

K-Means Clustering

Evelyn Yosiana / 13522083

1. Konsep K-means clustering intinya mencari centroid (jumlahnya sebanyak cluster yang ingin dibuat) kemudian untuk tiap titik, cari centroid dengan jarak yang paling dekat.

Proses train: untuk proses train, ambil centroid random sebanyak jumlah cluster yang ingin dibuat. Kemudian untuk tiap titik data, hitung jarak ke tiap centroid kemudian cari centroid dengan jarak terdekat. Setelah itu, titik-titik data akan membentuk kelompok. Untuk tiap kelompok, cari mean dari data-data kelompok tersebut. Jika mean yang dihasilkan sama dengan centroid sebelumnya, maka proses train akan dihentikan. Jika tidak, update centroid dengan mean dari tiap kelompok, kemudian lakukan hal yang sama sampai mean = centroid atau jumlah iterasi yang diinginkan.

Proses predict: untuk proses ini, tiap data yang akan diprediksi (X_{test}) dihitung jaraknya ke centroid-centroid yang sudah ada kemudian cluster data tersebut akan sama dengan centroid yang paling dekat dengannya.

Parameter:

- `n_cluster`: banyaknya cluster yang ingin dibentuk.
 - `max_iter`: banyaknya iterasi (auto berhenti saat centroid baru dan centroid lamanya sama).
4. Pada model ini, hasil perbandingan model yang saya buat dan model dari library sebenarnya sama, namun terkadang labelnya berbeda. Misalnya data dengan cluster 0 dari kode saya akan sama dengan cluster 1 dari library.

Lampiran

Contoh perhitungan manual:

Evelyn Yonana / 13522083

K-Means Clustering

$$X_{\text{train}} = \begin{bmatrix} 1 & 2 \\ 1 & 4 \\ 1 & 0 \\ 10 & 2 \\ 10 & 4 \\ 10 & 0 \end{bmatrix} \quad \text{cluster} = 2$$

ambil centroid random:

$$c_1 = [1, 2]$$

$$c_2 = [10, 2]$$

contoh produk $[3, 7]$:

$$[3, 7] \text{ ke } [1, 2] = 5,39$$

$$[3, 7] \text{ ke } [10, 2] = 8,6$$

karena $5,39 < 8,6$ jadi $[3, 7]$ masuk

cluster 1.

hitung jarak & cari centroid paling dekat.

$$\begin{array}{l} [1, 2] \text{ ke } c_1 = 0 \\ [1, 2] \text{ ke } c_2 = 9 \end{array} \left. \vphantom{\begin{array}{l} [1, 2] \text{ ke } c_1 = 0 \\ [1, 2] \text{ ke } c_2 = 9 \end{array}} \right\} c_1$$

$$\begin{array}{l} [1, 4] \text{ ke } c_1 = 2 \\ [1, 4] \text{ ke } c_2 = 9,22 \end{array} \left. \vphantom{\begin{array}{l} [1, 4] \text{ ke } c_1 = 2 \\ [1, 4] \text{ ke } c_2 = 9,22 \end{array}} \right\} c_1$$

$$\begin{array}{l} [1, 0] \text{ ke } c_1 = 2 \\ [1, 0] \text{ ke } c_2 = 9,22 \end{array} \left. \vphantom{\begin{array}{l} [1, 0] \text{ ke } c_1 = 2 \\ [1, 0] \text{ ke } c_2 = 9,22 \end{array}} \right\} c_1$$

$$\begin{array}{l} [10, 2] \text{ ke } c_1 = 9 \\ [10, 2] \text{ ke } c_2 = 0 \end{array} \left. \vphantom{\begin{array}{l} [10, 2] \text{ ke } c_1 = 9 \\ [10, 2] \text{ ke } c_2 = 0 \end{array}} \right\} c_2$$

$$\begin{array}{l} [10, 4] \text{ ke } c_1 = 9,22 \\ [10, 4] \text{ ke } c_2 = 2 \end{array} \left. \vphantom{\begin{array}{l} [10, 4] \text{ ke } c_1 = 9,22 \\ [10, 4] \text{ ke } c_2 = 2 \end{array}} \right\} c_2$$

$$\begin{array}{l} [10, 0] \text{ ke } c_1 = 9,22 \\ [10, 0] \text{ ke } c_2 = 2 \end{array} \left. \vphantom{\begin{array}{l} [10, 0] \text{ ke } c_1 = 9,22 \\ [10, 0] \text{ ke } c_2 = 2 \end{array}} \right\} c_2$$

update centroid:

$$c_1 = \text{mean}([1, 2], [1, 4], [1, 0]) = [1, 2]$$

$$c_2 = \text{mean}([10, 2], [10, 4], [10, 0]) = [10, 2]$$

udah sama kayak centroid sebelumnya
jadi stop sampai ini.

NOTE: perhitungan pakai euclidean, ga ditulis biar singkat.