K-means with Matrices

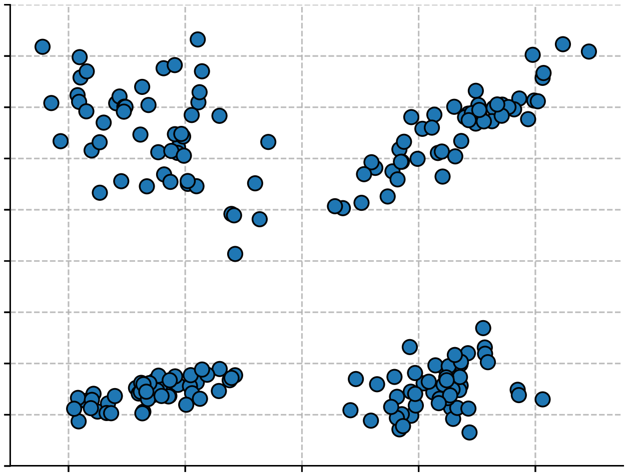
Implementing K-means in **Python** without **Sklearn**

Faculty of Mathematics and Computer Science, University of Bucharest
and
Sparktech Software

Academic Year 2018/2019, 1st Semester

K-means with matrices

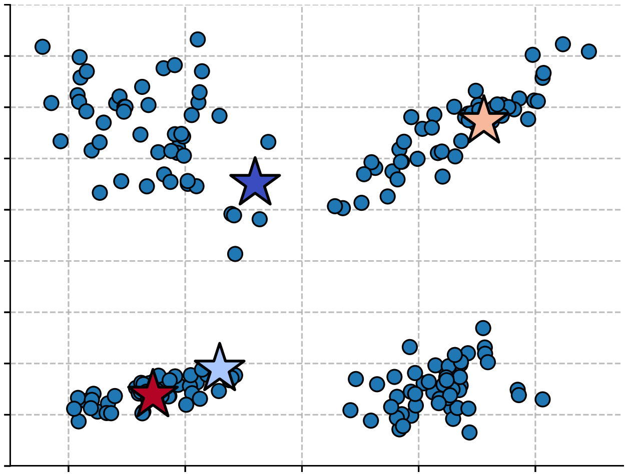
- $X = \{\vec{x}^{(1)}, \dots, \vec{x}^{(m)}\} \subset \mathbb{R}^n$ and $k \in \mathbb{N}^+$



$m = 200, \quad n = 2, \quad k = 4$

K-means with matrices

- $X = \{\vec{x}^{(1)}, \dots, \vec{x}^{(m)}\} \subset \mathbb{R}^n$ and $k \in \mathbb{N}^+$
- Initial cluster centers $\mathrm{M} = \{\vec{\mu}^{(1)}, \dots, \vec{\mu}^{(k)}\} \subset X$



$m = 200, \quad n = 2, \quad k = 4$

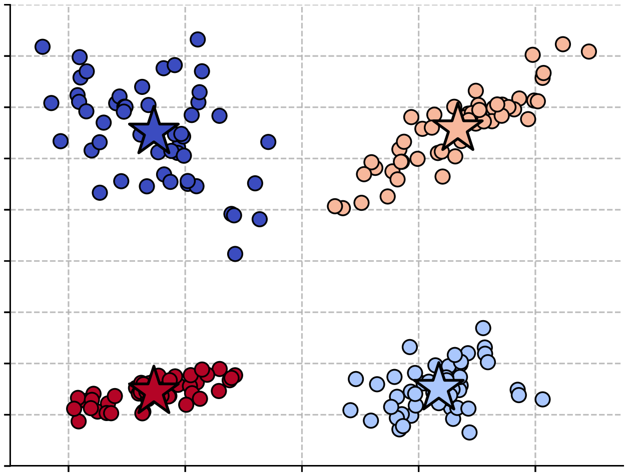
K-means with matrices

- $X = \{\vec{x}^{(1)}, \dots, \vec{x}^{(m)}\} \subset \mathbb{R}^n$ and $k \in \mathbb{N}^+$
- Initial cluster centers $\mathrm{M} = \{\vec{\mu}^{(1)}, \dots, \vec{\mu}^{(k)}\} \subset X$
- Expectation:

$$z_{ij^*} := \begin{cases} 1 & \text{if } j^* = \operatorname{argmin}_{\mathrm{j}} \left\| \vec{x}^{(i)} - \vec{\mu}^{(j)} \right\| \\ 0 & \text{otherwise} \end{cases}$$

