



SAPIENZA
UNIVERSITÀ DI ROMA

Applicazione di tecniche di analisi multivariata in fisica delle alte energie

Facoltà di Scienze Matematiche, Fisiche e Naturali
Corso di Laurea in Fisica

Candidato

Antonio Norelli

Matricola 1612487

Relatore

Prof. Stefano Giagu

Anno Accademico 2015/2016

Applicazione di tecniche di analisi multivariata in fisica delle alte energie

Tesi di Laurea. Sapienza – Università di Roma

© 2016 Antonio Norelli. Tutti i diritti riservati

Questa tesi è stata composta con L^AT_EX e la classe Sapthesis.

Email dell'autore: noranta4@gmail.com

*ad Anna Parisi e a mia mamma,
che ha insistito perché leggessi il suo primo libro*

Indice

Introduzione	2
1 Ricerca di fotoni dark con il rivelatore ATLAS presso LHC	3
1.1 Il Large Hadron Collider	3
1.2 Il rivelatore ATLAS	3
1.3 Processi di nuova fisica con fotoni dark	5
2 Il multilayer perceptron	7
2.1 Reti neurali artificiali	7
2.2 Il multilayer perceptron	7
2.3 Criticità nell'utilizzo delle reti neurali	9
2.4 Performance	10
3 Addestramento di un MLP per distinguere eventi di materia oscura da eventi di fondo	12
3.1 Software utilizzato	12
3.2 Variabili di input	12
3.3 Risultati	15
3.3.1 Configurazione A	15
3.3.2 Configurazione B	18
3.3.3 Configurazione C	20
3.4 Conclusioni	22

Introduzione

La *fisica delle alte energie* o *fisica delle particelle* si occupa dello studio dei costituenti fondamentali della materia, indaga sulle particelle elementari e sulle leggi che ne governano le interazioni per scoprire le proprietà prime del nostro universo. Uno degli strumenti principali di un fisico sperimentale delle alte energie è il moderno acceleratore, dove vengono prodotte particelle esotiche grazie ad urti di elevata energia fra particelle; osservare le particelle prodotte negli urti e misurarne le proprietà permette di avere una visione più chiara della natura delle forze fondamentali.

La maggior parte degli urti non produce nuove particelle. Per esempio, nonostante il Large Hadron Collider (LHC) produca approssimativamente 10^{11} urti per ora in ciascun punto di interazione, solo 300 di questi in media dà luogo ad un bosone di Higgs [1]. Ciò implica che una buona analisi dei dati dipenda criticamente da un'efficace distinzione fra gli urti che producono particelle di interesse (segnale) e quelli che producono altre particelle (fondo). La selezione può risultare difficile perché gli eventi di fondo possono avere caratteristiche simili a quelli di segnale, per esempio potrebbero avere gli stessi prodotti di decadimento.

A tal scopo negli ultimi anni nell'ambito della fisica delle alte energie si sono cominciate ad usare con successo tecniche di *machine learning* in uso in campi del tutto diversi, come la classificazione delle immagini: riconoscere se è presente un volto in una foto è un problema concettualmente equivalente a distinguere un evento di segnale da uno di fondo.

Il problema preso in esame in questa dissertazione riguarda processi di nuova fisica in studio ad ATLAS presso il Large Hadron Collider, in particolare si vuole costruire un classificatore che distingua efficacemente possibili eventi di nuova fisica con fotoni dark da noti eventi di cromodinamica quantistica (QCD) previsti dal Modello standard.

Nel primo capitolo dopo aver brevemente presentato l'acceleratore LHC e la struttura del rivelatore ATLAS, saranno discussi i processi di nuova fisica oggetto dell'analisi.

Nel secondo capitolo sarà descritto il classificatore utilizzato, il *multilayer perceptron*, una rete neurale *feedforward*. Se ne discuterà la struttura, la teoria dietro al processo di apprendimento e le criticità che presenta il suo utilizzo.

Nel terzo capitolo, cuore della dissertazione, saranno presentate le variabili scelte per la classificazione; ottenute da simulazioni MonteCarlo per gli eventi di nuova fisica (segnale) e da dati reali raccolti ad ATLAS per gli eventi di QCD (fondo). Verranno infine discussi i risultati ottenuti con tre diverse configurazioni del *multilayer perceptron*.

Capitolo 1

Ricerca di fotoni dark con il rivelatore ATLAS presso LHC

1.1 Il Large Hadron Collider

LHC è il più potente acceleratore di particelle attualmente in funzione. Si sviluppa lungo un tunnel circolare fra Francia e Svizzera lungo 27 km e ad una profondità media di 100 m. Nell'anello principale vengono accelerati fasci di protoni in direzioni opposte; è presente un campo magnetico di circa 8 T generato da magneti superconduttori che mantiene i protoni sul tracciato contenuti in due tubi a vuoto. In queste condizioni viene raggiunta un'energia nel centro di massa di circa 13 TeV, vicina al limite teorico della macchina di 14 TeV. Gli urti avvengono in corrispondenza dei quattro rivelatori principali: ATLAS (A Toroidal LHC ApparatuS), CMS (Compact Muon Solenoid), LHCb (LHC-beauty) e ALICE (A Large Ion Collider Experiment).

1.2 Il rivelatore ATLAS

Prima di descrivere la struttura del rivelatore è utile definire il sistema di coordinate e alcune grandezze di frequente utilizzo.

1.2.1 Concetti fondamentali

Il sistema di riferimento utilizzato ha origine nel punto dove ha luogo l'interazione, la direzione del fascio coincide con l'asse z , l'asse y punta verso l'alto e l'asse x verso il centro dell'acceleratore; definiamo l'angolo polare θ e l'angolo azimutale ϕ . Generalmente al posto di θ si utilizza la pseudorapidità definita come:

$$\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right] \quad (1.1)$$

perché la differenza di pseudorapidità di due particelle è un invariante lungo z , la direzione parallela al fascio. Si ha dunque $\eta = 0$ lungo y e $\eta = \infty$ lungo z .

Un cono con vertice nell'origine avrà apertura $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$.

Il numero di eventi che avvengono per unità di tempo in un acceleratore si può scrivere come:

$$\frac{dN}{dt} = \sigma \mathcal{L} \quad (1.2)$$

dove σ è la sezione d'urto del processo in esame e corrisponde ad una misura della probabilità che il processo in questione avvenga a seguito di un'interazione, mentre \mathcal{L} è la luminosità istantanea ed è un parametro che quantifica le prestazioni di un acceleratore.

La luminosità istantanea è definita in questo modo:

$$\mathcal{L} = \frac{N_1 N_2 f_i}{A} \quad (1.3)$$

con N_1 e N_2 numero di protoni presenti nei pacchetti, f_i frequenza di collisione, che sarà pari alla frequenza di rotazione dei pacchetti nell'anello per il numero di pacchetti del fascio, mentre A è la sezione trasversale media dei fasci.

Spesso si utilizza anche la luminosità integrata \mathcal{L}_{int} , utile per rappresentare il numero di eventi ottenuti in un certo intervallo di tempo T . Si ha infatti:

$$N = \sigma \int_0^T \mathcal{L} dt = \sigma \mathcal{L}_{int} \quad (1.4)$$

1.2.2 L'apparato di rivelazione

ATLAS (A Toroidal LHC ApparatuS) è il più grande rivelatore di particelle mai realizzato, è lungo 45 metri, alto 25 metri e dal peso complessivo di circa 7000 tonnellate. È un rivelatore di indirizzo generale, non si concentra su un particolare processo fisico; è stato piuttosto progettato per rivelare tutti i possibili prodotti dell'interazione p-p, da cui la complessità dell'apparato. Il rivelatore si articola in vari cilindri concentrici attorno all'asse del fascio. I componenti fondamentali di ATLAS sono (vedi Fig. 1.2):

- Un rivelatore interno (ID), il più vicino al tubo a vuoto in cui avvengono gli urti, composto da tre rivelatori di traccia di precisione, uno realizzato con pixel di silicio, un altro con microstrisce di silicio che copre il range di pseudorapidità $\eta < 2.5$, mentre l'ultimo è un rivelatore a radiazione di transizione. Un campo magnetico diretto lungo l'asse dei fasci fa curvare le particelle cariche, la direzione della curva e il grado di curvatura permettono di misurare carica e quantità di moto.
- Un calorimetro elettromagnetico (EMCAL) dello spessore di circa 23 lunghezze di radiazione (X_0) che copre un range $\eta < 3.2$, composto da strati di argon liquido e piombo, preceduti da un pre-campionatore che ha la funzione di misurare con precisione il punto di impatto delle particelle sul calorimetro e facilitare l'interazione di queste.
- Un calorimetro adronico (HCAL) dello spessore di circa 10 lunghezze di radiazione (λ) che copre un range $\eta < 4.9$.

