leftarrow \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2 \text{ (Actualiza el segundo vector de momento)} \\ \hat{m}_t &\leftarrow \frac{m_t}{1 - \beta_1^t} \text{ (Corrección del primer momento)} \\ \hat{v}_t &\leftarrow \frac{v_t}{1 - \beta_2^t} \text{ (Corrección del segundo momento)} \\ \theta_t &\leftarrow \theta_{t-1} - \text{LR} \cdot \frac{\hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}} \text{ (Actualización de los parámetros)} \end{aligned} \tag{6.1}$$

1115 En el Algoritmo 6.1, β_i^t denota la potencia t de β_i , para $i = 1, 2$. En $t = 0$, se toman os valores
1116 iniciales $m_0 = 0$, $v_0 = 0$, y un cierto parámetro θ_0 aleatorio. El parámetro ϵ se toma $\epsilon = 10^{-8}$
1117 por defecto (evita divisiones por 0 en la parte final del algoritmo, y se toma pequeño para no
1118 influir en su convergencia).

1119 EL SGD empleado permite una variación previa a la ejecución del algoritmo sobre sus pa-
1120 rámetros LR, β_1 , β_2 , θ_0 y ϵ . No obstante, el análisis de su convergencia no se limita a la
1121 variación de estos parámetros. Se va a estudiar también la variación del número de observa-
1122 ciones evaluadas en cada iteración (*batch size*) así como técnicas que evitan la producción de
1123 *overfitting* sobre la red.

1124 Siguiendo el método de la referencia [67], el primer parámetro que se optimiza es el *learning*
1125 *rate* (LR). Es una de las características de la red que más impacto tienen en su entrenamiento.
1126 Un valor de LR demasiado grande puede provocar que el algoritmo no converja hasta el
1127 valor óptimo, mientras que un parámetro demasiado bajo puede provocar una convergencia
1128 demasiado lenta que requiera de más épocas (y por tanto, más recursos).

1129 Se realiza un estudio amplio sobre este parámetro, representado en la Figura 6.8. El objetivo
1130 en la optimización de hiperparámetros consiste en llegar a valor lo menor posible en la función
1131 pérdida.

1132 El siguiente hiperparámetro que se estudia en la optimización de la red, se corresponde con el
1133 *batch size*. De entre toda la muestra de entrenamiento, la evaluación de cada observación no
1134 puede realizarse (en general) de manera simultánea en todas ellas para tamaños de muestra
1135 lo suficientemente grandes. Las observaciones con las que se opera a la vez se denotan por
1136 *batch*.

1137 Un tamaño menor en el *batch* añade mayor tiempo de entrenamiento, pero lo compensa con
1138 una mayor estabilidad de la red (esto se debe a que la convergencia del *SGD* resulta más

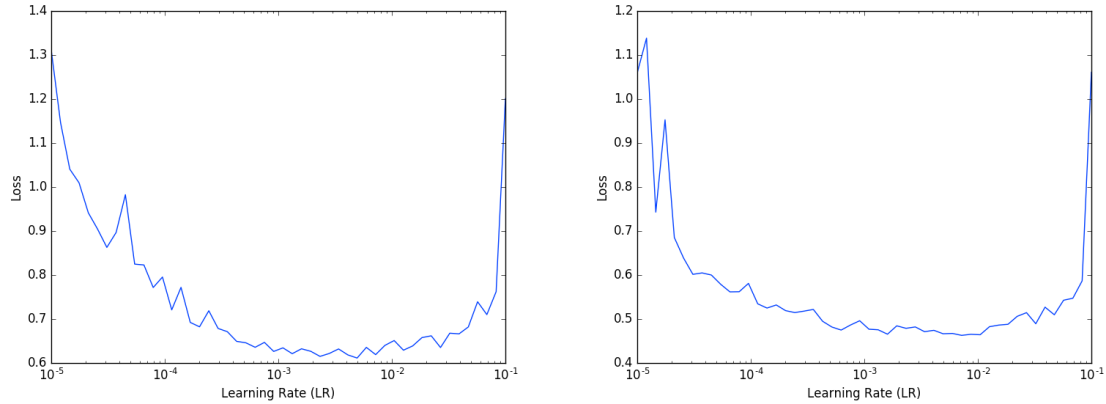


Figura 6.8: Estudio del *learning rate* de las redes neuronales (a la izquierda, la que identifica electrones; a la derecha, la que identifica muones). Se toma como valor óptimo $LR = 6 \times 10^{-3}$ para leptones e y $LR = 4 \times 10^{-3}$ para leptones μ .

1139 lenta). Por tanto, la elección del tamaño tampoco debería ser arbitraria en la red. Un estudio
 1140 sobre el *batch size* se representa en la Figura 6.9. El análisis anterior fue realizado con un
 1141 *batch* de tamaño 128.

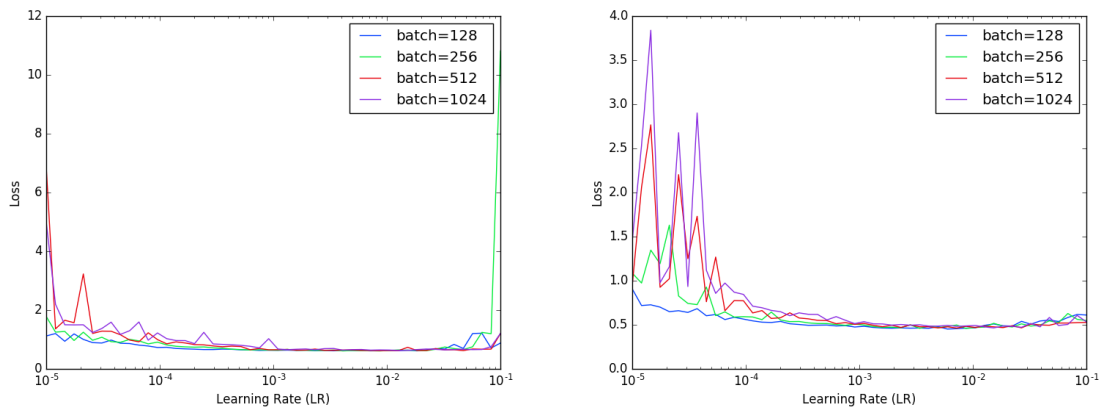


Figura 6.9: Estudio del *batch size* de las redes neuronales (a la izquierda, la que identifica electrones; a la derecha, la que identifica muones). Se toman potencias de 2 porque el costo computacional es menor de esta manera. Además, se estudia el efecto en conjunto con el LR.

1142 El momento es otro hiperparámetro que puede determinar la convergencia de un algoritmo,
 1143 de manera análoga al LR, e íntimamente relacionado con él. En esta red neuronal, su
 1144 optimización ha empleado dos parámetros de momento: β_1 y β_2 .

1145 Un estudio sobre ambos parámetros de momento se puede observar en la Figura 6.10. En
 1146 ella se observan los valores por defecto de dichos parámetros con los que se trabajó hasta el
 1147 momento, y se toman los valores de LR óptimos calculados anteriormente y un *batch* de 256.