- Maximization:

$$\vec{\mu}^{(j)} := \frac{1}{|C_j|} \sum_{\vec{x} \in C_j} \vec{x}$$



$m = 200, \quad n = 2, \quad k = 4$

K-means with matrices

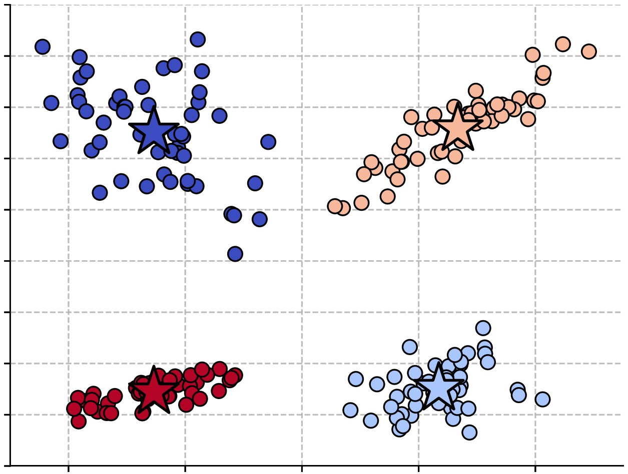
- $X = \{\vec{x}^{(1)}, \dots, \vec{x}^{(m)}\} \subset \mathbb{R}^n$ and $k \in \mathbb{N}^+$
- Initial cluster centers $\mathrm{M} = \{\vec{\mu}^{(1)}, \dots, \vec{\mu}^{(k)}\} \subset X$
- Expectation:

$$y_i = \mathrm{argmin}_\mathrm{j} \left\| \vec{x}^{(i)} - \vec{\mu}^{(j)} \right\|$$

$$z_{ij^*} \coloneqq \begin{cases} 1 & \text{if } j^* = y_i \\ 0 & \text{otherwise} \end{cases}$$

- Maximization:

$$\vec{\mu}^{(j)} \coloneqq \frac{1}{|C_j|} \sum_{\vec{x} \in C_j} \vec{x}$$



$m = 200, \quad n = 2, \quad k = 4$

K-means with matrices

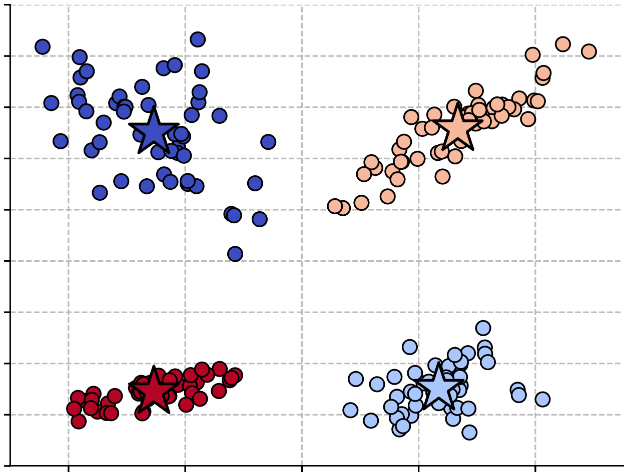
- $X = \{\vec{x}^{(1)}, \dots, \vec{x}^{(m)}\} \subset \mathbb{R}^n$ and $k \in \mathbb{N}^+$
- Initial cluster centers $\mathrm{M} = \{\vec{\mu}^{(1)}, \dots, \vec{\mu}^{(k)}\} \subset X$
- Expectation:

$$y_i = \mathrm{argmin_j} \sqrt{\langle \vec{x}^{(i)} - \vec{\mu}^{(j)}, \vec{x}^{(i)} - \vec{\mu}^{(j)} \rangle}$$

$$z_{ij^*} := \begin{cases} 1 & \text{if } j^* = y_i \\ 0 & \text{otherwise} \end{cases}$$

- Maximization:

$$\vec{\mu}^{(j)} := \frac{1}{|C_j|} \sum_{\vec{x} \in C_j} \vec{x}$$



$m = 200, \quad n = 2, \quad k = 4$

K-means with matrices

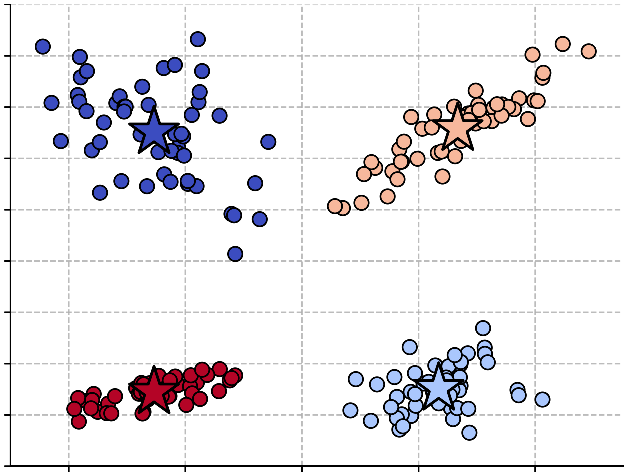
- $X = \{\vec{x}^{(1)}, \dots, \vec{x}^{(m)}\} \subset \mathbb{R}^n$ and $k \in \mathbb{N}^+$
- Initial cluster centers $\mathrm{M} = \{\vec{\mu}^{(1)}, \dots, \vec{\mu}^{(k)}\} \subset X$
- Expectation:

$$y_i = \mathrm{argmin_j} \sqrt{\langle \vec{x}^{(i)}, \vec{x}^{(i)} \rangle - 2\langle \vec{x}^{(i)}, \vec{\mu}^{(j)} \rangle + \langle \vec{\mu}^{(j)}, \vec{\mu}^{(j)} \rangle}$$

$$z_{ij^*} := \begin{cases} 1 & \text{if } j^* = y_i \\ 0 & \text{otherwise} \end{cases}$$