- Uno spettrometro muonico (MS) che copre il range $\eta < 2.7$; i muoni infatti, essendo particelle al minimo di ionizzazione, depositano solo una piccola parte di energia nei calorimetri, e richiedono un rivelatore più esterno che ne misuri la quantità di moto mediante la deflessione indotta da un campo magnetico per lo più ortogonale alla traiettoria dei muoni, e che quindi grazie alla forza di Lorentz cambia solo la direzione del moto e non la velocità delle particelle. Le coordinate della traiettoria dei muoni vengono quindi misurate per quasi tutto il range di pseudorapidità da *tubi a drift*.
- Un sistema di magneti toroidali superconduttori che fornisce il campo magnetico necessario allo spettrometro muonico.

Avvenendo ogni 25 ns l'incrocio fra due pacchetti, circostanza che dà luogo in media a 30 urti simultanei, al fine di limitare la memorizzazione di dati inutili, ATLAS è inoltre dotato di un complesso sistema di trigger che seleziona in tempo reale gli eventi interessanti. Questo è articolato su tre livelli: L1, che ricerca rapidamente particelle con una grande componente trasversale della quantità di moto, L2 che opera una selezione ricevendo informazioni da tutti i detector di ATLAS, ed un filtro eventi (EF) la cui analisi avviene dopo la ricostruzione completa dell'evento, e la cui approvazione determina la scrittura in memoria.

1.3 Processi di nuova fisica con fotoni dark

Tra le possibili estensioni del modello standard che prevedano l'esistenza della materia oscura, molte predicono l'esistenza di un settore dark che sia debolmente accoppiato con quello visibile. A seconda della struttura di questo settore dark e del suo accoppiamento con il modello standard, alcuni stati dark instabili potrebbero essere prodotti negli acceleratori e decadere in particelle del modello standard con un rapporto di ramificazione (BR) consistente.

1.3.1 Il caso dei fotoni dark

Un caso ampiamente studiato è quello in cui i due settori sono accoppiati da un cosiddetto portale vettoriale, dove un fotone dark (γ_d) con massa equivalente compresa fra il MeV e il GeV si accoppia ad un fotone standard. Se il fotone dark fosse lo stato più leggero nel settore dark, non potrà che decadere in particelle del modello standard, principalmente leptoni ma anche mesoni leggeri. A causa della sua debole interazione con il modello standard il fotone dark può avere un tempo di vita medio non trascurabile.

Ad LHC questi fotoni dark sarebbero tipicamente prodotti con un elevato impulso, a causa della loro ridotta massa, generando strutture simili a jet molto collimati contenenti coppie di leptoni o coppie di adroni leggeri (Jet di nuova fisica, LJ). A causa della vita media non trascurabile questi LJ potrebbero avere origine lontano dal punto di interazione (ordine del metro).

La versatilità e l'elevata risoluzione del rivelatore di ATLAS lo rendono ideale per questo tipo di ricerca. Inoltre il sistema di tracciamento interno di ATLAS può essere utilizzato per stabilire criteri basati sulle tracce utili per ridurre in maniera significativa l'altrimenti smisurato fondo dato da eventi del modello standard.

1.3.2 Eventi considerati

La ricerca dei Jet di nuova fisica di cui si occuperà questa dissertazione impiega dati collezionati da ATLAS durante il 2015 in cui LHC ha funzionato ad una energia nel centro di massa di $\sqrt{s} = 13$ TeV, ed ha collezionato una luminosità integrata $\mathcal{L}_{int} = 3.4 \text{ fb}^{-1}$. In particolare saranno presi in considerazione LJ dovuti a fotoni dark che decadono in adroni o elettroni, cioè quelli in cui si hanno solamente jet di elettroni e pioni.

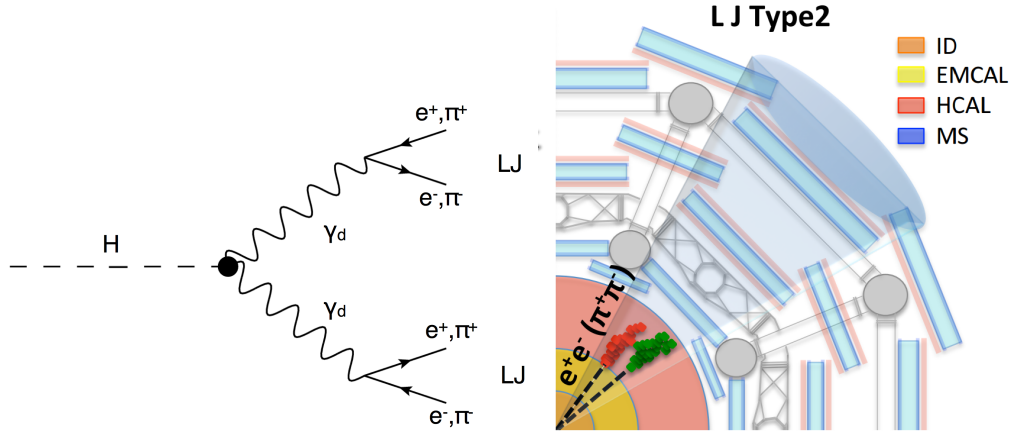


Figura 1.1. Il diagramma di Feynman del modello considerato in questa dissertazione. L'Higgs dark decade in due fotoni dark che danno origine ai LJ.

Figura 1.2. Rappresentazione schematica di un LJ adronico in ATLAS.

È possibile ora descrivere nel dettaglio gli eventi di segnale e fondo oggetto della classificazione. Gli eventi di segnale sono eventi da simulazione Monte Carlo di fotoni dark con massa equivalente di 400 MeV che generano LJ adronici composti completamente da pioni:

$$p + p \rightarrow H \rightarrow \gamma_d + \gamma_d \rightarrow (\pi_+ + \pi_-) + (\pi_+ + \pi_-) \quad (1.5)$$

Mentre come fondo è stato utilizzato un campione di dati reali collezionati da ATLAS di multi-jet di QCD in cui il processo dominante è:

$$p + p \rightarrow q + q \rightarrow 2 \text{ jet adronici} \quad (1.6)$$

I prodotti di decadimento di questi processi sono dunque molto simili. Presentano tuttavia caratteristiche differenti: i pioni di fondo sono meno collimati di quelli da fotoni dark e soprattutto sono prodotti esclusivamente nel punto dove avviene l'urto dei protoni, il vertice primario di interazione.

Tali caratteristiche saranno sfruttate nel seguito per addestrare un algoritmo multivariato al fine di separare in modo ottimale i due processi fisici.