1148 Un hiperparámetro importante en la regularización de la red es el *weight drop-out* (WD). Es

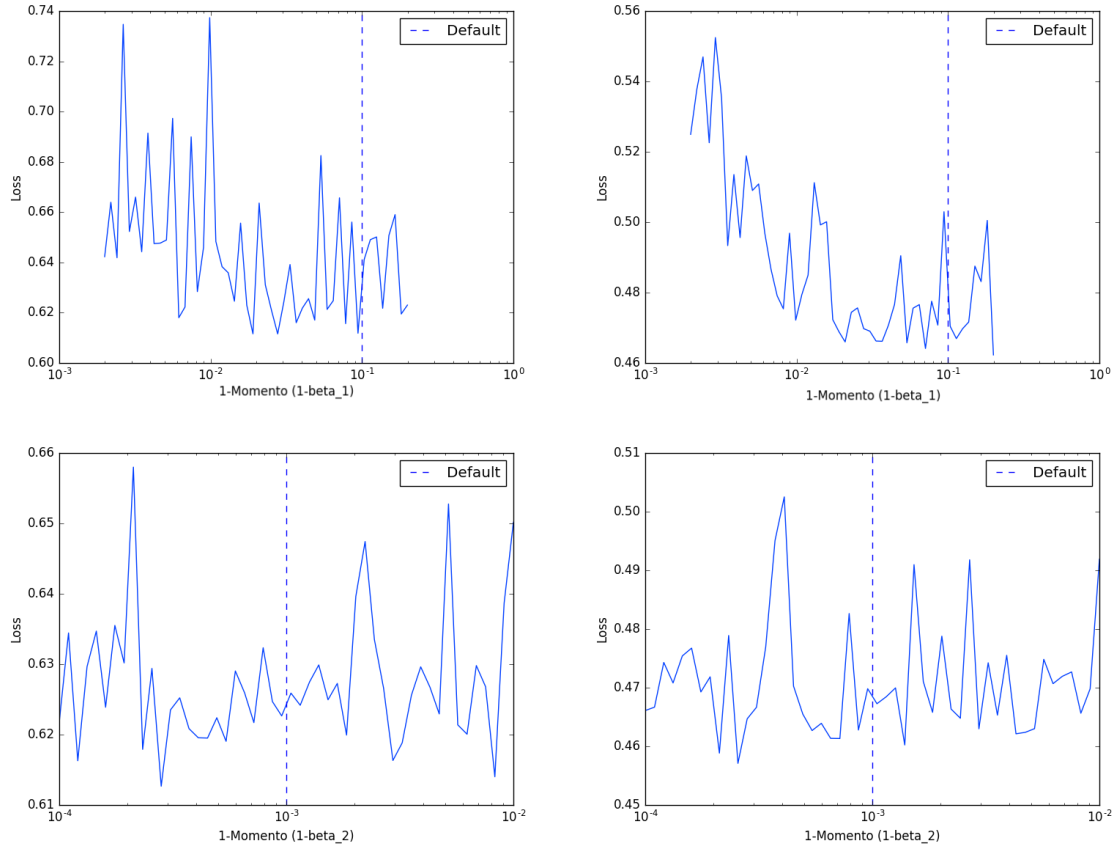


Figura 6.10: Estudio de los momentos de las redes neuronales (a la izquierda, la que identifica electrones; a la derecha, la que identifica muones). Se representa el valor $1 - \beta$ en el eje X . Las gráficas superiores estudian β_1 y las inferiores β_2 . La línea discontinua marca el valor inicial. No hay indicios de mejora al tomar valores distintos del predeterminado.

1149 una técnica empleada en cada capa de la red, desarrollada en 2014 [68].

1150 La técnica se basa en ignorar una proporción de neuronas de manera aleatoria en cada
 1151 ciclo de actualización de la red (iteración en 6.1). Al no actuar sobre el algoritmo final, las
 1152 demás neuronas deben evolucionar de manera que no dependan de las neuronas ignoradas,
 1153 provocando que el algoritmo evolucione a uno con unidades ocultas con mayor independencia
 1154 las unas de las otras.

1155 Se ha comprobado el efecto del *drop-out* sobre la red, en la Figura 6.11. Los valores que se
 1156 toman de este parámetro representan la cantidad de neuronas ignoradas en cada iteración. El
 1157 estudio se practica con los valores de LR tomados hasta ahora, un *batch* de 256 y el momento
 1158 predeterminado.

1159 También se han comprobado otros hiperparámetros no mencionados en [67]. Es el caso de la
 1160 *regularización en la actividad* de las unidades ocultas. Como su nombre indica, se trata de
 1161 una técnica de regularización de la red para evitar sobreentrenamiento.

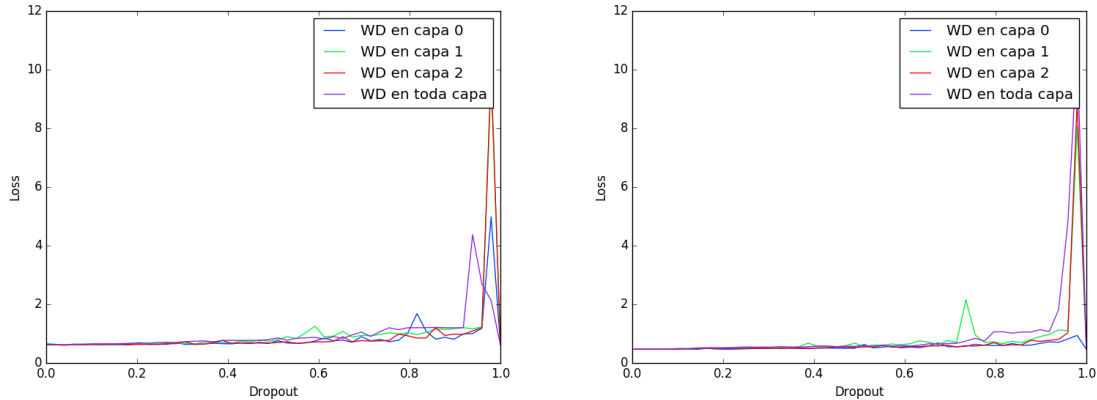


Figura 6.11: Estudio del efecto del *drop-out* en las redes neuronales (a la izquierda, la que identifica electrones; a la derecha, la que identifica muones). No se observan mejoras apreciables en la red al emplear esta herramienta.

1162 Es una técnica similar a las técnicas de regularización sobre los pesos en las unidades ocultas,
 1163 las cuales logran una mayor independencia entre neuronas. En este caso, se penaliza en la
 1164 función pérdida aquellas capas que tengan una alta actividad (es decir, cuyos pesos en las
 1165 unidades ocultas sean muy altos).

1166 Para el cálculo de la actividad de la capa, se emplean tres expresiones distintas. En el estudio
 1167 de la Figura 6.12 vienen referidas como l1, la cual toma la suma de los valores absolutos de
 1168 los pesos de las neuronas; l2, la cual toma la suma de los cuadrados de dichos pesos; y l1_l2,
 1169 la cual combina las dos anteriores.

