

$$\text{np.linalg.norm}(X[:, \text{np.newaxis}] - \text{centroids}, \text{axis} = 2) \Rightarrow \text{distance} (50, 3)$$

$(X \Rightarrow [50, 2])$
 $\Rightarrow [50, 1, 2]$

$\hookrightarrow n_clusters = 3 \Rightarrow (3, 2)$
 $[3, 2]$ originally

Broadcasting

$$[50, 3, 2] - [50, 3, 2]$$

hidden step: $(n_samples, n_clusters, n_features)$

diff: $\Rightarrow ([50, 3, 2])$

$\underline{d}_{ic_k} \Rightarrow$ distance b/w i th sample & c_k th centroid.

	c_1	c_2	c_3
s_1	\underline{d}_{1c_1}	\underline{d}_{1c_2}	\underline{d}_{1c_3}
s_2	\underline{d}_{2c_1}	\underline{d}_{2c_2}	\underline{d}_{2c_3}
\vdots	\vdots	\vdots	\vdots
\vdots	\vdots	\vdots	\vdots
\vdots	\vdots	\vdots	\vdots
	\underline{d}_{50c_1}	\underline{d}_{50c_2}	\underline{d}_{50c_3}

$\underline{d}_{ic_k} \Rightarrow [d_{ic_k f_1}, d_{ic_k f_2}]$
 2 features each

Euclidean norm: $\|\underline{x}\|_2 := \sqrt{x_1^2 + x_2^2 + x_3^2 + \dots + x_n^2}$

where n is the dimension of \underline{x} .

dimension of $\underline{d}_{ic_k} \Rightarrow 2 \Rightarrow$ hence we take norm along axis = 2.

$$\text{labels} = \text{np.argmin}(\text{distances}, \text{axis}=1)$$

$$(50, 3) \Rightarrow$$

$$\begin{array}{c} s_1 \\ s_2 \\ \vdots \\ \vdots \\ \vdots \end{array} \begin{bmatrix} c_1 & c_2 & c_3 \\ d_{1c_1} & d_{1c_2} & d_{1c_3} \\ d_{2c_1} & d_{2c_2} & d_{2c_3} \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ d_{50c_1} & d_{50c_2} & d_{50c_3} \end{bmatrix} \begin{array}{l} \rightarrow \text{argmin} \Rightarrow c_1 \\ \rightarrow \text{argmin} \Rightarrow c_2 \\ \rightarrow \text{argmin} \Rightarrow c_3 \end{array} \left. \vphantom{\begin{array}{c} s_1 \\ s_2 \\ \vdots \\ \vdots \\ \vdots \end{array}} \right\} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ \vdots \\ \vdots \\ c_3 \end{bmatrix} (50, 1)$$

Assign each point s to closest centroid.

new-centroids =

$$\text{np.array}(\underbrace{X[\text{labels} == i].\text{mean}(\text{axis}=0)}_{\text{mean of points in cluster } i}) \text{ for } i \text{ in range}(n\text{-clusters})$$

$$X[\text{labels} == i]$$

$$(i=c_1) \Rightarrow [\text{True False} \dots \text{False}] (50, 1)$$

$$X[\text{True False} \dots] \Rightarrow [\underline{x_1}, 0, \dots, 0] (50, 1)$$

mean of 50 point belonging to cluster 1. \leftarrow mean along axis = 0