Capitolo 2

Il multilayer perceptron

2.1 Reti neurali artificiali

Generalizzando il più possibile, con rete neurale artificiale ci si riferisce ad una qualsiasi collezione di neuroni interconnessi in cui ogni neurone produce una certa risposta ad un certo numero di segnali in input. Applicando un segnale esterno ad alcuni neuroni, detti di input, la rete assume un determinato stato che può essere misurato dalla risposta di uno o più neuroni, detti di output. Si può dunque vedere la rete come una funzione che mappa uno spazio di n dimensioni, dove n è il numero di variabili in input x_1, \dots, x_n (*features*), in uno spazio di m dimensioni, dove m è il numero di variabili in output y_1, \dots, y_m ; nel caso della discriminazione fra segnale e fondo lo spazio di output sarà unidimensionale e $m = 1$. La mappatura sarà non lineare se almeno un neurone ha una risposta non lineare nelle sue variabili di input, come nel caso del *multilayer perceptron*.

Il comportamento della rete neurale è quindi determinato dalla disposizione dei neuroni, dai pesi degli archi e dalla loro risposta all'input, descritta dalla *funzione di risposta del neurone* ρ .

2.2 Il multilayer perceptron

In teoria una rete neurale con n neuroni può avere n^2 archi orientati che connettono tra loro i vari nodi, tale complessità può essere ridotta organizzando i neuroni in strati e permettendo solo connessioni che siano dirette da uno strato verso il successivo, tale proprietà è detta *feed-forward*. Se poi ad eccezione di quelli di input e output tutti gli altri neuroni presentano una ρ non lineare allora la rete è del tipo *multilayer perceptron*. Il primo strato è lo strato di input mentre l'ultimo è quello di output, tutti gli altri sono strati nascosti (vedi Fig. 2.1).

2.2.1 Funzione di risposta del neurone ρ

La funzione di risposta del neurone $\rho : \mathbf{R}^n \rightarrow \mathbf{R}$ mappa gli input i_0, i_1, \dots, i_n nell'output del neurone y_j^s (j -esimo neurone dello strato s). Detto w_{kj}^{s-1} il peso dell'arco fra il neurone k dello strato $s-1$ e il neurone j dello strato s , gli input i_h sono definiti come il segnale mandato sull'arco dal neurone dello strato precedente moltiplicato per il peso dell'arco, in formule:

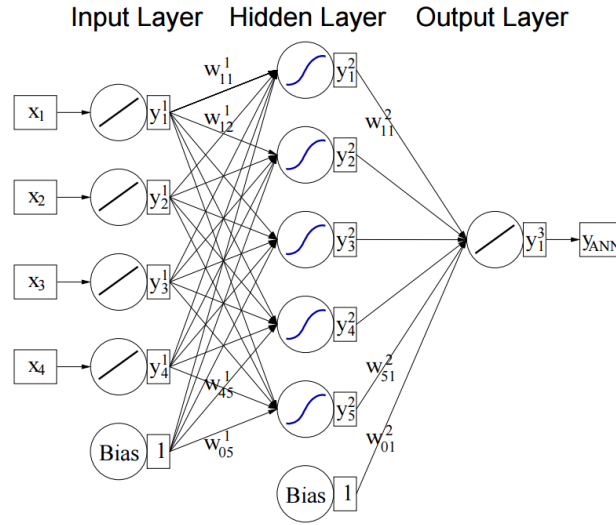


Figura 2.1. Multilayer perceptron con un solo strato nascosto; y_{ANN} è l'output, in questo caso unidimensionale, della rete; mentre x_1, x_2, x_3 e x_4 sono le variabili di input. All'interno di ciascun neurone è rappresentata la relativa *funzione di attivazione* α .

$$i_0 = w_{0j}^{s-1}; \quad i_h = w_{hj}^{s-1} y_h^{s-1} \quad \text{con } h = 1, \dots, n \quad (2.1)$$

Nel caso più comune, che corrisponde a quello utilizzato in questa dissertazione, il neurone fa una semplice somma degli input $x = i_0 + i_1 + \dots + i_n$ e trasforma x secondo una certa *funzione di attivazione* α che sarà lineare per i neuroni in input e output e non lineare per gli altri. Le funzioni non lineari più utilizzate sono la tangente iperbolica e la sigmoide:

$$\alpha(x) = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad \text{o} \quad \alpha(x) = \frac{1}{1 + e^{-x}} \quad (2.2)$$

2.2.2 Allenamento della rete neurale

L'algoritmo più utilizzato per tarare i pesi degli archi in modo che ottimizzino la classificazione della rete neurale, cioè la probabilità di assegnare la classe corretta ad un dato evento di input, è la cosiddetta *back-propagation* (BP). Appartiene alla famiglia dei metodi di apprendimento supervisionato: la rete è allenata con un set di dati di cui si conosce a priori la classificazione corretta. Sia \mathcal{B} tale insieme composto da U elementi, allora $\mathcal{B} = ((\mathbf{x}_1, \hat{y}_1), \dots, (\mathbf{x}_\alpha, \hat{y}_\alpha), \dots, (\mathbf{x}_U, \hat{y}_U))$ con $\mathbf{x}_\alpha = (x_1, \dots, x_{n_I})_\alpha$ evento etichettato $\hat{y}_\alpha \in \{0, 1\}$ (1 per gli eventi di segnale e 0 per quelli di fondo).

Per l'analisi di cui si occupa questa dissertazione è stato usato sia l'algoritmo di BP semplice che una sua evoluzione detta BFGS (Broyden-Fletcher-Goldfarb-Shannon), qui verrà esposto l'algoritmo di BP semplice, più comune e concettualmente più limpido.

L'output di un *multilayer perceptron*, qui per semplicità con un singolo strato nascosto, output unidimensionale e con funzione di attivazione $\alpha(x) = \tanh(x)$ (usare come riferimento la figura 2.1), è dato da:

$$y_{\text{ANN}} = \sum_{h=1}^{n_N} y_h^{s=2} w_{h1}^{s=2} = \sum_{h=1}^{n_N} \tanh\left(\sum_{l=1}^{n_I} x_l w_{lh}^{s=1}\right) w_{h1}^{s=2} \quad (2.3)$$

dove n_N e n_I sono rispettivamente il numero di neuroni sullo strato nascosto e su quello di input.

Si prendano a caso T elementi $(\mathbf{x}_a, \hat{y}_a)$ da \mathcal{B} , solitamente $T = \frac{U}{2}$, questo nuovo insieme è detto *training set*. Confronteremo ora l'output della rete neurale di ciascuno di questi eventi $(y_{\text{ANN},a})$ con l'output vero (\hat{y}_a) .

Costruiamo dunque una *funzione di errore* E così definita:

$$E(\mathbf{x}_1, \dots, \mathbf{x}_T, \mathbf{w}) = \sum_{a=1}^T E_a(\mathbf{x}_a, \mathbf{w}) = \sum_{a=1}^T \frac{1}{2} (y_{\text{ANN},a} - \hat{y}_a)^2 \quad (2.4)$$

dove \mathbf{w} è l'insieme di tutti i pesi regolabili della rete. Tanto più l'output della rete corrisponde all'output vero negli eventi del *training set*, tanto più la funzione di errore si avvicina a zero. Il set di pesi che minimizza la funzione di errore può essere trovato usando il metodo di discesa del gradiente, dato che la funzione di risposta del neurone ρ è differenziabile rispetto alle sue variabili di input. Assegnando inizialmente ad ogni peso un numero casuale $\mathbf{w}^{(\beta)}$, questi sono aggiornati compiendo piccoli passi nello spazio dei \mathbf{w} nella direzione $-\nabla_{\mathbf{w}} E$, cioè dove E decresce più rapidamente:

$$\mathbf{w}^{(\beta+1)} = \mathbf{w}^{(\beta)} - \eta \nabla_{\mathbf{w}} E \quad (2.5)$$

il numero positivo η determina la lunghezza del passo ed è detto *learning rate*.

L'allenamento termina quando il gradiente è nullo, cioè si è raggiunto un punto di minimo per E , oppure dopo un numero stabilito di passi, come nella maggior parte dei casi.

