**Esercizio 7**

“Si generino due classi di eventi: la prima classe chiamata segnale con pdf gaussiana bidimensionale centrata in (x,y=0,0) e σx=σy=0.3, ρ=0.5, la seconda classe chiamata fondo con pdf gaussiana bidimensionale centrata in (x,y)=(4,4) e σx=σy=1, ρ=0.4. Si utilizzi uno o più metodi MVA illustrati a lezione per separare le due classi di eventi e si caratterizzi il risultato in termini di purezza del segnale e rigezione del fondo.”

**Introduzione**

La separazione del segnale dal fondo è uno dei tipici problemi dell’analisi multivariata. Si supponga di avere due categorie di eventi H0 e H1. Si possono individuare delle osservabili che permettono la discriminazione delle categorie, che avverrà sottoforma di veri e propri “tagli” nello spazio dei parametri per separarli. Per trovare il “taglio” adeguato bisogna ricercare un modello funzionale al set di dati in esame.

I parametri importanti per decidere il taglio sono:

* efficienza: frazione di eventi di segnale accettato sul totale degli eventi di segnale
* purezza: frazione di eventi segnali presenti nel campione selezionato

Per trovare il taglio più adeguato si può sfruttare la relazione

ove , possono essere generate tramite simulazioni MC dei due tipi di eventi. Se non si possono generare le due funzioni si possono fare comunque delle ipotesi (“ansatz”) sulla forma funzionale della statistica di test t(x). Come esempio una funzione lineare , si può dimostrare che la separazione massima nel caso di funzioni gaussiane si ha con la statistica di test “discriminante lineare di Fisher”, definiti:

Con con . Nel caso di gaussiane è molto facile perché il discriminante lineare di Fisher è proporzionale a con s per il segnale e b per il fondo. In tutti gli altri casi bisognerà cercare soluzioni alternative come le reti neuronali.

Per la risoluzione dell’esercizio è stato utilizzato il pacchetto TMVA di Root che fornisce gli strumenti per l’analisi multivariata ma soprattutto vari metodi per il discrimine delle classi. Il pacchetto è munito di un apparato interno di Machine Learning, ossia i suoi algoritmi effettueranno un “train” prima su dei set di dati d’esempio e poi grazie all’intelligenza artificiale applicare lo stesso al caso in esame.

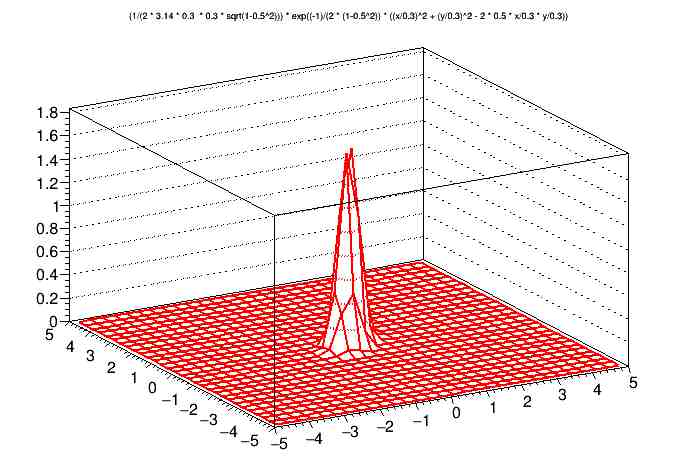
Per la risoluzione dell’esercizio è stato utilizzato tra i metodi quello LD (Linear discriminant analysis) che è un metodo equivalente al discriminante di Fisher e che si basa su effettuare una classificazione dei dati con un modello lineare, dove con lineare ci si riferisce alla funzione discriminante y(x) che è lineare nei parametri β



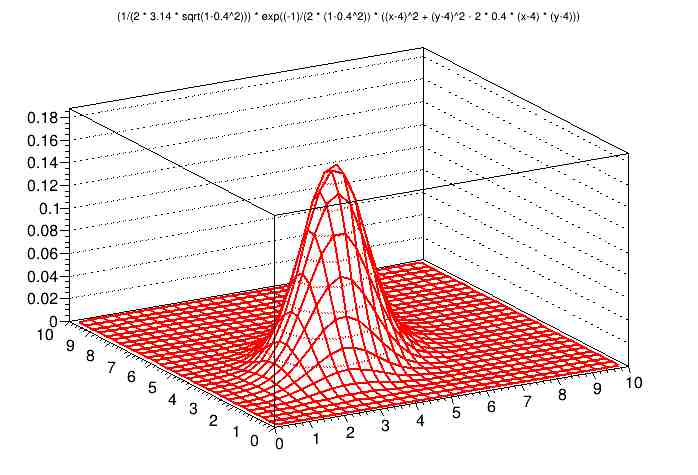
dove il bias β0 è regolato in modo tale che y(x) sia maggiore di zero per il segnale e minore di zero per il fondo.

