Corso di Laurea in Fisica
Prova di esame - Laboratorio di Calcolo e Statistica

24 febbraio 2025

Indicazioni generali

Si risolva il seguente esercizio, scrivendo un programma in `C++` o in `Python` ed organizzando il codice sorgente in modo che le funzioni utilizzate risultino implementate in librerie separate del programma principale. Ai fini della valutazione, il primo criterio che deve essere soddisfatto è che il codice sia eseguibile senza errori (inclusi quelli di compilazione, nel caso del `C++`) realizzando le funzionalità richieste dal testo. Per la valutazione sarà inoltre tenuto in considerazione il fatto che i codici sorgente siano scritti con ordine, utilizzando opportunamente l'**indentazione** e i **commenti**. Per gli svolgimenti in `C++`, si richiede infine di iniziare i codici con una riga di commento contenente il comando necessario per creare l'eseguibile.

Receiver operating characteristic

Per decidere se ritenere valido un fit con il metodo dei minimi quadrati si utilizza il test del χ^2, che fissa una soglia di accettazione Q_0^2 sul valore della somma degli scarti quadratici, Q^2, che caratterizza il singolo fit. Questa tecnica ha il costo di accettare soltanto una frazione dei fit effettuati, che viene definita in italiano specificità (*true positive rate* in inglese).
Quando il test viene applicato al caso in cui ci sono due possibili ipotesi sotto indagine, anche la probabilità che il Q^2 di un fit effettuato con il modello sbagliato sia minore della soglia Q_0^2 non è nulla ed indica il tasso di falsi positivi (*false positive rate* in inglese).
La curva che mostra l'andamento della specificità rispetto al tasso di falsi positivi è chiamata *Receiver Operating Characteristic* (spesso abbreviato in ROC) e rappresenta graficamente l'efficacia di un test di ipotesi.

1. Si crei una libreria che implementi la generazione di numeri pseudo-casuali Gaussiani utilizzando la tecnica del teorema centrale del limite e la si utilizzi per generare dieci coppie di punti (x_i, y_i) tali per cui:
$$y_i = \varphi(x_i, \theta) + \varepsilon_i \; , \tag{1}$$
dove i numeri ε_i sono indipendenti, identicamente distribuiti secondo una distribuzione di densità di probabilità Gaussiana con media $\mu = 0$ e deviazione standard $\sigma = 1.5$, mentre la funzione $\varphi(x, \theta)$ ha la seguente forma:
$$\varphi(x, \theta) = 0.5\, x^2 + 1 \; , \tag{2}$$
con le x_i distribuite a distanza regolare fra 0 e 10.
2. Si faccia il disegno dei punti così generati e si esegua il fit dei punti utilizzando la più generica parabola possibile, verificando il successo e la bontà del fit.
3. Utilizzando le tecnica dei *toy experiment* si costruisca la distribuzione del Q^2 atteso dal fit con 1000 diversi tentativi e se ne disegni l'istogramma, scegliendone con un algoritmo opportuno gli estremi ed il binning.

4. In modo analogo, si costruisca la distribuzione del Q^2 atteso nel caso in cui la funzione di fit utilizzata sia:

$$\psi(x, \theta) = a\, e^{bx} + c \tag{3}$$

e la si disegni sovrapposta all'istogramma del punto precedente.
5. Si disegni, al variare della soglia di rigetto Q^2_0, la probabilità di accettare il risultato di un fit nel caso in cui sia fatto con il modello corretto (parabolico) rispetto alla probabilità di accettarlo nel caso del modello sbagliato (esponenziale).

Gli studenti affetti da disturbi specifici dell'apprendimento (DSA) potranno tralasciare il punto 5 o, alternativamente, svolgere tutto il compito in 4 ore di tempo, dichiarando la propria preferenza all'inizio della prova. Questi dovranno anche consegnare, oltre allo svolgimento del tema, una copia del proprio Progetto Universitario Individualizzato (P.Uo.I).