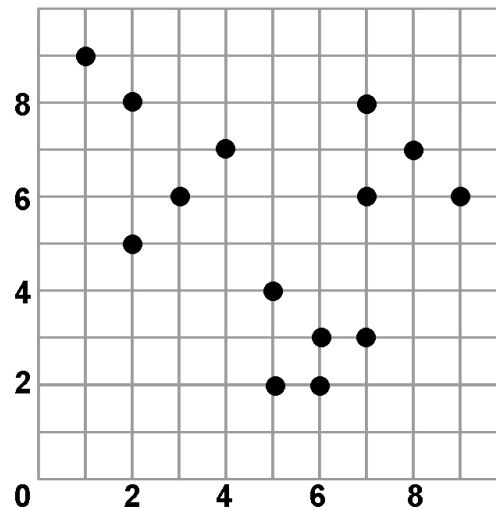


Artificial Intelligence III: Artificial Intelligence and Deep Learning

Ch06 – Unsupervised Learning

Tutorial

1. Given the following dataset:



- By using Single Linkage approach, group the samples into THREE clusters. Euclidean distance should be used.
- By using Complete Linkage approach, group the samples into THREE clusters. Euclidean distance should be used.
- Assume Sum-of-Squared-Error (SSE) ($S = \sum_{i=1}^c \sum_{x \in D_i} \|x - m_i\|^2$) as a criterion function. By setting initial points to (0, 0), (0, 10) and (8, 8), find the clusters using K-mean ($k = 3$).

2. Given the following dataset:

Sample	Feature 1	Feature 2
s ₁	1	0
s ₂	2	0
s ₃	3	0
s ₄	5	6
s ₅	6	6
s ₆	7	6

- a) Calculate the standardized dataset (A)
 - $a = A - \mu$
- b) Calculate the **covariance matrix (M)** for the features in the dataset
 - $Cov(D) = M = \frac{a^T a}{n}$
- c) Calculate the **eigenvalues** (λ_i) and **eigenvectors** (x_i, y_i) for the covariance matrix.
 - $det(M - I\lambda) = 0$
 - $M \begin{bmatrix} x_i \\ y_i \end{bmatrix} = \lambda_i \begin{bmatrix} x_i \\ y_i \end{bmatrix}$
 - Put $x_i = 1$ and convert to unit vector: $\sqrt{x_i^2 + y_i^2} = 1$
- d) **Pick top k eigenvalues, where k=2 (Λ_k)**
 - $\Lambda_k = \begin{bmatrix} x_1 & x_2 & \dots & x_k \\ y_1 & y_2 & \dots & y_k \end{bmatrix}$, where x_i and y_i correspond to the i^{th} highest λ .
- e) Transform the original space to the principal component space (X)
 - $M\Lambda_k = X$