1170 El enfoque en este estudio, dado que hay cuatro capas en las que se puede aplicar esta
 1171 regularización, fue de manera independiente a cada una de ellas. Se pudo comprobar como el
 1172 empleo de esta técnica no resultaba favorable a una red más optimizada en ninguna capa, y
 1173 se descartó un estudio conjunto. El valor de LR fue el tomado hasta ahora, el batch fue 256,
 1174 el momento no se varió y no se aplicó *drop-out*.

1175 Por último, se pretendió estudiar el efecto de valores de LR que varían en cada época (*Cyclical*
 1176 *Learning Rates*). Se midió el efecto en la red al aumentar el valor de este hiperparámetro
 1177 hasta 10 veces su valor inicial de manera proporcional en cada época, así como el efecto en
 1178 su disminución hasta una décima parte.

1179 Se estudiaron distintas maneras de variar el LR entre un valor mínimo y uno máximo ([69]),
 1180 encontrando que estas variaciones resultan equivalentes y por lo tanto recomendando la más
 1181 simple, que es la variación lineal practicada en este estudio (Figura 6.13). En particular, se
 1182 varían los LR considerados como óptimos hasta el momento, denotados por $LR_{inicial}$, hasta
 1183 un cierto valor, LR_{final} . En el estudio se representa el valor $LR_{final}/LR_{inicial}$ tomando un
 1184 rango de valores desde 0.1 hasta 10, con *batch* 256 y momentos por defecto.

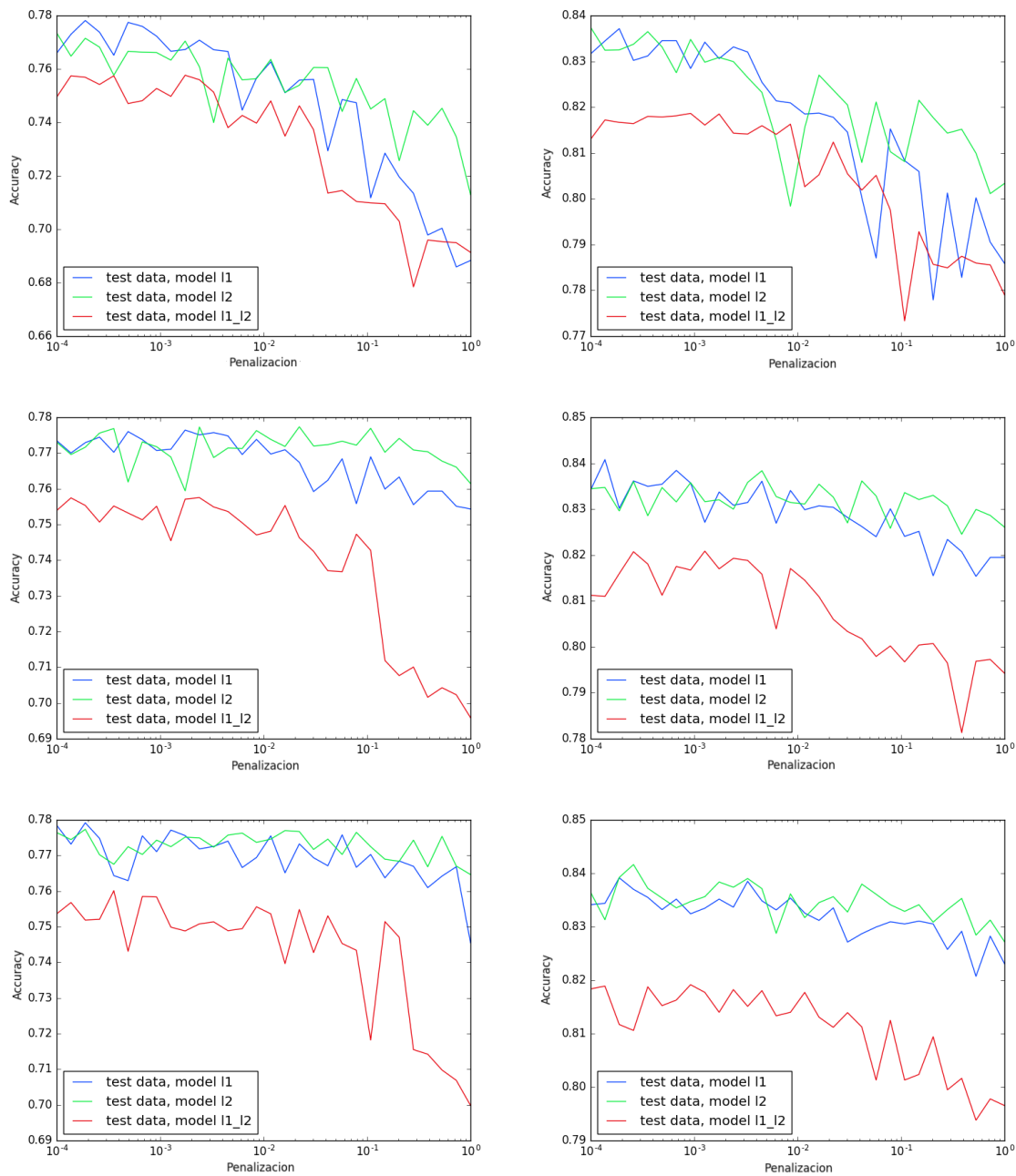


Figura 6.12: Estudio independiente del efecto de la regularización de la actividad en cada una de las capas de las redes. Se representa la penalización aplicada en cada técnica frente a la función de precisión. El empleo de la técnica no produce mejoras en el modelo.

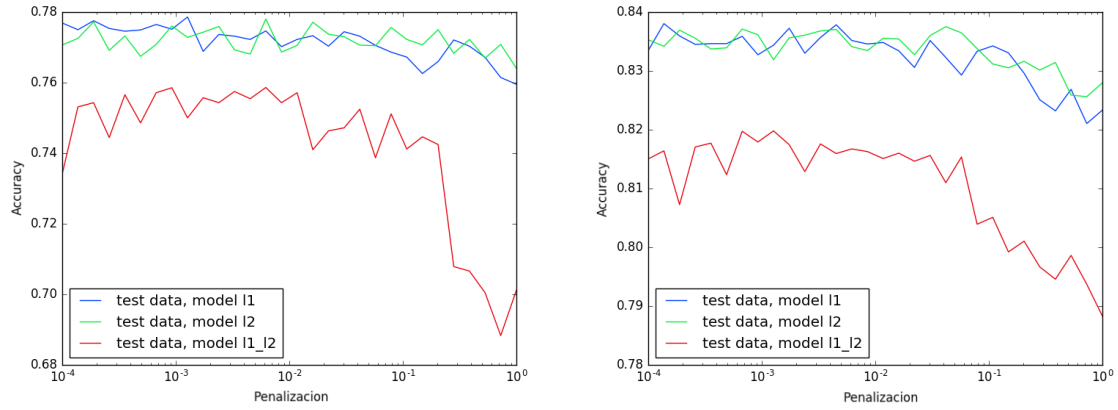


Figura 6.12: Estudio independiente del efecto de la regularización de la actividad en cada una de las capas de las redes (continuación).