- Maximization:

$$\vec{\mu}^{(j)} := \frac{1}{|C_j|} \sum_{\vec{x} \in C_j} \vec{x}$$



$m = 200, \quad n = 2, \quad k = 4$

K-means with matrices

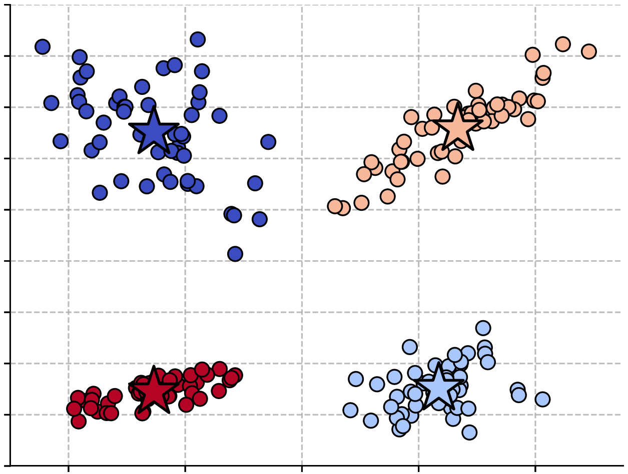
- $X = \{\vec{x}^{(1)}, \ldots, \vec{x}^{(m)}\} \subset \mathbb{R}^n$ and $k \in \mathbb{N}^+$
- Initial cluster centers $\mathrm{M} = \{\vec{\mu}^{(1)}, \ldots, \vec{\mu}^{(k)}\} \subset X$
- Expectation:

$$y_i = \mathrm{argmin}_{\mathrm{j}} \left(-2\langle \vec{x}^{(i)}, \vec{\mu}^{(j)} \rangle + \left\|\vec{\mu}^{(j)}\right\|^2\right)$$

$$z_{i\,j^*} := \begin{cases} 1 & \text{if } j^* = y_i \\ 0 & \text{otherwise} \end{cases}$$

- Maximization:

$$\vec{\mu}^{(j)} := \frac{1}{|C_j|} \sum_{\vec{x} \in C_j} \vec{x}$$



$m = 200, \quad n = 2, \quad k = 4$

K-means with matrices

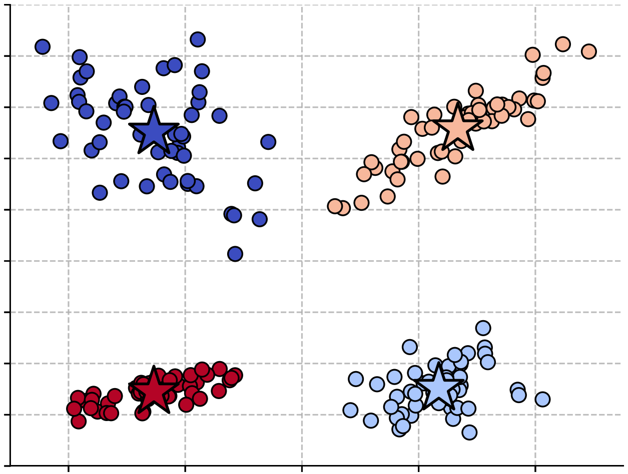
- $X = \{\vec{x}^{(1)}, \dots, \vec{x}^{(m)}\} \subset \mathbb{R}^n$ and $k \in \mathbb{N}^+$
- Initial cluster centers $\mathrm{M} = \{\vec{\mu}^{(1)}, \dots, \vec{\mu}^{(k)}\} \subset X$
- Expectation:

$$y_i = \mathrm{argmin_j}\left(-2\langle\vec{x}^{(i)}, \vec{\mu}^{(j)}\rangle + \left\|\vec{\mu}^{(j)}\right\|^2\right)$$

$$z_{ij^*} \coloneqq \begin{cases} 1 & \text{if } j^* = y_i \\ 0 & \text{otherwise} \end{cases}$$

- Maximization:

$$\vec{\mu}^{(j)} \coloneqq \frac{1}{\sum_{\mathrm{i}} z_{ij}} \sum_{\mathrm{i}} z_{ij}\vec{x}^{(i)}$$