2.3 Criticità nell'utilizzo delle reti neurali

Quando si sceglie di utilizzare una rete neurale, o un altro classificatore, ciò che si cerca di ottenere è un modello che descriva il fenomeno dal quale emerge la differenza fra le classi in esame. Non si tratta di un semplice *fit* in cui la forma funzionale, il modello, è conosciuto a priori; è piuttosto un processo induttivo, perché per stabilire il modello si considerano solo un numero limitato di casi particolari, il *training set*.

L'unica cosa che è necessario conoscere a priori per utilizzare una rete neurale è la *complessità* del modello cercato.¹ A seconda della complessità attesa si utilizzerà una rete neurale con più o meno neuroni.

Si può dimostrare che una rete neurale del tipo del *multilayer perceptron* può approssimare con precisione arbitraria qualsiasi funzione $f: \mathbf{R}^n \rightarrow \mathbf{R}^m$ se dotata di un numero sufficiente di neuroni, anche con un singolo strato nascosto [4].

Se la rete è troppo semplice rispetto al modello cercato non saranno presi in considerazione tutti gli aspetti del fenomeno e non si otterrà un buon risultato, si parla in questo caso di *underfitting*.

¹Se si utilizza un regolatore (vedi fine sezione) è possibile avere una conoscenza meno precisa della complessità del modello.

Se al contrario la rete è troppo complessa rispetto al modello cercato, data la natura limitata e imperfetta del *training set*, questa potrebbe sviluppare un modello più complesso che comprenda il rumore o le fluttuazioni statistiche degli eventi del *training set*, si parla in questo caso di *overfitting*.

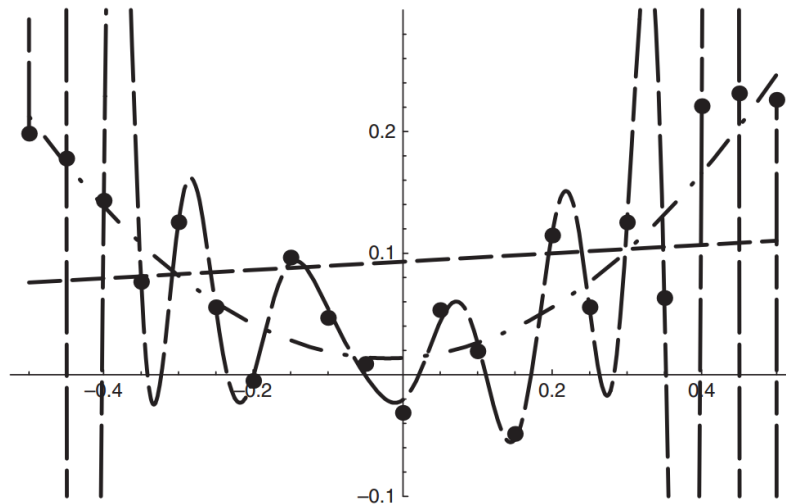


Figura 2.2. Esempio di underfitting e overfitting su un campione rumoroso. Il modello che descrive i dati è una parabola, 3 parametri da stimare, il modello lineare è troppo semplice, 2 parametri, mentre quello del polinomio di grado alto troppo complesso, $n \gg 3$ parametri.

Per completezza si riporta che una delle tecniche principali per contrastare l'*overfitting*, oltre alla semplificazione della rete, è l'utilizzo di un regolatore, il cui effetto è "smussare" le ipersuperfici che separano le classi nello spazio a n_I dimensioni delle variabili di input, ma la cui teoria va oltre lo scopo di questa dissertazione.

2.4 Performance

Per misurare le prestazioni della rete e controllare la presenza di un eventuale *overfitting* si utilizzano gli elementi di \mathcal{B} che non appartengono al *training set*, questo nuovo insieme viene chiamato *test set*.

Un'indicazione immediata sulla bontà della classificazione si può ottenere dalla *curva ROC*, che rappresenta per ogni possibile soglia su y_{ANN} il falso negativo (*background rejection*) e il vero positivo (*signal efficiency*) sugli eventi del *test set*. Una soglia per cui queste due quantità sono uguali a 1 garantisce una classificazione perfetta, in generale più la *curva ROC* è vicina a questo punto più la classificazione è buona (esempio in figura (3.6)).

Per avere un'idea ancora più chiara della qualità della classificazione si può osservare la distribuzione dell'output della rete neurale (y_{ANN}) per gli eventi di segnale e fondo: minore l'area comune alle due distribuzioni, migliore la classificazione. Se le distribuzioni vengono ulteriormente suddivise fra eventi del *training set* e del *test set* allora è possibile verificare se c'è stato un *overfitting*, questo si ha se le

distribuzioni di segnale e fondo del *training set* non coincidono entro l'errore con quelle del *test set* (esempio in figura (3.11)).

Per sapere in tempo reale durante l'allenamento se ci si trova in *overfitting* si può confrontare l'errore E sul *training set* con quello sul *test set*; se quello sul *test set* comincia a crescere è probabile che la rete stia cominciando a modellizzare rumore e fluttuazioni statistiche del *training set* (esempio in figura (3.10)).

Capitolo 3

Addestramento di un MLP per distinguere eventi di materia oscura da eventi di fondo

3.1 Software utilizzato

Il software utilizzato per l'analisi è TMVA (Toolkit for Multivariate Data Analysis) un toolkit per l'analisi multivariata interno a ROOT, un pacchetto software di analisi dei dati sviluppato dal CERN.

TMVA implementa tre diversi tipi di rete neurale fra cui il *multilayer perceptron* (MLP), i principali parametri da controllare sono: le variabili da considerare in input, il numero di strati e di neuroni su ogni strato, la funzione di risposta del neurone ρ , il numero di cicli di training, il tipo di algoritmo di apprendimento, l'uso o meno di un regolatore, il *learning rate* η ed infine il *decay rate* se si vuole $\eta \rightarrow 0$ al crescere del numero di cicli di apprendimento effettuati.

Al termine del training TMVA stila inoltre una classifica delle variabili più *importanti* considerando la somma dei pesi al quadrato degli archi uscenti dai neuroni di input. L'importanza I_i di una variabile di input i è data quindi da:

$$I_i = \bar{x}_i^2 \sum_{j=1}^{n_N} (w_{ij}^{s=1})^2 \quad \text{con } i = 1, \dots, n_I \quad (3.1)$$

Dove \bar{x}_i è la media della variabile di input i sui dati del campione analizzato.

3.2 Variabili di input

Il sistema di riferimento utilizzato e le grandezze menzionate sono descritti nella sezione (1.2.1), mentre gli eventi presi in considerazione sono trattati nella sezione (1.3.2), ogni evento può generare 2 o più jet.

Le variabili di input, o *features*, scelte per allenare la rete neurale sono:

- **ptL**: quantità di moto trasversa, cioè la quantità di moto proiettata sul piano $x - y$, del jet più energetico (*leading*).

- **ptF**: quantità di moto trasversa del jet più lontano angularmente da quello *leading*.
- **sumptL**: somma delle quantità di moto trasverse di tutte le tracce cariche ricostruite in un cono di apertura $\Delta R = 0.4$ centrato intorno alla direzione del jet *leading*.
- **sumptF**: analogo a *sumptL* con cono centrato intorno alla direzione del jet più lontano angularmente da quello *leading*.
- **deltaPhi**: distanza in angolo azimutale ($\Delta\phi$) fra il jet *leading* e quello angularmente più lontano.
- **sumptMax**: massimo fra *sumptL* e *sumptF*.
- **timeL**: intervallo di tempo fra l'interazione dei protoni e l'arrivo nei calorimetri (HCAL) del jet *leading*.
- **timeF**: analogo a *timeL* per il jet più lontano angularmente da quello *leading*.
- **emfrF**: frazione di energia rilasciata nel calorimetro elettromagnetico (EM-CAL) dal jet più lontano angularmente da quello *leading*.
- **wF**: larghezza trasversale del jet più lontano angularmente da quello *leading*.
- **jvtF**: frazione di tracce cariche appartenenti al jet che vengono associate al vertice primario di interazione; più il jet è prodotto vicino al vertice primario maggiore sarà la frazione di tracce associate dall'apparato di rivelazione al vertice.