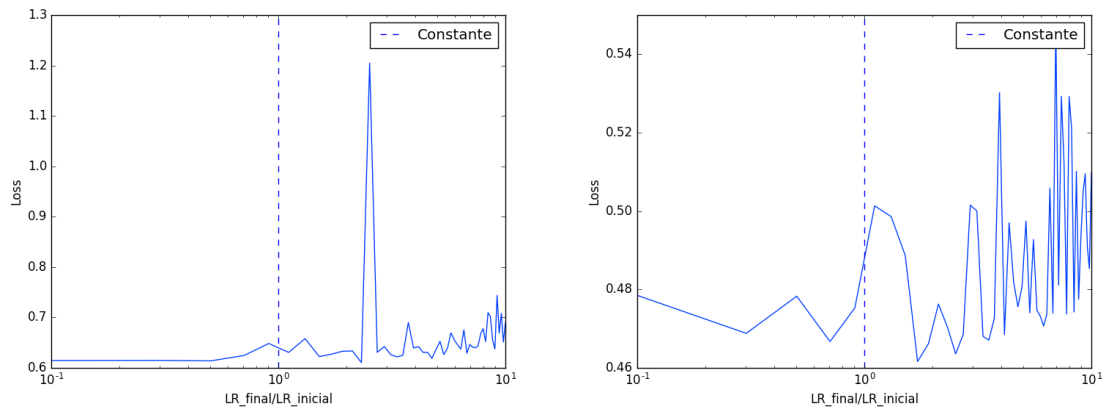


Figura 6.13: Estudio de la variación (lineal) por época del LR , donde se representa la pérdida obtenida según su valor final. La figura de la izquierda representa el estudio en electrones, y la de la derecha en muones. No se obtienen aparentes mejoras al aplicar un LR no constante (CLR).

1185 Capítulo 7

1186 Resultados obtenidos y conclusiones

1187 Finalmente, se evalúa el modelo construido para comprobar su validez en un entorno práctico.
1188 En este capítulo se compara su rendimiento con el de un *BDT* en el mismo contexto, y se
1189 discute el modelo final.

1190 7.1. Resultados obtenidos

1191 El modelo generado por la red neuronal ha resultado ser muy efectivo: se obtuvo un valor de
1192 $AUC = 0.968$ en el algoritmo discriminante de electrones, y $AUC = 0.982$ en el caso aplicado
1193 a muones. Estos valores resultan ser algo superiores a los de referencia ($AUC = 0.966$ en
1194 electrones y $AUC = 0.980$ en muones, obtenidos del *BDT*). Esto indica que el algoritmo
1195 entrenado está en condiciones de ser aplicado en el mismo contexto que el algoritmo *BDT*.

1196 Para comprobar su actuación sobre la muestra de validación, se ha representado en figuras
1197 la respuesta del programa ante los sucesos pertenecientes a cada clase (Figuras 7.2 y 7.3).
1198 Se representa en el eje X la probabilidad del algoritmo asignada a la clase, y en el eje Y la
1199 proporción de observaciones en cada intervalo de probabilidad (se trazan de distintos colores
1200 las distintas clasificaciones del programa para dicha muestra). Por ejemplo, una situación
1201 ideal en este tipo de histograma en la clasificación de elementos donde se ha producido un
1202 *heavy fake*, colocaría todo el peso de dicha categoría en los valores cercanos a 1 de la X , y el
1203 resto de categorías se encontrarían con valores cercanos a 0.

1204 A la vista de las gráficas, se puede observar como en la clasificación de sucesos *prompt*, el
1205 algoritmo distingue con éxito la señal de fondo, pero con algo menos de rigurosidad los su-
1206 cesos *prompt* de los *prompt* τ . Esto ocurre en ambos algoritmos. Más aún, en la clasificación
1207 de sucesos *prompt* τ el clasificador de electrones no llega a clasificar con seguridad las ob-
1208 servaciones en esta categoría, efecto que se ve ciertamente mitigado pero no anulado en el
1209 clasificador de muones.

1210 Los sucesos que producen *heavy fakes* son correctamente clasificados la mayoría de veces,
1211 análogamente al caso de los leptones *prompt*. Por otra parte, sucesos que producen *light*
1212 *fakes* no son clasificados, en general, con probabilidades cercanas a 1 por el algoritmo en su
1213 categoría correspondiente en el caso de electrones.

1214 7.2. Conclusiones y perspectivas

1215 Las redes neuronales presentadas han resultado ser completamente competentes en una tarea
1216 de distinción entre sucesos de señal y de fondo en la producción de pares $t\bar{t}$.

1217 La arquitectura del problema ha permitido un buen rendimiento de la red con la imple-
1218 mentación de cuatro capas. La consideración de redes neuronales con un mayor número de
1219 capas podría producir mejores resultados, si bien es cierto que el margen de mejora no será
1220 muy amplio a la vista de estos resultados. La simpleza de la red, al entrenarse en tan pocas
1221 dimensiones no solo es responsable del número de capas implementado; además, evita que
1222 tengan que aplicarse herramientas de regularización, como se ha comprobado en la sección
1223 6.2.2, donde en ocasiones incluso empeoraba el rendimiento del programa por su uso.

1224 Tras un estudio exhaustivo de las propiedades de cada red, se determinó que bastaba con
1225 un ajuste de cada factor de aprendizaje correspondiente, Un exceso en la regularización del
1226 problema tiene como inconveniente la aparición de *underfitting* en el algoritmo, algo que es
1227 preciso evitar.

1228 Problemas más complejos podrían requerir una mayor cantidad de capas en la red para que
1229 funcione a una velocidad aceptable (la cual depende del contexto en el que se aplique cada
1230 red, y de la que dependerá el número de épocas de entrenamiento que podrá realizar). En
1231 estos problemas el estudio anterior podría resultar completamente distinto, y herramientas
1232 como el *drop-out* se pueden volver muy determinantes en la optimización de hiperparámetros
1233 de la red.

1234 Por último, la decisión sobre el empleo de la red neuronal en un análisis real de leptones
1235 producidos en sucesos $t\bar{t}$ no solo debe tener en cuenta la efectividad del algoritmo. Es preciso
1236 notar que la mejora ha sido baja con respecto a la técnica anterior en este contexto, así como
1237 señalar la rapidez computacional y facilidad en la interpretación que se obtienen en el empleo
1238 de un *BDT*.