$m = 200, \quad n = 2, \quad k = 4$

K-means with matrices

$$X = \begin{pmatrix} x_1^{(1)} & x_2^{(1)} & \cdots & x_n^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \cdots & x_n^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{(m)} & x_2^{(m)} & \cdots & x_n^{(m)} \end{pmatrix}_{m \times n}$$

$$M \in \mathbb{R}^{k \times n} = \begin{pmatrix} \mu_1^{(1)} & \mu_2^{(1)} & \cdots & \mu_n^{(1)} \\ \mu_1^{(2)} & \mu_2^{(2)} & \cdots & \mu_n^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_1^{(k)} & \mu_2^{(k)} & \cdots & \mu_n^{(k)} \end{pmatrix}_{k \times n}$$

K-means with matrices

$$X = \begin{pmatrix} x_1^{(1)} & x_2^{(1)} & \cdots & x_n^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \cdots & x_n^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{(m)} & x_2^{(m)} & \cdots & x_n^{(m)} \end{pmatrix}_{m \times n}$$

$$M \in \mathbb{R}^{k \times n} = \begin{pmatrix} \mu_1^{(1)} & \mu_2^{(1)} & \cdots & \mu_n^{(1)} \\ \mu_1^{(2)} & \mu_2^{(2)} & \cdots & \mu_n^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_1^{(k)} & \mu_2^{(k)} & \cdots & \mu_n^{(k)} \end{pmatrix}_{k \times n}$$

$$XM^T = \begin{pmatrix} \langle \vec{x}^{(1)}, \vec{\mu}^{(1)} \rangle & \langle \vec{x}^{(1)}, \vec{\mu}^{(2)} \rangle & \cdots & \langle \vec{x}^{(1)}, \vec{\mu}^{(k)} \rangle \\ \langle \vec{x}^{(2)}, \vec{\mu}^{(1)} \rangle & \langle \vec{x}^{(2)}, \vec{\mu}^{(2)} \rangle & \cdots & \langle \vec{x}^{(2)}, \vec{\mu}^{(k)} \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \vec{x}^{(m)}, \vec{\mu}^{(1)} \rangle & \langle \vec{x}^{(m)}, \vec{\mu}^{(2)} \rangle & \cdots & \langle \vec{x}^{(m)}, \vec{\mu}^{(k)} \rangle \end{pmatrix}_{m \times k}$$

K-means with matrices

$$X = \begin{pmatrix} x_1^{(1)} & x_2^{(1)} & \cdots & x_n^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \cdots & x_n^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{(m)} & x_2^{(m)} & \cdots & x_n^{(m)} \end{pmatrix}_{m \times n}$$

$$M \in \mathbb{R}^{k \times n} = \begin{pmatrix} \mu_1^{(1)} & \mu_2^{(1)} & \cdots & \mu_n^{(1)} \\ \mu_1^{(2)} & \mu_2^{(2)} & \cdots & \mu_n^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_1^{(k)} & \mu_2^{(k)} & \cdots & \mu_n^{(k)} \end{pmatrix}_{k \times n}$$

$$XM^T = \begin{pmatrix} \langle \vec{x}^{(1)}, \vec{\mu}^{(1)} \rangle & \langle \vec{x}^{(1)}, \vec{\mu}^{(2)} \rangle & \cdots & \langle \vec{x}^{(1)}, \vec{\mu}^{(k)} \rangle \\ \langle \vec{x}^{(2)}, \vec{\mu}^{(1)} \rangle & \langle \vec{x}^{(2)}, \vec{\mu}^{(2)} \rangle & \cdots & \langle \vec{x}^{(2)}, \vec{\mu}^{(k)} \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \vec{x}^{(m)}, \vec{\mu}^{(1)} \rangle & \langle \vec{x}^{(m)}, \vec{\mu}^{(2)} \rangle & \cdots & \langle \vec{x}^{(m)}, \vec{\mu}^{(k)} \rangle \end{pmatrix}_{m \times k}$$

$$M_* = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix}_{m \times n} \cdot (M^T)^2$$

Point-wise squaring

K-means with matrices

$$X = \begin{pmatrix} x_1^{(1)} & x_2^{(1)} & \cdots & x_n^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \cdots & x_n^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{(m)} & x_2^{(m)} & \cdots & x_n^{(m)} \end{pmatrix}_{m \times n}$$

$$M \in \mathbb{R}^{k \times n} = \begin{pmatrix} \mu_1^{(1)} & \mu_2^{(1)} & \cdots & \mu_n^{(1)} \\ \mu_1^{(2)} & \mu_2^{(2)} & \cdots & \mu_n^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_1^{(k)} & \mu_2^{(k)} & \cdots & \mu_n^{(k)} \end{pmatrix}_{k \times n}$$

