



UNIVERSITÀ  
DEGLI STUDI DI BARI  
ALDO MORO

Dipartimento di Informatica

Corso di laurea in Informatica

---

**Metodi Avanzati di Programmazione**

Kmeans

**Documentazione**

Progetto di:

**Davide Cirilli** (760412) d.cirilli2@studenti.uniba.it

**Mattia Curri** (758306) m.curri8@studenti.uniba.it

**Emanuele Fontana** (758344) emanuele.fontana7@studenti.uniba.it

---

Anno Accademico **2022-2023**



# Indice

1	Introduzione	1
---	--------------	---

# 1 Introduzione

Il *K-means* è un algoritmo di clustering, una tecnica di apprendimento non supervisionato utilizzata per suddividere un insieme di dati in gruppi omogenei chiamati *cluster*. L'obiettivo del *K-means* è di assegnare ogni dato al cluster più vicino, in modo che i punti all'interno di ciascun cluster siano simili tra loro e i punti tra cluster diversi siano diversi. Ecco come funziona l'algoritmo K-means:

1. **Inizializzazione:** Si inizia scegliendo il numero  $k$  desiderato di cluster, e si selezionano casualmente  $k$  punti come centroidi iniziali. Un centroide rappresenta il centro del cluster.
2. **Assegnazione:** Per ogni punto dati, viene calcolata la distanza tra il punto e i centroidi. Il punto viene assegnato al cluster rappresentato dal centroide più vicino in base alla distanza.
3. **Aggiornamento dei centroidi:** Una volta assegnati tutti i punti ai cluster, i centroidi vengono aggiornati calcolando la media delle posizioni dei punti all'interno di ciascun cluster. Questa media diventa il nuovo centroide per il cluster corrispondente.
4. **Ripetizione:** I passi 2 e 3 vengono ripetuti fino a quando i centroidi smettono di cambiare o si raggiunge un numero massimo di iterazioni. In generale, l'algoritmo converge verso una soluzione stabile, anche se la soluzione ottenuta può essere un minimo locale invece del minimo globale.
5. **Risultato:** Alla fine delle iterazioni, si ottiene un insieme di  $k$  centroidi e i punti dati assegnati a ciascun cluster.