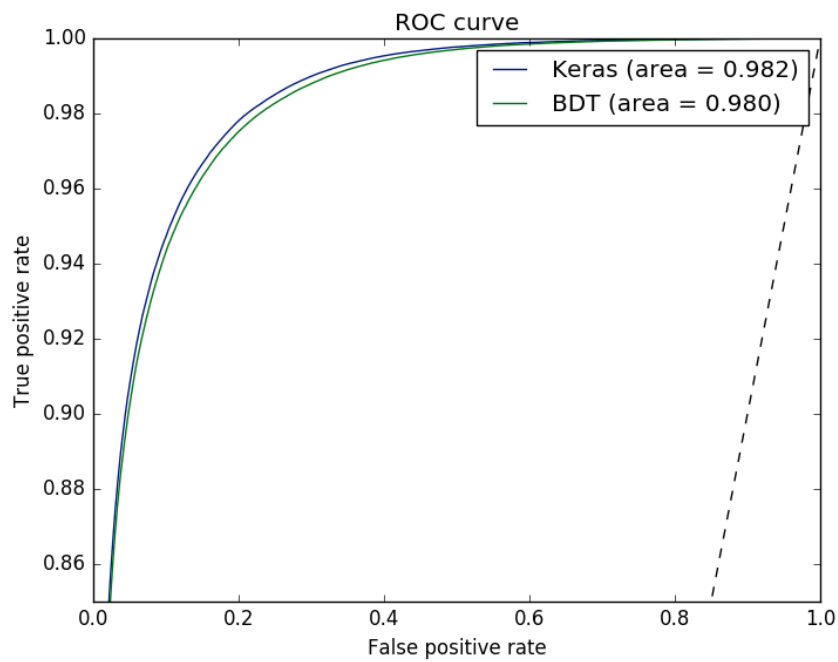
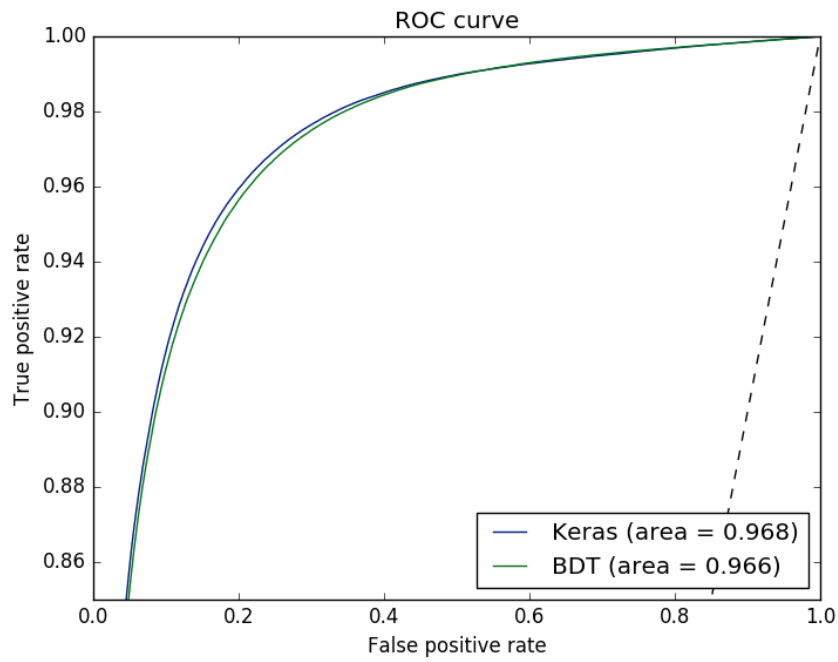


Figura 7.1: Curvas *ROC* obtenidas a partir de la red neuronal (azul) y el *BDT* de referencia (verde). Se representa arriba el algoritmo especializado en electrones, y abajo el de muones. Los valores de la red resultan ser ligeramente superiores a los del árbol de decisión.

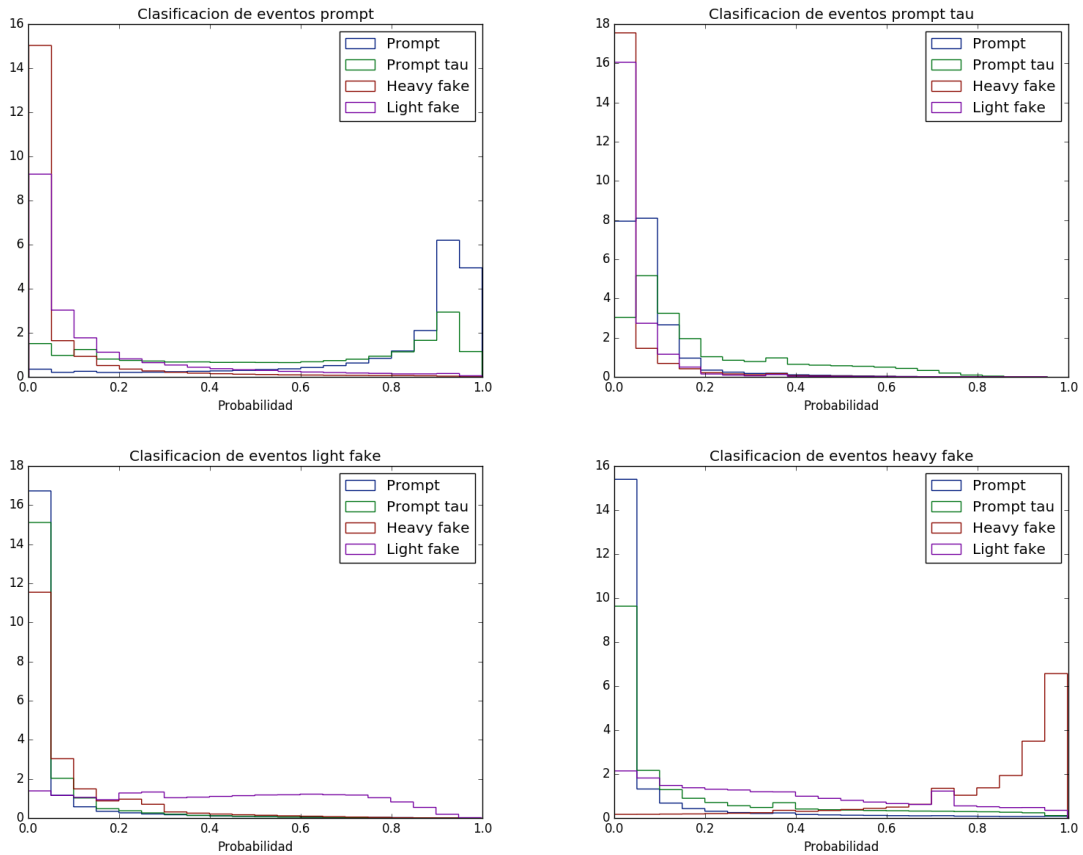


Figura 7.2: Representación de la clasificación del programa de los sucesos. Cada gráfica representa una clase de observaciones, etiquetada con anterioridad al programa. En el eje X se representa la probabilidad asignada por el algoritmo a que la observación pertenezca a la clase denotada por el color de la línea. Algoritmo especializado en la identificación de electrones.

