

# Probabilità e Statistica in Alta Dimensione:

## Esercitazione 3 ( $k$ -means algorithm)

2 Dicembre 2024

Disponibile all'indirizzo:  
<https://albertofachechi.com/HDP/Esercitazioni/Tracce/ES3.pdf>

1. **Task 1.**

Implementare una funzione che generi un dataset in  $d = 2$  dimensioni costituito da  $n_c = 3$  nubi Gaussiane (con media e covarianza casuali) di  $n_p = 500$  punti ognuna.

2. **Task 2.**

Scrivere il codice dell'algoritmo  $k$ -means e determinare il clustering finale fissando  $k = n_c$ .

Inviare al docente il codice insieme ad un testo in cui viene spiegato come è stato inizializzato l'algoritmo, una rappresentazione grafica del clustering ottenuto, insieme ad un grafico che mostra l'evoluzione temporale della funzione obiettivo (ogni passo dell'algoritmo corrisponde ad un diverso valore della funzione obiettivo, la funzione obiettivo non può crescere con le iterazioni).

3. **Task 3.**

Ripetere l'algoritmo diverse volte con diverse condizioni iniziali e stabilire se il clustering ottenuto è sempre lo stesso o se esso varia al variare della condizione iniziale. Verificare le prestazioni dell'algoritmo al variare del numero di componenti  $k$  da identificare.

Spiegare in un breve testo il risultato dell'analisi svolta.