
Come funziona K-Means (passo passo):

1. Scegli il numero di gruppi (“cluster”) K

- Es: vuoi dividere i dati in 3 gruppi? Imposti $K = 3$.

2. Il computer piazza “a caso” K punti nel grafico

- Questi sono i “centri” iniziali dei gruppi.

3. Ogni punto viene assegnato al centro più vicino

- Il modello guarda ogni dato e lo assegna al gruppo del centro più vicino (in base alla distanza).

4. I centri si spostano

- Ogni gruppo calcola la posizione “media” dei suoi punti, e il centro si sposta lì.

5. Ripeti

- Si ripetono i passaggi 3 e 4:
 - Riassegna i punti
 - Risposta i centri
- ...fino a che i centri non cambiano più posizione (cioè, i gruppi sono “stabili”).

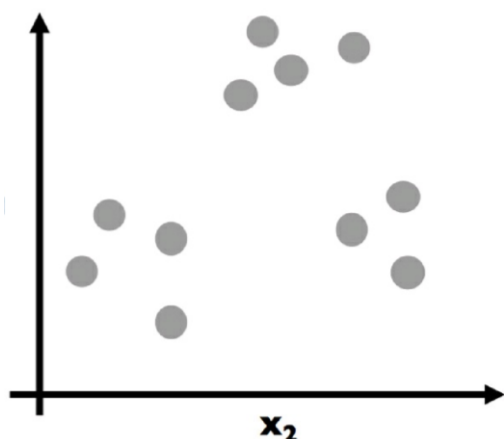


Esempio didattico

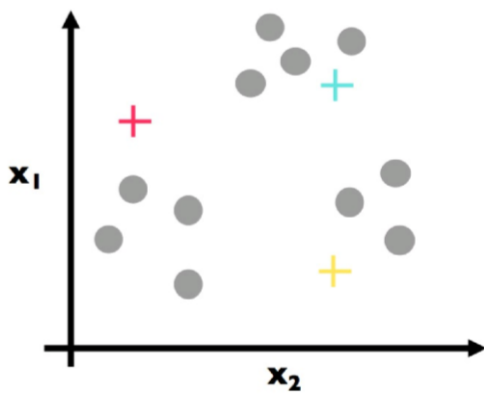
- Immagina una mappa dove vuoi raggruppare città in 3 regioni.
 - Metti 3 “bandierine” a caso.
 - Ogni città va sotto la bandierina più vicina.
 - Sposta le bandierine nel punto “medio” (ovvero la posizione media di tutte le città che appartengono a quel gruppo) delle città che rappresentano.
 - Alcune città potrebbero essere più vicine a una bandierina diversa da quella precedente
 - Continua finché le bandierine non si spostano più e i raggruppamenti si sono stabilizzati.
-

ESEMPIO VISUALE

Immaginate di avere i punti dei dati rappresentati nella Figura, in uno spazio bidimensionale. Ogni punto è colorato con lo stesso grigio per simboleggiare il fatto che non presumiamo alcun raggruppamento prima di applicare l'algoritmo K-means. L'obiettivo, qui, è quello di colorare diversamente i vari punti e di creare dei raggruppamenti (cluster).

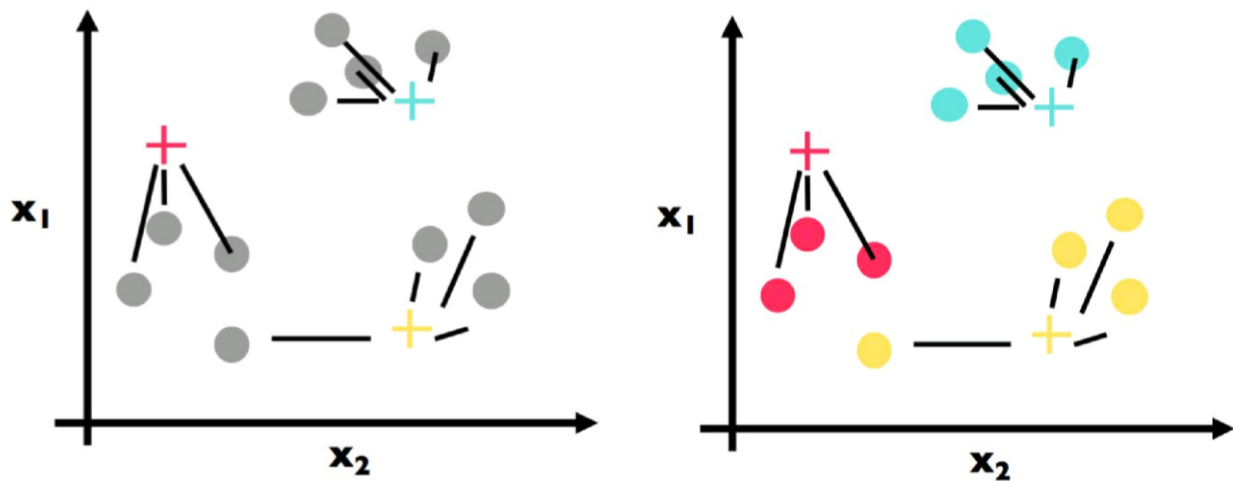


Nella Figura sotto è stato applicato il passo 2. Abbiamo scelto casualmente tre centroidi (i simboli "+" rappresentati con tre diversi livelli di grigio). La maggior parte degli algoritmi K-means colloca casualmente i centroidi iniziali, ma esistono anche altri metodi per collocare i centroidi iniziali.

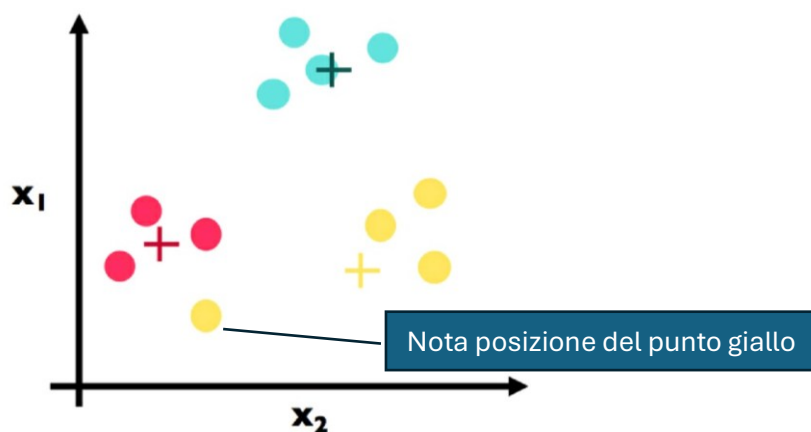


Nelle Figure sotto è stata applicata la prima parte del passo 3. Per ogni punto dei dati, abbiamo trovato il centroide più simile (più vicino).

Nella Figura a fianco è stata applicata anche una colorazione diversa per i punti che hanno ora una sfumatura di colore che concorda con quella del relativo centroide.



La Figura sotto rappresenta il passo 4, che è il punto cruciale di K- Means. Notate che abbiamo spostato i centroidi in modo che si trovino al centro del loro cluster. Per ogni colore abbiamo calcolato il punto medio e lì abbiamo collocato ogni centroide.



Proseguiamo con il nostro algoritmo ripetendo i passi 3 e 4. Abbiamo trovato il centro più vicino per ogni punto. Potete notare una variazione: il punto cerchiato nella Figura di sinistra apparteneva a un altro gruppo, di colore giallo, ma ora ha cambiato colore perché il suo centroide si è spostato più vicino agli altri punti del cluster.