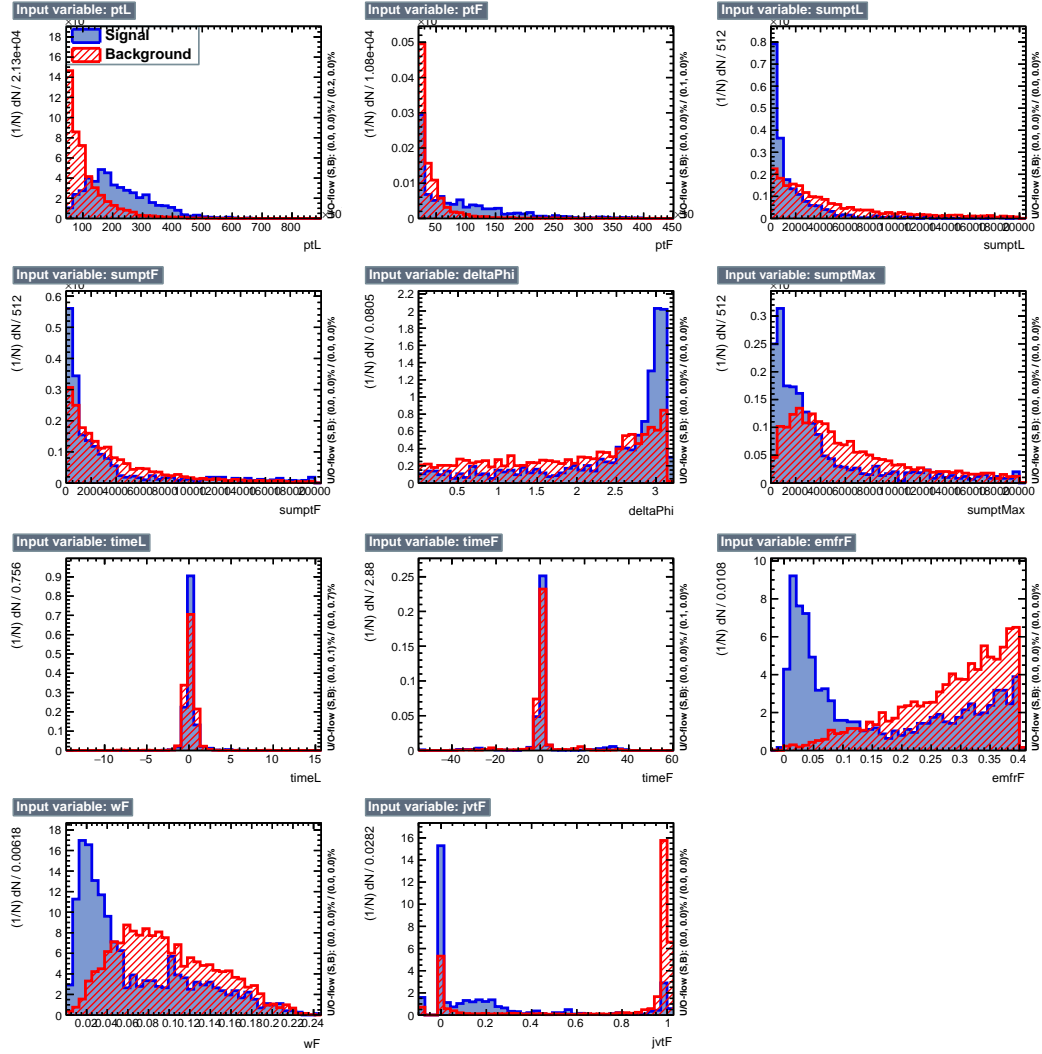


Figura 3.1. Istogrammi delle frequenze delle variabili di input per gli eventi di segnale e fondo dei dati a disposizione.

Con riferimento agli istogrammi in figura (3.1) e alle matrici di correlazione in figura (3.2) seguono alcune osservazioni sui dati a disposizione.

Come aspettato le variabili ptL e ptF hanno un buon potere separatore, infatti la ridotta massa equivalente dei fotoni dark (400 MeV) rispetto all'impulso fa sì che i pioni generati siano molto collimati, più di quelli da eventi di fondo.

Anche le variabili $sumptL$, $sumptF$ e $sumptMax$ garantiscono una buona separazione, ciò è dovuto al punto di origine dei jet: poiché la maggior parte dei jet da fotoni dark è generato lontano dal punto di interazione è più difficile ricostruire le tracce cariche nell'ID, dunque per gli eventi di segnale questa grandezza tenderà ad assumere valori più piccoli.

Le distribuzione di $deltaPhi$ degli eventi di segnale mostra che quando vengono prodotti fotoni dark questi sono prodotti a volte in configurazione *back to back*, generando coppie di jet con $\Delta\phi = 180^\circ$.

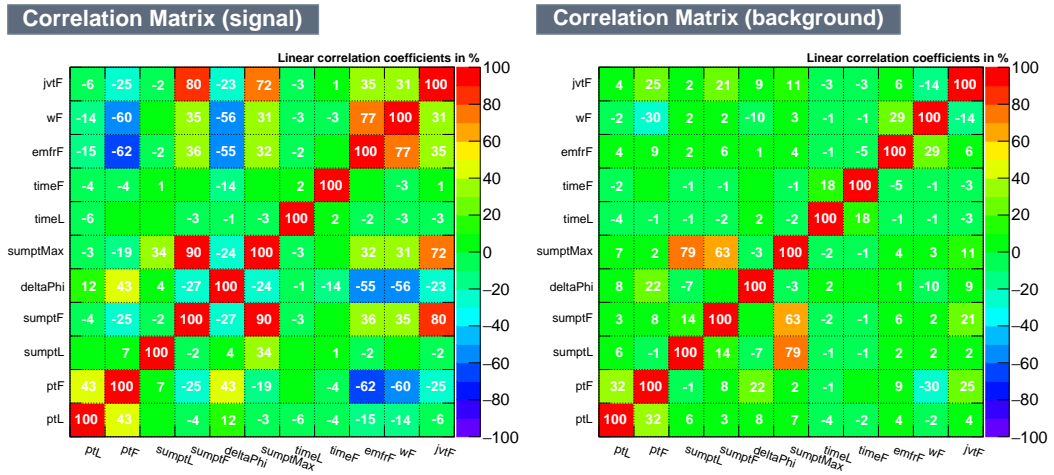


Figura 3.2. Matrici di correlazione delle variabili di input per gli eventi di segnale e di fondo

Per finire anche le variabili $emfrF$, wF e $jvtF$ offrono un ottimo potere separatore. Sono tutte sensibili alla distanza fra il punto di interazione e l'origine del jet, infatti rispettivamente un fotone dark che decade lontano: rilascerà meno energia nel calorimetro elettromagnetico, farà sì che il jet abbia meno tempo per allargarsi, renderà meno probabile l'associazione di tracce cariche del jet al vertice primario.

Come intuibile alcune di queste variabili risultano correlate perché sensibili allo stesso aspetto del fenomeno, per esempio la lontananza del punto di origine del jet dal vertice primario di interazione, o semplicemente perché funzione di altri variabili di input ($sumptMax$).

3.3 Risultati

La rete è stata allenata più volte con diverse configurazioni dei parametri salvando in tutto 15 configurazioni significative. Di seguito sono riportate quella che restituisce la migliore classificazione, una con un limitato numero di *features* ma che garantisce comunque buoni risultati ed infine una con più strati in cui è evidente la presenza di *overfitting*.