**Risultati e discussione**

Queste sono le due classi di eventi (le due pdf), la prima è la gaussiana bidimensionale centrata in (x,y=0,0) e σx=σy=0.3, ρ=0.5 e rappresenta la classe segnale:

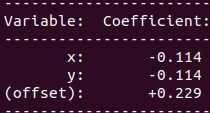


Questa è la seconda classe, che denota invece il fondo ossia la gaussiana bidimensionale centrata in (x,y)=(4,4) e σx=σy=1, ρ=0.4.



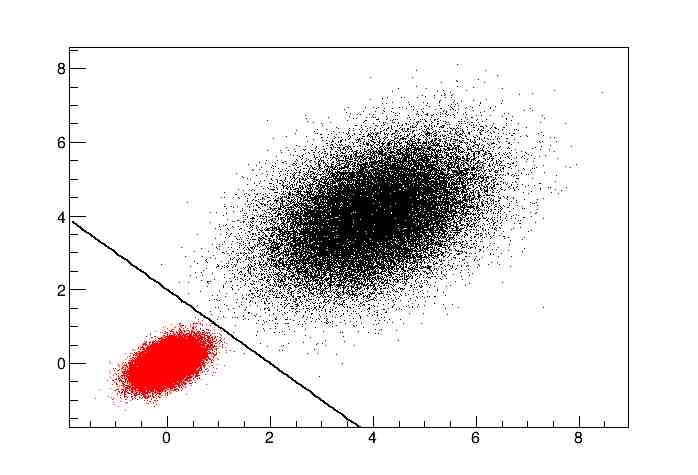
Tramite un metodo MC (generatore di numeri casuali utilizzando la funzione RAND()) si è andato a generare un set di dati per il segnale e per il fondo.

Si è sfruttato successivamente il metodo LD del pacchetto TMVA per andare a ricavare i parametri della retta (funzione lineare) in grado di andare a separare le due classi di dati. Il risultato è stato il seguente:

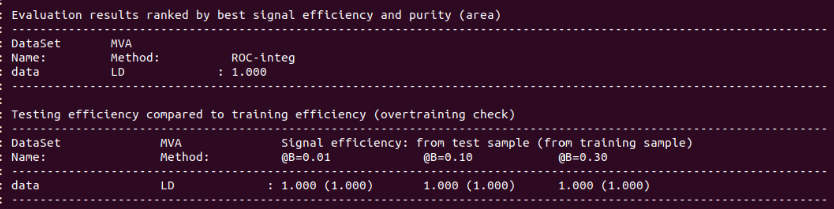


Denotando un iperpiano , vogliamo ovviamente ridurre di una dimensione per rappresentarlo come retta per questo procediamo all’intersezione con trovando quindi la retta .

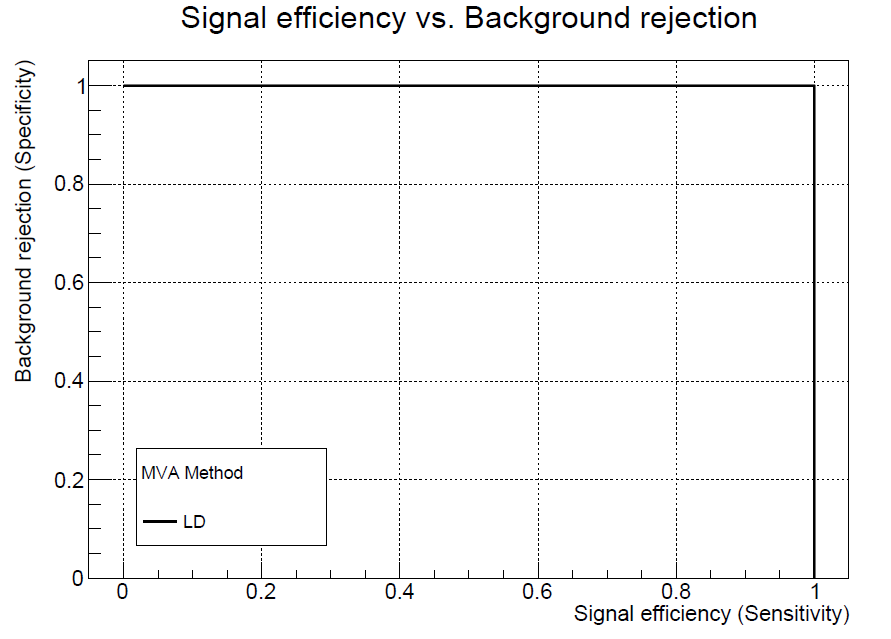
Adesso è possibile rappresentare i set di dati del fondo (colore nero) e quelli del segnale (colore rosso) ben separati con la retta ottenuta utilizzando l’analisi discriminante:

Si può notare graficamente come i due set di dati sia perfettamente separato col modello utilizzato.

Per quanto riguarda la bontà dell’analisi discriminante effettuata possiamo vedere i valori di efficienza del segnale stampati nel terminale:



Andando a creare un file .root e utilizzando l’interfaccia GUI fornita dal pacchetto TMVA è possibile estrarre ulteriori dati interessanti come ad esempio curva ROC che mette sull’asse x l’efficienza del segnale e sull’asse y il rigetto del fondo



Confermando l’efficacia del metodo utilizzato.