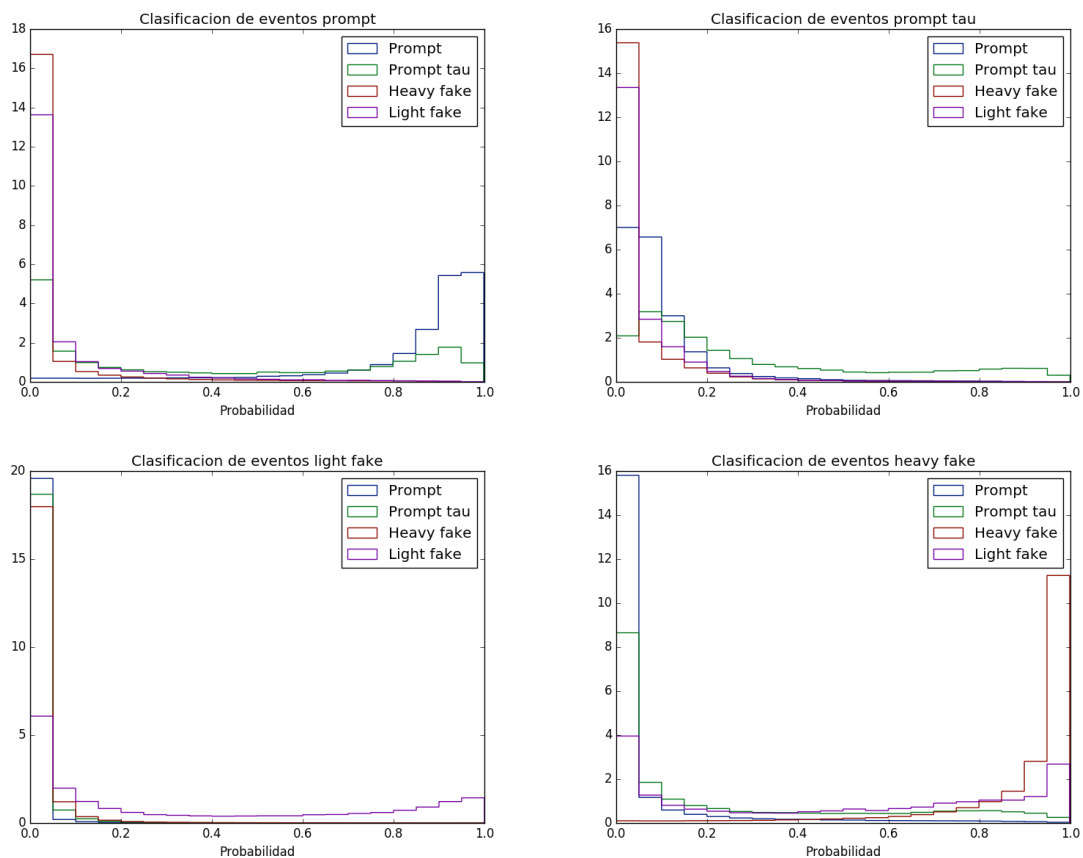


Figura 7.3: Representación de la clasificación del programa de los sucesos. Cada gráfica representa una clase de observaciones, etiquetada con anterioridad al programa. En el eje X se representa la probabilidad asignada por el algoritmo a que la observación pertenezca a la clase denotada por el color de la línea. Algoritmo especializado en la identificación de muones.

Bibliografía

1239

- 1240 [1] Standard Model of Elementary Particles [imagen online]. Tomada el 12 de octubre de
1241 2019 de https://en.wikipedia.org/wiki/Standard_Model.
- 1242 [2] Riordan, M. (1992). The Discovery of Quarks. *Science*, 256(5061). doi:10.1126/scien-
1243 ce.256.5061.1287
- 1244 [3] Rohlf, J. W. (1994). Modern Physics from a to Z0. Nueva York: John Wiley & Sons.
- 1245 [4] Tanabashi M. et al. (Particle Data Group) (2018). Review of Particle Physics. *Physical*
1246 *Review D*, 98(030001). doi:10.1103/PhysRevD.98.030001.
- 1247 [5] Álvarez González, B. (2019). Fenomenología del Modelo Estándar y sus exten-
1248 siones [Diapositivas]. Tomado de [http://www.hep.uniovi.es/balvarez/FMEye1920/](http://www.hep.uniovi.es/balvarez/FMEye1920/Intro-FMEye(1).pdf)
1249 [Intro-FMEye\(1\).pdf](http://www.hep.uniovi.es/balvarez/FMEye1920/Intro-FMEye(1).pdf).
- 1250 [6] The CMS Collaboration (2012). Observation of a new boson at a mass of 125 GeV
1251 with the CMS experiment at the LHC. *Physics Letters B*, 716(1). doi:10.1016/j.phys-
1252 letb.2012.08.021.
- 1253 [7] The CMS Collaboration (2013). The Discovery of the Higgs Boson with the CMS Detector
1254 and its Implications for Supersymmetry and Cosmology. *Time and Matter 2013*. Tomado
1255 de <http://inspirehep.net/record/1252561?ln=es>.
- 1256 [8] Tonon, N. (2019). *Recherche de la production d'un quark top solitaire associé à un boson*
1257 *Z ou un boson de Higgs avec l'expérience CMS au LHC* (tesis doctoral). Universidad de
1258 Estrasburgo, Alsacia, Francia.
- 1259 [9] The CMS Collaboration (2019). Combined measurements of Higgs boson couplings
1260 in proton–proton collisions at $\sqrt{s} = 13\text{TeV}$. *The European Physical Journal C*, 79(5).
1261 doi:10.1140/epjc/s10052-019-6909-y.
- 1262 [10] Swart, J. G., Bertone, G. & van Dongen, J. (2017). How dark matter came to matter.
1263 *Nature Astronomy*, 1(3). doi:10.1038/s41550-017-0059.
- 1264 [11] Dark Energy, Dark Matter. Tomado el 13 de octubre de 2019 de [https://science.](https://science.nasa.gov/astrophysics/focus-areas/what-is-dark-energy)
1265 [nasa.gov/astrophysics/focus-areas/what-is-dark-energy](https://science.nasa.gov/astrophysics/focus-areas/what-is-dark-energy).
- 1266 [12] Queiroz, F. S. (2017). WIMP Theory Review. *European Physical Society Conference on*
1267 *High Energy Physics*. Tomado de <https://arxiv.org/abs/1711.02463>