3.3.1 Configurazione A: 11 *features* e singolo strato nascosto

In questa configurazione sono state utilizzate tutte e 11 le *features* disponibili nei dati, una rete composta da un singolo strato nascosto di 14 neuroni con funzione di attivazione α di tipo sigmoide, BFGS come algoritmo di allenamento, un *learning rate* $\eta = 0.02$ e regolatore attivo.

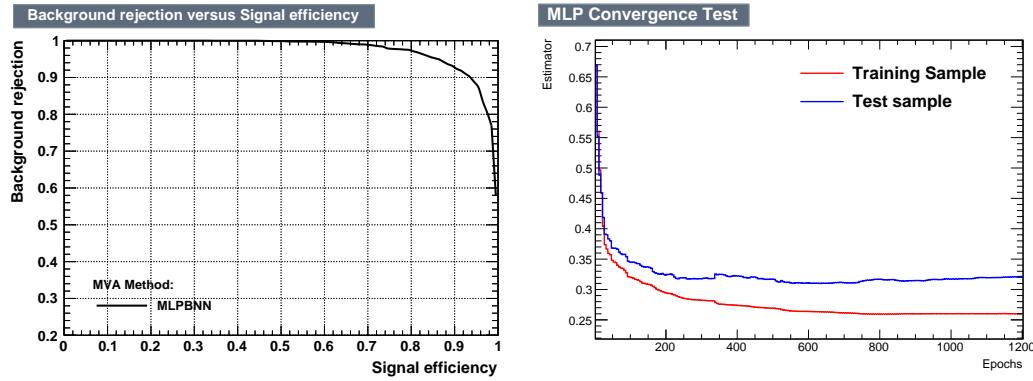


Figura 3.3. Curva ROC e andamento di E per *training set* e *test set* nella configurazione A.

La configurazione in esame è risultata essere la migliore fra quelle testate. Dal profilo della curva ROC in figura (3.3) si può vedere che il classificatore individua come segnale più del 90% degli eventi da fotoni dark ad una reiezione del 94% di eventi di QCD. La distinzione netta fra le due classi è evidente anche nella distribuzione dell'output della rete in figura (3.4a). Osservando lo stesso grafico in scala logaritmica (3.4b) è possibile accertarsi della presenza o meno di *overfitting*, in questo caso evidentemente non presente data la sovrapponibilità entro l'errore fra le distribuzioni degli eventi di *training set* e *test set*.

La classifica delle variabili ordinate per importanza è:

Rank	Variable	Importance
1	ptF	6.51E+01
2	ptL	4.58E+01
3	jvtF	2.09E+01
4	sumptF	5.18E+00
5	deltaPhi	3.06E+00
6	sumptL	2.79E+00
7	emfrF	2.18E+00
8	timeF	3.97E-01
9	wF	3.96E-02
10	sumptMax	1.15E-03
11	timeL	6.67E-05

Come aspettato ai primi posti sono presenti le variabili per cui si riscontra maggiore differenza fra eventi di segnale e di fondo (vedi Fig. 3.1) mentre agli ultimi oltre a variabili meno discriminanti (*timeL*, *timeF*) si trovano anche variabili correlate con altre più in alto in classifica (*sumptMax*, *wf*, vedi Fig. 3.2).

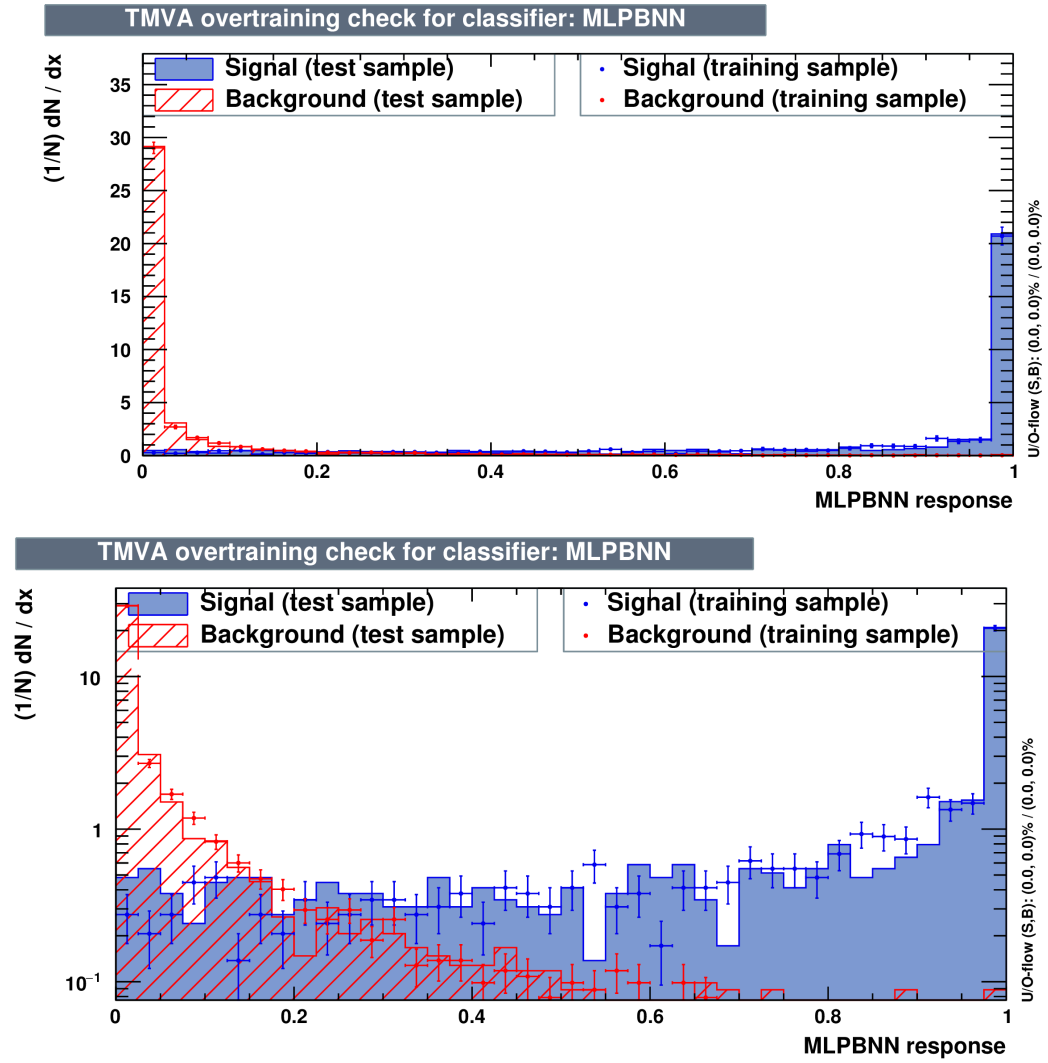


Figura 3.4. Distribuzioni dell'output della rete neurale della configurazione A in scala lineare e log.

3.3.2 Configurazione B: 7 *features* e singolo strato nascosto

In questa configurazione sono state utilizzate solo 7 delle *features* disponibili nei dati, eliminando variabili con un ridotto potere di discriminazione o correlate con altre, sono state rimosse in particolare *sumptL*, *timeF*, *timeL*, *deltaPhi*. La rete è composta da un singolo strato nascosto di 12 neuroni con funzione di attivazione $\alpha(x) = \tanh(x)$, BFGS come algoritmo di allenamento, un *learning rate* $\eta = 0.02$ e regolatore attivo.