$$XM^T = \begin{pmatrix} \langle \vec{x}^{(1)}, \vec{\mu}^{(1)} \rangle & \langle \vec{x}^{(1)}, \vec{\mu}^{(2)} \rangle & \cdots & \langle \vec{x}^{(1)}, \vec{\mu}^{(k)} \rangle \\ \langle \vec{x}^{(2)}, \vec{\mu}^{(1)} \rangle & \langle \vec{x}^{(2)}, \vec{\mu}^{(2)} \rangle & \cdots & \langle \vec{x}^{(2)}, \vec{\mu}^{(k)} \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \vec{x}^{(m)}, \vec{\mu}^{(1)} \rangle & \langle \vec{x}^{(m)}, \vec{\mu}^{(2)} \rangle & \cdots & \langle \vec{x}^{(m)}, \vec{\mu}^{(k)} \rangle \end{pmatrix}_{m \times k}$$

$$\boxed{1 \cdot {\mu_1^{(1)}}^2 + 1 \cdot {\mu_2^{(1)}}^2 + \cdots + 1 \cdot {\mu_n^{(1)}}^2}$$

$$M_* = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix}_{m \times n} \cdot (M^T)^2 = 1_{m \times n} \cdot \begin{pmatrix} {\mu_1^{(1)}}^2 & \cdots & {\mu_1^{(k)}}^2 \\ {\mu_2^{(1)}}^2 & \cdots & {\mu_2^{(k)}}^2 \\ \vdots & \ddots & \vdots \\ {\mu_n^{(1)}}^2 & \cdots & {\mu_n^{(k)}}^2 \end{pmatrix} = \begin{pmatrix} \lVert \vec{\mu}^{(1)} \rVert^2 & & \\ & & \\ & & \end{pmatrix}_{m \times k}$$

Point-wise squaring

K-means with matrices

$$X = \begin{pmatrix} x_1^{(1)} & x_2^{(1)} & \cdots & x_n^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \cdots & x_n^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{(m)} & x_2^{(m)} & \cdots & x_n^{(m)} \end{pmatrix}_{m\times n} \qquad M \in \mathbb{R}^{k\times n} = \begin{pmatrix} \mu_1^{(1)} & \mu_2^{(1)} & \cdots & \mu_n^{(1)} \\ \mu_1^{(2)} & \mu_2^{(2)} & \cdots & \mu_n^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_1^{(k)} & \mu_2^{(k)} & \cdots & \mu_n^{(k)} \end{pmatrix}_{k\times n}$$

$$XM^T = \begin{pmatrix} \langle \vec{x}^{(1)}, \vec{\mu}^{(1)} \rangle & \langle \vec{x}^{(1)}, \vec{\mu}^{(2)} \rangle & \cdots & \langle \vec{x}^{(1)}, \vec{\mu}^{(k)} \rangle \\ \langle \vec{x}^{(2)}, \vec{\mu}^{(1)} \rangle & \langle \vec{x}^{(2)}, \vec{\mu}^{(2)} \rangle & \cdots & \langle \vec{x}^{(2)}, \vec{\mu}^{(k)} \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \vec{x}^{(m)}, \vec{\mu}^{(1)} \rangle & \langle \vec{x}^{(m)}, \vec{\mu}^{(2)} \rangle & \cdots & \langle \vec{x}^{(m)}, \vec{\mu}^{(k)} \rangle \end{pmatrix}_{m\times k}$$

$$M_* = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix}_{m\times n} \cdot (M^T)^2 = 1_{m\times n} \cdot \begin{pmatrix} {\mu_1^{(1)}}^2 & \cdots & {\mu_1^{(k)}}^2 \\ {\mu_2^{(1)}}^2 & \cdots & {\mu_2^{(k)}}^2 \\ \vdots & \ddots & \vdots \\ {\mu_n^{(1)}}^2 & \cdots & {\mu_n^{(k)}}^2 \end{pmatrix} = \begin{pmatrix} \|\vec{\mu}^{(1)}\|^2 & \|\vec{\mu}^{(2)}\|^2 & \cdots & \|\vec{\mu}^{(k)}\|^2 \\ \|\vec{\mu}^{(1)}\|^2 & \|\vec{\mu}^{(2)}\|^2 & \cdots & \|\vec{\mu}^{(k)}\|^2 \\ \vdots & \vdots & \ddots & \vdots \\ \|\vec{\mu}^{(1)}\|^2 & \|\vec{\mu}^{(2)}\|^2 & \cdots & \|\vec{\mu}^{(k)}\|^2 \end{pmatrix}_{m\times k}$$