- 1268 [13] NASA's Goddard Space Flight Center (20 de Septiembre de 2016). GMS: Content of the
1269 Universe Pie Chart [Archivo de vídeo]. Tomado de <https://svs.gsfc.nasa.gov/12307>.
- 1270 [14] Höcker, A. (2006). CP Violation and the CKM Matrix. *Annual Review of Nuclear and*
1271 *Particle Science*, 56(501-567). doi:10.1146/annurev.nucl.56.080805.140456.
- 1272 [15] LHC Season 2: facts & figures. Tomado el 10 de Noviembre de 2019 de <https://run2-13tev.web.cern.ch/background/lhc-season-2-facts-figures>.
1273
- 1274 [16] Luminosity? Why don't we just say collision rate? (2011, 2 de
1275 marzo), *CERN*. Tomado de [https://home.cern/news/opinion/cern/](https://home.cern/news/opinion/cern/luminosity-why-dont-we-just-say-collision-rate)
1276 [luminosity-why-dont-we-just-say-collision-rate](https://home.cern/news/opinion/cern/luminosity-why-dont-we-just-say-collision-rate).
- 1277 [17] CERN Map [Imagen]. Tomado el 12 de octubre de 2019 de <https://lhcb-public.web.cern.ch/lhcb-public/en/Detector/Detector-en.html>.
1278
- 1279 [18] The Large Electron-Positron Collider. Tomado el 10 de Noviembre de 2019 de <https://home.cern/science/accelerators/large-electron-positron-collider>
1280
- 1281 [19] The CMS Collaboration (2018). The CERN accelerator complex [Imagen]. Tomado de
1282 <http://cds.cern.ch/record/2636343/files/>
- 1283 [20] Release the beams! Linac 4 hits the 50 MeV mark (2015). Tomado de <https://home.cern/news/news/accelerators/release-beams-linac-4-hits-50-mev-mark>.
1284
- 1285 [21] The Proton Synchrotron Booster. Tomado el 10 de Noviembre de 2019 de <https://home.cern/science/accelerators/proton-synchrotron-booster>.
1286
- 1287 [22] LHC injection tests to begin (2015). Tomado de 2019 de [https://home.cern/news/](https://home.cern/news/news/accelerators/lhc-injection-tests-begin)
1288 [news/accelerators/lhc-injection-tests-begin](https://home.cern/news/news/accelerators/lhc-injection-tests-begin).
- 1289 [23] Taking a closer look at LHC - LHC layout. Tomado el 10 de Noviembre de 2019 de
1290 https://www.lhc-closer.es/taking_a_closer_look_at_lhc/0.lhc_layout.
- 1291 [24] Rossi, L. (2003). The LHC superconducting magnets. *20th IEEE Particle Accelera-*
1292 *tor Conference*. Tomado de [http://accelconf.web.cern.ch/accelconf/p03/papers/](http://accelconf.web.cern.ch/accelconf/p03/papers/toab001.pdf)
1293 [toab001.pdf](http://accelconf.web.cern.ch/accelconf/p03/papers/toab001.pdf).
- 1294 [25] CMS. Tomado el 10 de Noviembre de 2019 de [https://home.cern/science/](https://home.cern/science/experiments/cms)
1295 [experiments/cms](https://home.cern/science/experiments/cms).
- 1296 [26] ATLAS. Tomado el 10 de Noviembre de 2019 de [https://home.cern/science/](https://home.cern/science/experiments/atlas)
1297 [experiments/atlas](https://home.cern/science/experiments/atlas).
- 1298 [27] Taylor, L. (2011). CMS detector design. Tomao de [http://cms.web.cern.ch/news/](http://cms.web.cern.ch/news/cms-detector-design)
1299 [cms-detector-design](http://cms.web.cern.ch/news/cms-detector-design)
- 1300 [28] Magnets & Detectors I. Tomado el 23 de Octubre de 2019 de [https://www.lhc-closer.](https://www.lhc-closer.es/taking_a_closer_look_at_lhc/0.magnets__detectors_i)
1301 [es/taking_a_closer_look_at_lhc/0.magnets__detectors_i](https://www.lhc-closer.es/taking_a_closer_look_at_lhc/0.magnets__detectors_i).
- 1302 [29] Taylor, L. (2011). Superconducting Magnet. Tomado de [http://cms.web.cern.ch/](http://cms.web.cern.ch/news/superconducting-magnet)
1303 [news/superconducting-magnet](http://cms.web.cern.ch/news/superconducting-magnet).

- 1304 [30] Taylor, L. (2011). Tracker detector. Tomado de [http://cms.web.cern.ch/news/
1305 tracker-detector](http://cms.web.cern.ch/news/tracker-detector).
- 1306 [31] Silicon Pixels. Tomado el 20 de Octubre de 2019 de [https://cms.cern/detector/
1307 identifying-tracks/silicon-pixels](https://cms.cern/detector/identifying-tracks/silicon-pixels).
- 1308 [32] Silicon Strips. Tomado el 20 de Octubre de 2019 de [http://cms.web.cern.ch/news/
1309 silicon-strips](http://cms.web.cern.ch/news/silicon-strips).
- 1310 [33] Lima, R. (2017). Overview of Energy Reconstruction, and Electron and Photon Perfor-
1311 mances with the CMS ECAL in RUN II. *Journal of Physics: Conference Series*, 928(1).
1312 doi:10.1088/1742-6596/928/1/012005.
- 1313 [34] Chatrchyan, S. et al. (2008). The CMS experiment at the CERN LHC. *Journal of Ins-
1314 trumentation*, 3(Agosto 2008). doi:10.1088/1748-0221/3/08/S08004.
- 1315 [35] Colaleo, A., Safonov, A., Sharma, A. & Tytgat, M. (2015). CMS Technical Design Report
1316 for the Muon Endcap GEM Upgrade. *Technical Design Report CMS*, 13.
- 1317 [36] Min Suk, K. (2013). CMS reconstruction improvement for the muon tracking by
1318 the RPC chambers. *Journal of Instrumentation*, 8(Marzo 2013). doi:10.1088/1748-
1319 0221/8/03/T03001.
- 1320 [37] The CMS Collaboration (2013). The performance of the CMS muon detector in
1321 proton-proton collisions at $\sqrt{s} = 7\text{TeV}$ at the LHC. *Journal of Instrumentation*, 8(2).
1322 doi:10.1088/1748-0221/8/11/P11002.
- 1323 [38] Williams, J. (2007). CMS Particle Detection Summary [imagen]. Tomado de [https:
1324 //twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookCMSExperiment](https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookCMSExperiment).
- 1325 [39] The CMS Collaboration (2016). The CMS trigger system. *Journal of Instrumentation*,
1326 12(Enero 2017). doi:10.1088/1748-0221/12/01/P01020
- 1327 [40] Guest, D., Cranmer, K. & Whiteson, D. (2018). Deep Learning and its Applica-
1328 tion to LHC Physics. *Annual Review of Nuclear and Particle Science*, 68(161-181).
1329 doi:10.1146/annurev-nucl-101917-021019.
- 1330 [41] Sjostrand, T., Mrenna, S. & Skands, P. Z. (2007). A Brief Introduction to PYTHIA 8.1.
1331 *Computer Physics Communications*, 178(11). doi:10.1016/j.cpc.2008.01.036.
- 1332 [42] The Geant4 Collaboration (2019). Introduction to Geant4. Tomado
1333 de [http://geant4-userdoc.web.cern.ch/geant4-userdoc/UsersGuides/
1334 IntroductionToGeant4/fo/IntroductionToGeant4.pdf](http://geant4-userdoc.web.cern.ch/geant4-userdoc/UsersGuides/IntroductionToGeant4/fo/IntroductionToGeant4.pdf).
- 1335 [43] The CMS Collaboration (2009). Particle-Flow Event Reconstruction in CMS and Per-
1336 formance for Jets, Taus, and MET. Tomado de <https://cds.cern.ch/record/1194487>.
- 1337 [44] The CMS Collaboration (2018). Performance of reconstruction and identification of τ
1338 leptons decaying to hadrons and ν_τ in pp collisions at $\sqrt{s} = 13\text{ TeV}$. *Journal of Instru-
1339 mentation*, 13(Octubre 2018). doi:10.1088/1748-0221/13/10/P10005
- 1340 [45] Cacciari, M., Salam, G. P. & Soyez, Gregory (2008). The anti- k_t jet clustering algorithm.
1341 *Journal of High Energy Physics*, 2008(4). doi:10.1088/1126-6708/2008/04/063.