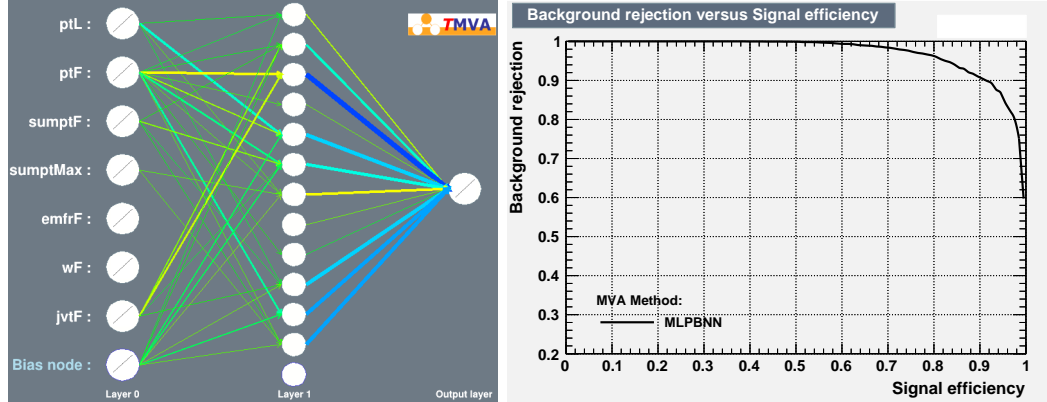


Figura 3.5. Struttura della rete neurale della configurazione B. Lo spessore delle linee è proporzionale al peso dell'arco.

Figura 3.6. Curva ROC per la configurazione B.

Nonostante il ridotto numero di variabili si può vedere dal profilo della curva ROC in figura (3.6) che il classificatore individua come segnale il 90% degli eventi da fotoni dark ad una reiezione del 90% di eventi di QCD. La distinzione netta fra le due classi è evidente anche nella distribuzione dell'output della rete in figura (3.7a). Dall'osservazione dello stesso grafico in scala logaritmica (3.7b) si può concludere che non si riscontra *overfitting*.

La classifica delle variabili ordinate per importanza è:

Rank	Variable	Importance
1	ptF	3.99E+01
2	ptL	1.51E+01
3	jvtF	8.07E+00
4	sumptF	3.44E+00
5	sumptMax	4.95E-01
6	emfrF	2.54E-01
7	wF	1.24E-03

La classifica non è molto differente da quella della configurazione A. Come aspettato ai primi posti sono presenti le variabili per cui si riscontra maggiore differenza fra eventi di segnale e di fondo mentre agli ultimi dopo aver eliminato le variabili meno discriminanti restano variabili correlate con altre più in alto in classifica (*emfrF*, *wf*).

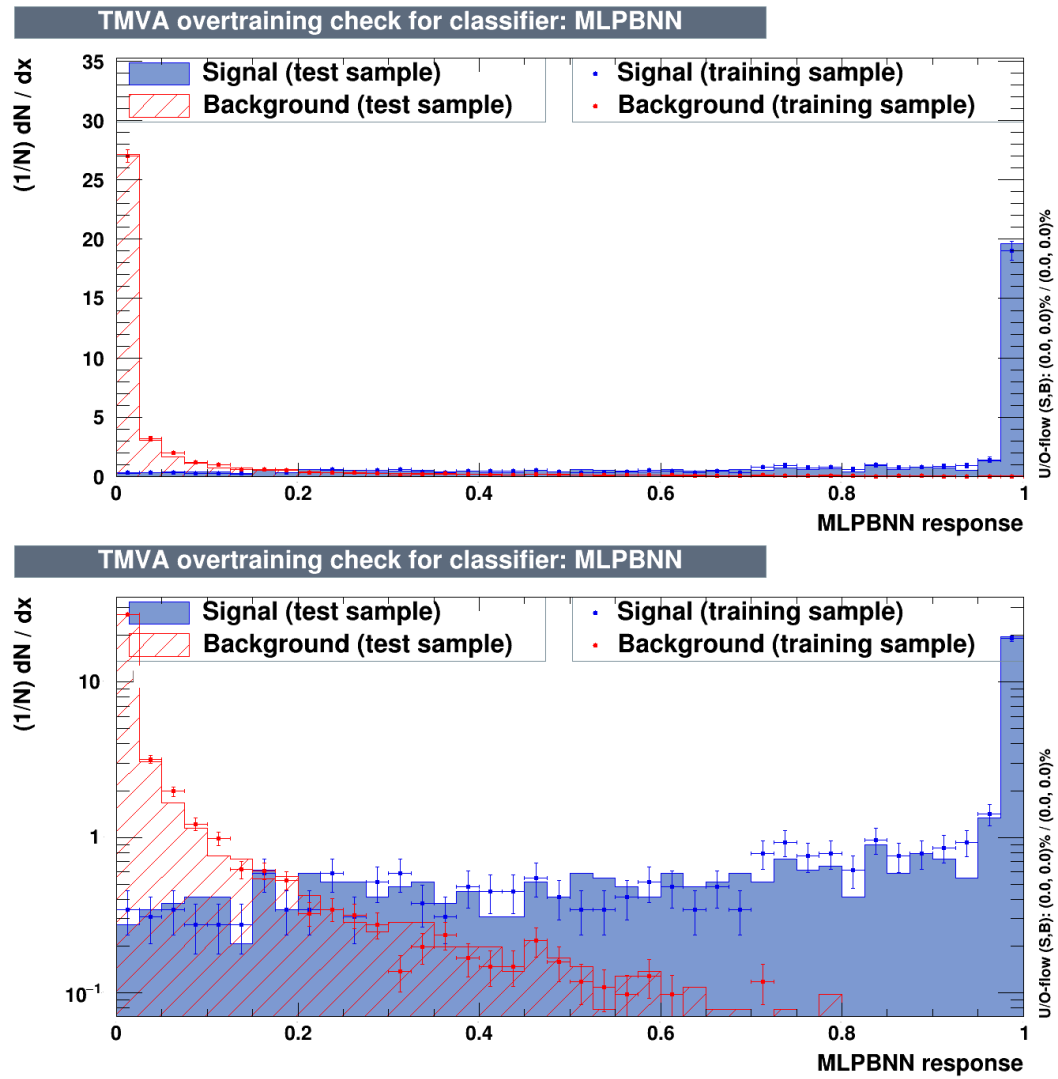


Figura 3.7. Distribuzioni dell'output della rete neurale della configurazione B in scala lineare e log.

Nonostante il ridotto numero di *features* i risultati ottenuti sono soddisfacenti, inoltre il tempo necessario per l'addestramento di questa rete è circa la metà di quello impiegato per addestrare la rete della configurazione A.

3.3.3 Configurazione C 11 *features* e più strati nascosti

In questa configurazione sono state utilizzate tutte e 11 le *features* disponibili nei dati. La rete è composta da 4 strati nascosti, 12 neuroni nei primi tre e 4 l'ultimo. La funzione di attivazione α è di tipo sigmoide, è stato utilizzato BP come algoritmo di allenamento e un *learning rate* $\eta = 0.02$ con regolatore disattivato.

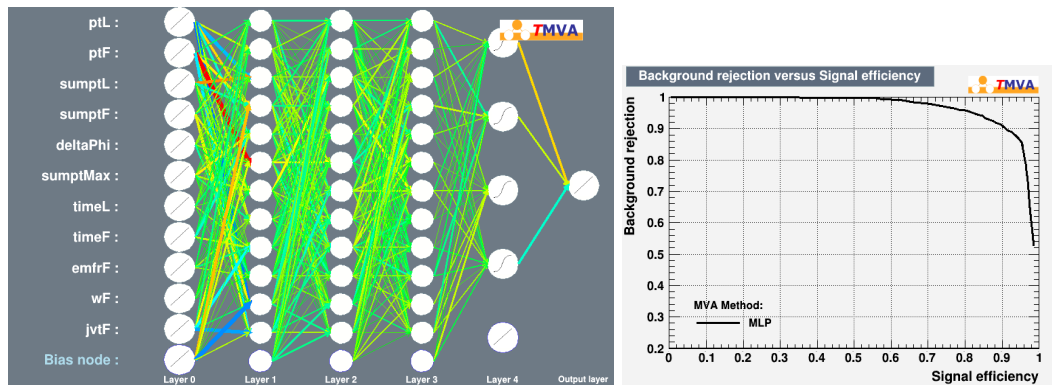


Figura 3.8. Struttura della rete neurale della configurazione C. Lo spessore delle linee è proporzionale al peso dell'arco.

