SIMULAZIONE D'ESAME DI METODI STATISTICI PER LA BIOINGEGNERIA 24/25

PARTE 2: DOMANDE APERTE

Durata: 20 minuti

Domanda 1 (6 punti):

a. Data una matrice $\mathbf{X} \in \mathbb{R}^{N \times M}$, dove N è il numero di osservazioni ed M il numero di variabili, si descriva il metodo di clustering K-means riportando in particolare la funzione obiettivo che viene utilizzata nel caso di utilizzo della distanza euclidea, specificando il significato di tutti i termini utilizzati.
b. Supponendo che i dati vengano suddivisi in K cluster, quanti saranno i centroidi e che dimensione avranno?

Domanda 2 (6 punti):

a. Si scriva l'equazione del modello di regressione logistica univariata, ovvero la sua versione più semplice in cui compare una sola variabile indipendente X_1, specificando il significato dei termini coinvolti.
b. Si scriva la formula dell'odds ratio associato alla variabile X_1.
c. Spiegare come settando una soglia si può risolvere un problema di classificazione binaria tramite regressione logistica.