- 1342 [46] The CMS Collaboration (2016). Identification of b quark jets at the CMS Experiment
1343 in the LHC Run 2. Tomado de <https://cds.cern.ch/record/2138504>.
- 1344 [47] The CMS Collaboration (2018). Observation of $t\bar{t}H$ production. *Physical Review Letter*,
1345 *120*(231801). doi:10.1103/PhysRevLett.120.231801.
- 1346 [48] Ilisie, V. (2011). S.M. Higgs Decay and Production Channels (*Trabajo de Fin de Máster*).
1347 Universidad de Valencia, Valencia, España.
- 1348 [49] The CMS Collaboration (2014). Measurement of the ratio $\mathcal{B}(t \rightarrow Wb)/\mathcal{B}(t \rightarrow Wq)$ in pp
1349 collisions at $\sqrt{s} = 8\text{TeV}$. *Physics Letters B*, *736*(1). doi:10.1016/j.physletb.2014.06.076.
- 1350 [50] University of Zurich. Semileptonic decays (2018). Tomado de <https://www.fysik.uzh.ch/en/researcharea/lhcb/B-physics/semileptonic-decays.html>.
1351
- 1352 [51] Leptones Drell-Yan [imagen] (2017). Tomado de [https://wiki.fysik.uzh.ch/cms/
1353 _detail/latex:drell-yan_leptons.png?id=latex%3Afeynman](https://wiki.fysik.uzh.ch/cms/_detail/latex:drell-yan_leptons.png?id=latex%3Afeynman).
- 1354 [52] D. de Florian et al. (2016). Handbook of LHC Higgs Cross Sections: 4. Deciphering the
1355 Nature of the Higgs Sector. *CERN Yellow Reports: Monographs*, *2*. doi:10.23731/CYRM-
1356 2017-002
- 1357 [53] The CMS Collaboration (2018). Measurement of the associated production of a Higgs
1358 boson with a top quark pair in final states with electrons, muons and hadronically decaying
1359 τ leptons in data recorded in 2017 at $\sqrt{s} = 13\text{ TeV}$. Tomado de [http://cds.cern.ch/
1360 record/2649199](http://cds.cern.ch/record/2649199).
- 1361 [54] The CMS Collaboration (2017). Measurement of the $t\bar{t}$ production cross section using
1362 events in the $e\mu$ final state in pp collisions at $\sqrt{s} = 13\text{ TeV}$. *European Physical Journal C*,
1363 *77*(3). doi:10.1140/epjc/s10052-017-4718-8
- 1364 [55] The CMS Collaboration, 2017. Measurement of the $t\bar{t}$ production cross section using
1365 events with one lepton and at least one jet in pp collisions at $\sqrt{s} = 13\text{ TeV}$. *Journal of
1366 High Energy Physics*, *2017*(9). doi:10.1007/JHEP09(2017)051
- 1367 [56] Landsberg, G. (2013). Summary of the cross section measurements of Standard Model
1368 processes [Imagen]. Tomado de [https://twiki.cern.ch/twiki/bin/view/CMSPublic/
1369 PhysicsResultsCombined](https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsCombined).
- 1370 [57] Dorney, B. (2011). Anatomy of a Jet in CMS [Entrada de blog]. Tomado de [https:
1371 //www.quantumdiaries.org/2011/06/01/anatomy-of-a-jet-in-cms/](https://www.quantumdiaries.org/2011/06/01/anatomy-of-a-jet-in-cms/).
- 1372 [58] Rehermann, K. & Tweedie, B. (2010). Efficient Identification of Boosted Se-
1373 mileptonic Top Quarks at the LHC. *Journal of High Energy Physics*, *2011*(3).
1374 doi:10.1007/JHEP03(2011)059.
- 1375 [59] Charitos, P. (2017). Deep learning and the quest for new phy-
1376 sics at the LHC. Tomado de [https://ep-news.web.cern.ch/content/
1377 deep-learning-and-quest-new-physics-lhc](https://ep-news.web.cern.ch/content/deep-learning-and-quest-new-physics-lhc).
- 1378 [60] Overfitting (2018) [Imagen]. Tomado de [https://www.tradesmart.cz/tag/
1379 overfitting/](https://www.tradesmart.cz/tag/overfitting/).

- 1380 [61] Tek Around (2018). Train, Validation, Test Set in Machine Learning— How to understand. Tomado de [https://medium.com/@tekaround/](https://medium.com/@tekaround/train-validation-test-set-in-machine-learning-how-to-understand-6cdd98d4a764)
1381 [train-validation-test-set-in-machine-learning-how-to-understand-6cdd98d4a764](https://medium.com/@tekaround/train-validation-test-set-in-machine-learning-how-to-understand-6cdd98d4a764).
1382
- 1383 [62] Hastie, T., Tibshirani, R. & Friedman, J. (2017). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Second Edition: Springer.
1384
- 1385 [63] Freund, Y. & Schapire, E. (1997). A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences*, 55(1).
1386 doi:10.1006/jcss.1997.1504
1387
- 1388 [64] Hornik, K., Stinchcombe, M., White, H. (1989). Multilayer feedforward networks are universal approximators. *ScienceDirect*, 2(5). doi:10.1016/0893-6080(89)90020-8.
1389
- 1390 [65] Krizhevsky, A., Sutskever, I. & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural network. *Advances in neural information processing systems*, 25(2).
1391 doi:10.1145/3065386.
1392
- 1393 [66] Narkhede, S. (2018). Understanding AUC - ROC Curve. Tomado de [https://](https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5)
1394 towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5.
- 1395 [67] Smith, L. N. (2018). A disciplined approach to neural network hyper-parameters: Part 1 – learning rate, batch size, momentum, and weight decay. *US Naval Research Laboratory Technical Report 5510-026*.
1396
1397
- 1398 [68] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, 15(1929-1958).
1399
1400
- 1401 [69] Smith, L.N. (2017). Cyclical Learning Rates for Training Neural Networks. *Winter Conference on Applications of Computer Vision 2017*. Tomado de [https://arxiv.org/abs/](https://arxiv.org/abs/1506.01186)
1402 [1506.01186](https://arxiv.org/abs/1506.01186).
1403
- 1404 [70] Rodríguez Bouza, V. (2017). *Estudio del proceso $t\bar{t}H$ en el experimento CMS del acelerador LHC (CERN) a 13 TeV* (tesis de pregrado). Universidad de Oviedo, Asturias, España.
1405
1406
- 1407 [71] Al Dallah, S. y Zwicky, F. (2012). On Supersymmetry and the Origin of Dark Matter. *Journal of Modern Physics*, 3(9A).
1408
- 1409 [72] Jowett, J. (2015). LHC report: Plumbing new heights. *CERN Bulletin*, 2015(49,50).