Figura 3.9. Curva ROC per la configurazione C.

Nonostante l'elevata complessità della rete le prestazioni del classificatore sono inferiori a quelle della configurazione A e pressoché equivalenti a quelle della configurazione B come si può vedere dal profilo della curva ROC in figura (3.9). Ciò è dovuto a due principali ragioni: il semplice algoritmo di BP implementato in TMVA non è adeguato per questo tipo di reti per cui servono algoritmi di allenamento più complessi e più esosi; lo studio di questi algoritmi e l'utilizzo di reti complesse a più strati concerne la branca del *machine learning* detta *deep learning*. La seconda motivazione riguarda la sovrastima della complessità del modello necessario a distinguere gli eventi di segnale e fondo in questione, in particolare è evidente la presenza di *overfitting*. Si può osservare a tal proposito l'andamento di E per il *test set* che presenta evidentemente un minimo (Fig. 3.10) e la non sovrapponibilità entro l'errore delle distribuzioni del *training set* e del *test set* (Fig. 3.11).

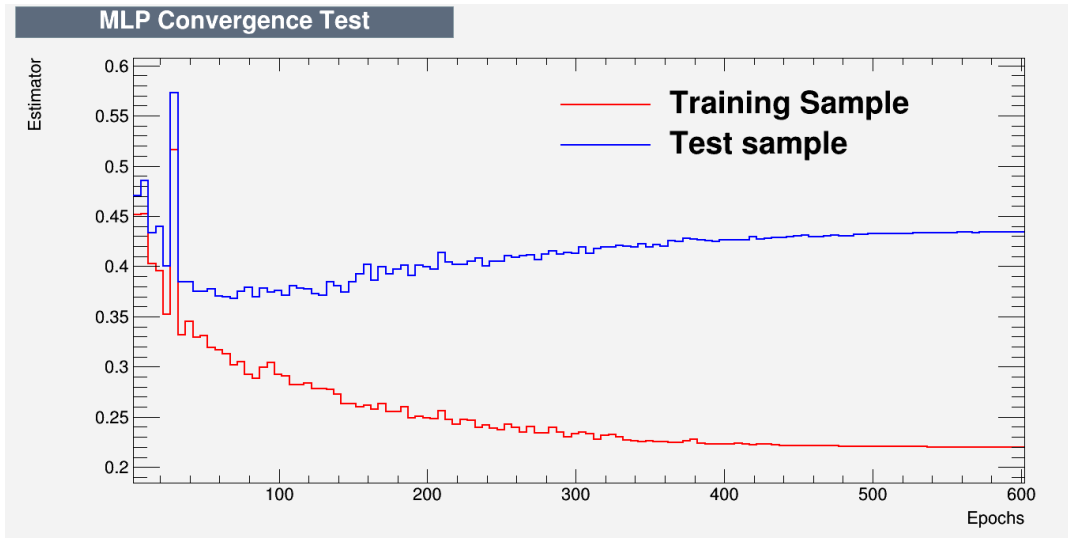


Figura 3.10. Andamento di E per *training set* e *test set* per la configurazione C.

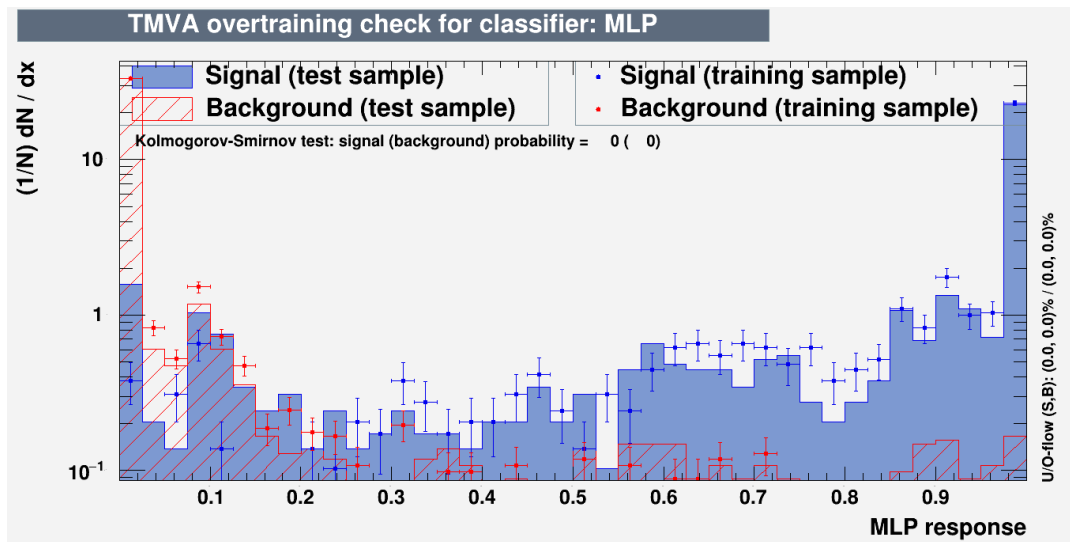


Figura 3.11. Distribuzioni dell'output della rete neurale della configurazione C in scala log.

3.4 Conclusioni

Il problema della discriminazione fra eventi da LJ adronici generati da fotoni dark e più comuni eventi di QCD è risolto con successo da una rete neurale a singolo strato del tipo del *multilayer perceptron*; in particolare se le variabili misurate coincidono con quelle utilizzate in questa dissertazione (sez. 3.2) la configurazione che garantisce la miglior classificazione è la A (sez. 3.3.1).

Algoritmi di questo tipo sono attualmente in uso nelle analisi di ricerca ad ATLAS e quello presentato in questa dissertazione potrebbe essere utilizzato per la ricerca di eventi da fotoni dark.

Le reti neurali del tipo del *multilayer perceptron* si dimostrano dunque ottimi classificatori, sono in grado di lavorare con dati rumorosi o correlati fra loro riducendo la necessità di un preventivo prefiltraggio da parte di un agente umano. Tale caratteristica è ancora più marcata per le reti *deep*, che cominciano ad essere utilizzate con successo anche nell'analisi multivariata in fisica delle alte energie [1].

Grazie allo studio teorico e applicativo del *machine learning* ed in particolare alle applicazioni legate all'uso di classificatori basati su reti neurali, la comprensione di tali strumenti e di come emergano le proprietà funzionali alla classificazione sta progredendo velocemente, rendendoli non più delle *black box*, e promuovendone l'utilizzo in numerosi campi della fisica.

Bibliografia

- [1] Baldi, Pierre, Peter Sadowski, and Daniel Whiteson. "Searching for exotic particles in high-energy physics with deep learning." *Nature communications* 5 (2014).
- [2] Whiteson, Shimon, and Daniel Whiteson. "Machine learning for event selection in high energy physics." *Engineering Applications of Artificial Intelligence* 22.8 (2009): 1203-1217.
- [3] Webb, Andrew R. "Statistical pattern recognition." John Wiley and Sons, 2003.
- [4] Hornik, Kurt, Maxwell Stinchcombe, and Halbert White. "Multilayer feedforward networks are universal approximators." *Neural networks* 2.5 (1989): 359-366.
- [5] ATLAS Collaboration. "Search for long-lived neutral particles decaying into displaced lepton jets in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector." *ATLAS-CONF-2016-042* (2016).
- [6] ATLAS collaboration. "The ATLAS experiment at the CERN large hadron collider." *Journal of Instrumentation* 3.S08003 (2008).
- [7] Hoecker, Andreas, et al. "TMVA-Toolkit for multivariate data analysis." *arXiv preprint physics/0703039* (2007).