Point-wise squaring

K-means with matrices

Point-wise operations	Matrix operations
$y_i = \text{argmin}_\text{j}\left(-2\langle \vec{x}^{(i)}, \vec{\mu}^{(j)}\rangle + \left\|\vec{\mu}^{(j)}\right\|^2\right)$	
$z_{ij^*} := \begin{cases} 1 & \text{if } j^* = y_i \\ 0 & \text{otherwise} \end{cases}$	
$\vec{\mu}^{(j)} := \frac{1}{\sum_\text{i} z_{ij}} \sum_\text{i} z_{ij} \vec{x}^{(i)}$	

$$XM^T = \begin{pmatrix} \langle \vec{x}^{(1)}, \vec{\mu}^{(1)} \rangle & \cdots & \langle \vec{x}^{(1)}, \vec{\mu}^{(k)} \rangle \\ \vdots & \ddots & \vdots \\ \langle \vec{x}^{(m)}, \vec{\mu}^{(1)} \rangle & \cdots & \langle \vec{x}^{(m)}, \vec{\mu}^{(k)} \rangle \end{pmatrix}$$

$$M_* = \begin{pmatrix} \left\|\vec{\mu}^{(1)}\right\|^2 & \cdots & \left\|\vec{\mu}^{(k)}\right\|^2 \\ \vdots & \ddots & \vdots \\ \left\|\vec{\mu}^{(1)}\right\|^2 & \cdots & \left\|\vec{\mu}^{(k)}\right\|^2 \end{pmatrix}$$

K-means with matrices

Point-wise operations	Matrix operations
$y_i = \mathrm{argmin_j} \left(-2\langle \vec{x}^{(i)}, \vec{\mu}^{(j)} \rangle + \left\Vert \vec{\mu}^{(j)} \right\Vert^2 \right)$	$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = \mathrm{rowwise\_argmin}(-2XM^T + M_*)$
$z_{ij^*} := \begin{cases} 1 & \text{if } j^* = y_i \\ 0 & \text{otherwise} \end{cases}$	
$\vec{\mu}^{(j)} := \frac{1}{\sum_{\mathrm{i}} z_{ij}} \sum_{\mathrm{i}} z_{ij} \vec{x}^{(i)}$	

$$XM^T = \begin{pmatrix} \langle \vec{x}^{(1)}, \vec{\mu}^{(1)} \rangle & \cdots & \langle \vec{x}^{(1)}, \vec{\mu}^{(k)} \rangle \\ \vdots & \ddots & \vdots \\ \langle \vec{x}^{(m)}, \vec{\mu}^{(1)} \rangle & \cdots & \langle \vec{x}^{(m)}, \vec{\mu}^{(k)} \rangle \end{pmatrix}$$

$$M_* = \begin{pmatrix} \left\Vert \vec{\mu}^{(1)} \right\Vert^2 & \cdots & \left\Vert \vec{\mu}^{(k)} \right\Vert^2 \\ \vdots & \ddots & \vdots \\ \left\Vert \vec{\mu}^{(1)} \right\Vert^2 & \cdots & \left\Vert \vec{\mu}^{(k)} \right\Vert^2 \end{pmatrix}$$

K-means with matrices

Point-wise operations	Matrix operations
$y_i = \mathrm{argmin_j}\left(-2\langle\vec{x}^{(i)}, \vec{\mu}^{(j)}\rangle + \left\|\vec{\mu}^{(j)}\right\|^2\right)$	$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = \mathrm{rowwise\_argmin}(-2XM^T + M_*)$
$z_{ij^*} := \begin{cases} 1 & \text{if } j^* = y_i \\ 0 & \text{otherwise} \end{cases}$	$Z = \begin{pmatrix} z_{11} & \cdots & z_{1k} \\ \vdots & z_{ij} & \vdots \\ z_{m1} & \cdots & z_{mk} \end{pmatrix} = \begin{pmatrix} \mathrm{one\_hot}(y_1) \\ \vdots \\ \mathrm{one\_hot}(y_m) \end{pmatrix}$
$\vec{\mu}^{(j)} := \frac{1}{\sum_\mathrm{i} z_{ij}} \sum_\mathrm{i} z_{ij}\vec{x}^{(i)}$	

$$XM^T = \begin{pmatrix} \langle\vec{x}^{(1)}, \vec{\mu}^{(1)}\rangle & \cdots & \langle\vec{x}^{(1)}, \vec{\mu}^{(k)}\rangle \\ \vdots & \ddots & \vdots \\ \langle\vec{x}^{(m)}, \vec{\mu}^{(1)}\rangle & \cdots & \langle\vec{x}^{(m)}, \vec{\mu}^{(k)}\rangle \end{pmatrix}$$

$$M_* = \begin{pmatrix} \left\|\vec{\mu}^{(1)}\right\|^2 & \cdots & \left\|\vec{\mu}^{(k)}\right\|^2 \\ \vdots & \ddots & \vdots \\ \left\|\vec{\mu}^{(1)}\right\|^2 & \cdots & \left\|\vec{\mu}^{(k)}\right\|^2 \end{pmatrix}$$

