P09 Clustering

Luis Sa-Couto[1] and Andreas Wichert[2]

INESC-ID, Instituto Superior Tecnico, Universidade de Lisboa
{luis.sa.couto,andreas.wichert}@tecnico.ulisboa.pt

1 K-means clustering

1) Consider the following training data without labels:

$$\left\{\mathbf{x}^{(1)} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \mathbf{x}^{(2)} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix} 0 \\ 2 \end{pmatrix}, \mathbf{x}^{(4)} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}\right\}$$

Also, consider the following initialization centroids for $k = 2$ clusters $\mu^1 = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$ and $\mu^2 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$.

a) Apply the k-means clustering algorithm until convergence.
b) Plot the data points and draw the clusters.

2) Consider the following training data without labels:

$$\left\{\mathbf{x}^{(1)} = \begin{pmatrix} 1.0 \\ 0.0 \\ 0.0 \end{pmatrix}, \mathbf{x}^{(2)} = \begin{pmatrix} 8.0 \\ 8.0 \\ 4.0 \end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix} 3.0 \\ 3.0 \\ 0.0 \end{pmatrix}, \mathbf{x}^{(4)} = \begin{pmatrix} 0.0 \\ 0.0 \\ 1.0 \end{pmatrix}, \mathbf{x}^{(5)} = \begin{pmatrix} 0.0 \\ 1.0 \\ 0.0 \end{pmatrix}, \mathbf{x}^{(6)} = \begin{pmatrix} 3.0 \\ 2.0 \\ 1.0 \end{pmatrix}\right\}$$

Everytime you need to initialize k clusters, do it by taking the first k points of the dataset and using them as centroids.

a) For $k = 2$ perform k-means clustering until convergence.
b) For $k = 3$ perform k-means clustering until convergence.
c) Which k provides a better clustering in terms of sum of intra-cluster euclidean distances.
d) Which k provides a better clustering in terms of mean inter-cluster centroid distance.

2 Expectation-Maximization Clustering

1) Consider the following training data with boolean features:

$$\left\{\mathbf{x}^{(1)} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \mathbf{x}^{(2)} = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \end{pmatrix}, \mathbf{x}^{(4)} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \mathbf{x}^{(5)} = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}\right\}$$

We want to model the data with three clusters. Initialize all priors uniformly and initialize using the following table:

	$p(x_1 = 1 \mid C = c)$	$p(x_2 = 1 \mid C = c)$	$p(x_3 = 1 \mid C = c)$	$p(x_4 = 1 \mid C = c)$
$c = 1$	0.8	0.5	0.1	0.1
$c = 2$	0.1	0.5	0.4	0.8
$c = 3$	0.1	0.1	0.9	0.2

Assume all features are conditionally independent given the cluster.

a) Perform one expectation maximization iteration.

b) Verify that after one iteration the probability of the data increased.

2) Consider the following training data without labels:

$$\left\{\mathbf{x}^{(1)} = \left(4\right), \mathbf{x}^{(2)} = \left(0\right), \mathbf{x}^{(3)} = \left(1\right)\right\}$$

We want to model the data with a mixture of two normal distributions. Initialize the likelihoods as follows:

$$p\left(\mathbf{x} \mid C = 1\right) = \mathcal{N}\left(\mu^1 = 0, \sigma^1 = 1\right)$$

$$p\left(\mathbf{x} \mid C = 2\right) = \mathcal{N}\left(\mu^2 = 1, \sigma^2 = 1\right)$$

Also, initialize the priors as follows:

$$p\left(C = 1\right) = 0.5$$

$$p\left(C = 2\right) = 0.5$$

a) Perform one expectation maximization iteration.

b) Plot the points and sketch the clusters.

3) Consider the following training data without labels:

$$\left\{\mathbf{x}^{(1)} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \mathbf{x}^{(2)} = \begin{pmatrix} 0 \\ 2 \end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}\right\}$$

We want to model the data with a mixture of two multivariate normal distributions. Initialize the likelihoods as follows:

$$p\left(\mathbf{x} \mid C = 1\right) = \mathcal{N}\left(\mu^1 = \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \mathbf{\Sigma}^1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$$

$$p\left(\mathbf{x} \mid C = 2\right) = \mathcal{N}\left(\mu^2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \mathbf{\Sigma}^2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$$

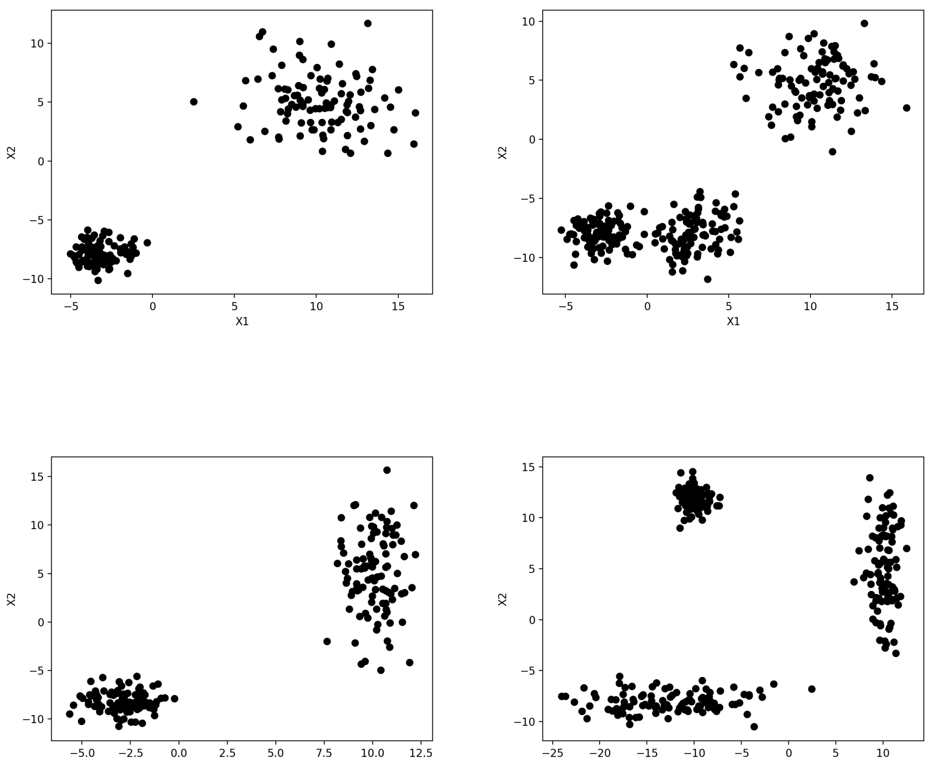
Also, initialize the priors as follows:

$$p\left(C=1\right)=0.6$$

$$p\left(C=2\right)=0.4$$

a) Perform one expectation maximization iteration.
b) Plot the points and sketch the clusters.
c) Verify that after one iteration the probability of the data increased.

4) Consider the following four scenarios of plotted data sets:



a) For each scenario justify wether or not k-means would be suitable.

b) Assuming you apply EM clustering to model all scenarios what would the means and covariances look like? For simplicity, assume all covariance matrices are diagonal.

3 Thinking Questions

a) Think about what measures would be desirable in a clustering. Think about distance between centroids and intra-cluster distances.

b) Think about initialization mechanisms and how they affect the final clustering.

c) Why do we need all those clustering indices?

d) Why do we need the covariance in EM? What is the difference between k-means and EM in terms of cluster shapes we can capture.

e) Is k-means really a kind of EM clustering?