

### **Domanda aperta 1 (punti max 6)**

- Data una matrice  $X \in \mathbb{R}^{N \times M}$ , 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.
- Supponendo che i dati vengano suddivisi in  $K$  cluster, quanti saranno i centroidi e che dimensione avranno?

### **Domanda aperta 2 (punti max 6)**

- 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.
- Si scriva la formula dell'odds ratio associato alla variabile  $X_1$ .
- Spiegare come settando una soglia si può risolvere un problema di classificazione binaria tramite regressione logistica.