K-means with matrices

Point-wise operations	Matrix operations
$y_i = \mathrm{argmin_j} \left(-2\langle \vec{x}^{(i)}, \vec{\mu}^{(j)}\rangle + \left\Vert \vec{\mu}^{(j)} \right\Vert^2\right)$	$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = \mathrm{rowwise\_argmin}(-2XM^T + M_*)$
$z_{ij^*} := \begin{cases} 1 & \text{if } j^* = y_i \\ 0 & \text{otherwise} \end{cases}$	$Z = \begin{pmatrix} z_{11} & \cdots & z_{1k} \\ \vdots & z_{ij} & \vdots \\ z_{m1} & \cdots & z_{mk} \end{pmatrix} = \begin{pmatrix} \mathrm{one\_hot}(y_1) \\ \vdots \\ \mathrm{one\_hot}(y_m) \end{pmatrix}$
$\vec{\mu}^{(j)} := \dfrac{1}{\sum_{\mathrm{i}} z_{ij}} \sum_{\mathrm{i}} z_{ij} \vec{x}^{(i)}$	

$$XM^T = \begin{pmatrix} \langle \vec{x}^{(1)}, \vec{\mu}^{(1)} \rangle & \cdots & \langle \vec{x}^{(1)}, \vec{\mu}^{(k)} \rangle \\ \vdots & \ddots & \vdots \\ \langle \vec{x}^{(m)}, \vec{\mu}^{(1)} \rangle & \cdots & \langle \vec{x}^{(m)}, \vec{\mu}^{(k)} \rangle \end{pmatrix}$$

$$M_* = \begin{pmatrix} \left\Vert \vec{\mu}^{(1)} \right\Vert^2 & \cdots & \left\Vert \vec{\mu}^{(k)} \right\Vert^2 \\ \vdots & \ddots & \vdots \\ \left\Vert \vec{\mu}^{(1)} \right\Vert^2 & \cdots & \left\Vert \vec{\mu}^{(k)} \right\Vert^2 \end{pmatrix}$$

One-hot encoding of y_i basically means row y_i of I_k (identity matrix of size $k \times k$):

$$I_4 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$k = 4 \implies$$
$$\mathrm{one\_hot}(3) = [0 \quad 0 \quad 1 \quad 0]$$
$$\mathrm{one\_hot}(1) = [1 \quad 0 \quad 0 \quad 0]$$

K-means with matrices

Point-wise operations	Matrix operations
$y_i = \mathrm{argmin_j} \left(-2\langle \vec{x}^{(i)}, \vec{\mu}^{(j)} \rangle + \left\lVert \vec{\mu}^{(j)} \right\rVert^2 \right)$	$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = \mathrm{rowwise\_argmin}(-2XM^T + M_*)$
$z_{ij^*} := \begin{cases} 1 & \text{if } j^* = y_i \\ 0 & \text{otherwise} \end{cases}$	$Z = \begin{pmatrix} z_{11} & \cdots & z_{1k} \\ \vdots & z_{ij} & \vdots \\ z_{m1} & \cdots & z_{mk} \end{pmatrix} = \begin{pmatrix} \mathrm{one\_hot}(y_1) \\ \vdots \\ \mathrm{one\_hot}(y_m) \end{pmatrix}$
$\vec{\mu}^{(j)} := \frac{1}{\sum_\mathrm{i} z_{ij}} \sum_\mathrm{i} z_{ij} \vec{x}^{(i)}$	

- The matrix Z is a permutation (with replacement) of I_k with rows selected according to Y.

$$XM^T = \begin{pmatrix} \langle \vec{x}^{(1)}, \vec{\mu}^{(1)} \rangle & \cdots & \langle \vec{x}^{(1)}, \vec{\mu}^{(k)} \rangle \\ \vdots & \ddots & \vdots \\ \langle \vec{x}^{(m)}, \vec{\mu}^{(1)} \rangle & \cdots & \langle \vec{x}^{(m)}, \vec{\mu}^{(k)} \rangle \end{pmatrix}$$

$$M_* = \begin{pmatrix} \left\lVert \vec{\mu}^{(1)} \right\rVert^2 & \cdots & \left\lVert \vec{\mu}^{(k)} \right\rVert^2 \\ \vdots & \ddots & \vdots \\ \left\lVert \vec{\mu}^{(1)} \right\rVert^2 & \cdots & \left\lVert \vec{\mu}^{(k)} \right\rVert^2 \end{pmatrix}$$

One-hot encoding of y_i basically means row y_i of I_k (identity matrix of size $k \times k$):

$$I_4 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$k = 4 \implies$$
$$\mathrm{one\_hot}(3) = [0 \quad 0 \quad 1 \quad 0]$$
$$\mathrm{one\_hot}(1) = [1 \quad 0 \quad 0 \quad 0]$$

K-means with matrices

Point-wise operations	Matrix operations
$y_i = \mathrm{argmin_j}\left(-2\langle\vec{x}^{(i)}, \vec{\mu}^{(j)}\rangle + \left\|\vec{\mu}^{(j)}\right\|^2\right)$	$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = \mathrm{rowwise\_argmin}(-2XM^T + M_*)$
$z_{ij^*} := \begin{cases} 1 & \text{if } j^* = y_i \\ 0 & \text{otherwise} \end{cases}$	$Z = \begin{pmatrix} z_{11} & \cdots & z_{1k} \\ \vdots & z_{ij} & \vdots \\ z_{m1} & \cdots & z_{mk} \end{pmatrix} = \begin{pmatrix} \mathrm{one\_hot}(y_1) \\ \vdots \\ \mathrm{one\_hot}(y_m) \end{pmatrix}$
$\vec{\mu}^{(j)} := \frac{1}{\sum_\mathrm{i} z_{ij}} \sum_\mathrm{i} z_{ij}\vec{x}^{(i)}$	$M = \frac{Z^T X}{Z^T 1_{m\times n}}$

- The matrix Z is a permutation (with replacement) of I_k with rows selected according to Y.

$$XM^T = \begin{pmatrix} \langle\vec{x}^{(1)}, \vec{\mu}^{(1)}\rangle & \cdots & \langle\vec{x}^{(1)}, \vec{\mu}^{(k)}\rangle \\ \vdots & \ddots & \vdots \\ \langle\vec{x}^{(m)}, \vec{\mu}^{(1)}\rangle & \cdots & \langle\vec{x}^{(m)}, \vec{\mu}^{(k)}\rangle \end{pmatrix}$$

$$M_* = \begin{pmatrix} \left\|\vec{\mu}^{(1)}\right\|^2 & \cdots & \left\|\vec{\mu}^{(k)}\right\|^2 \\ \vdots & \ddots & \vdots \\ \left\|\vec{\mu}^{(1)}\right\|^2 & \cdots & \left\|\vec{\mu}^{(k)}\right\|^2 \end{pmatrix}$$

One-hot encoding of y_i basically means row y_i of I_k (identity matrix of size $k \times k$):

$$I_4 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$k = 4 \implies$$
$$\mathrm{one\_hot}(3) = [0 \quad 0 \quad 1 \quad 0]$$
$$\mathrm{one\_hot}(1) = [1 \quad 0 \quad 0 \quad 0]$$

K-means with matrices

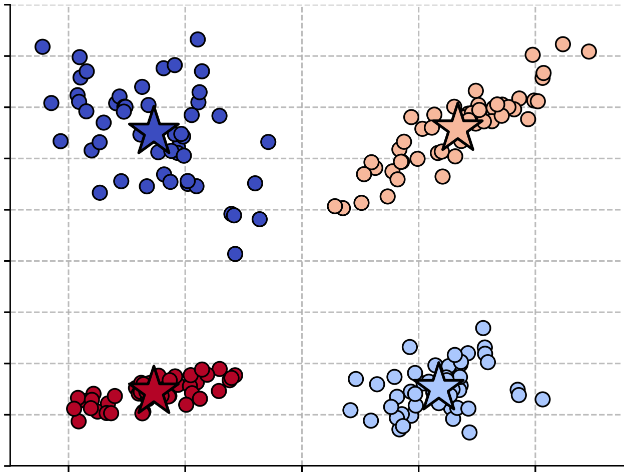
- $X \in \mathbb{R}^{m \times n}$ and $k \in \mathbb{N}^+$
- Initial cluster centers $\mathrm{M} \in \mathbb{R}^{k \times n}$

- Expectation:

$$M_* = 1_{m \times n}(M^T)^2$$
$$Y = \text{rowwise\_argmin}(-2XM^T + M_*)$$
$$Z = \text{rowwise\_onehot}(Y) = I_k[Y]$$

- Maximization:

$$M = \frac{Z^T X}{Z^T 1_{m \times n}}$$



$m = 200, \quad n = 2, \quad k = 4$

K-means in Python without Sklearn

```python
import numpy as np

m , n = X.shape # (200, 2)
k = 4
indices = np.arange(m) # indices = np.array([0, 1, 2, …, 199])
np.random.suffle(indices) # shuffles the indices in place
M = X[indices[:k]] # select the point of the first k indices (after shuffling)

n_steps = 50
for _ in range(n_steps):
    M_star = np.ones((m, n)).dot(M.T**2) # M_star.shape == (200, 4)
    y = np.argmin(-2*X.dot(M.T) + M_star, axis = 1) # y.shape == (200,)
    Z = np.eye(k)[y] # Expectation
    M = Z.T.dot(X) / Z.T.dot(np.ones((m, n))) # Maximization
```

K-means in Python with some Sklearn

```python
import numpy as np
from sklearn.metrics.pairwise import euclidean_distances
from sklearn.utils import resample

m , n = X.shape # (200, 2)
k = 4
M = resample(X, n_samples = k)

n_steps = 50
for _ in range(n_steps):
    y = np.argmin(euclidean_distances(X, M), axis = 1) # y.shape == (200,)
    Z = np.eye(k)[y]
    M = Z.T.dot(X) / Z.sum(axis = 0).reshape(-1, 1) # reshape turns shape (4,) into (4, 1)
```