Machine Learning Homework Sheet 11

Clustering

1 Gaussian Mixture Model

Problem 1: Consider a mixture of K Gaussians

$$p(\boldsymbol{x}) = \sum_k \pi_k \mathcal{N}(\boldsymbol{x} \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k).$$

Derive the expected value $\mathbb{E}[\boldsymbol{x}]$ and the covariance $\mathrm{Cov}[\boldsymbol{x}]$.

Hint: it is helpful to remember the identity $\mathrm{Cov}[\boldsymbol{x}] = \mathbb{E}[\boldsymbol{x}\boldsymbol{x}^T] - \mathbb{E}[\boldsymbol{x}]\mathbb{E}[\boldsymbol{x}]^T$.

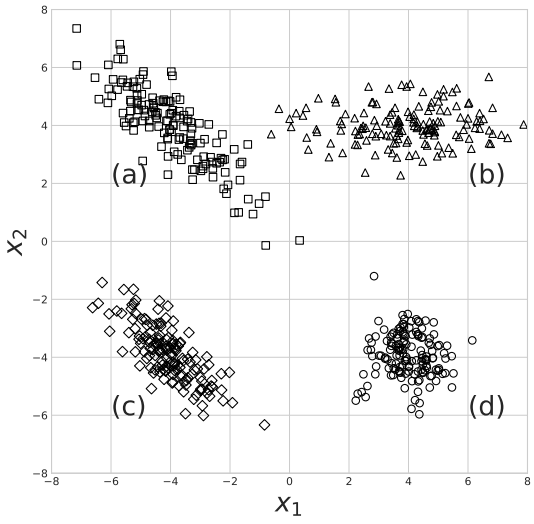
Problem 2: Consider a mixture of K isotropic Gaussians, all with the same *known* covariances $\boldsymbol{\Sigma}_k = \sigma^2 \boldsymbol{I}$.

Derive the EM algorithm for the case when $\sigma^2 \to 0$, and show that it's equivalent to Lloyd's algorithm for K-means.

Problem 3: The dataset displayed on the right has been generated using a Gaussian mixture model with $K = 4$ components, each with its own mean $\boldsymbol{\mu}_k$ and covariance matrix $\boldsymbol{\Sigma}_k$.

Match the covariance matrices in the table on the left with their corresponding Gaussian components in the plot on the right. Explain each of the answers with 1 sentence.

$\boldsymbol{\Sigma}_k$	Cluster
$\begin{bmatrix} 2 & -1.7 \\ -1.7 & 2 \end{bmatrix}$	
$\begin{bmatrix} 0.9 & -0.8 \\ -0.8 & 1.2 \end{bmatrix}$	
$\begin{bmatrix} 3 & 0 \\ 0 & 0.5 \end{bmatrix}$	
$\begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}$	



Problem 4:

a) Given is the dataset displayed in the figure below. Apply the K-means algorithm to this data using $K = 2$ and using the circled points as initial centroids.

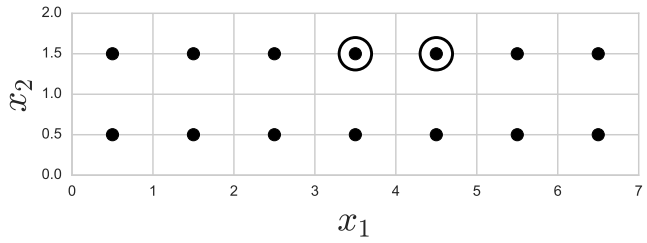


Figure 1: K-Means Dataset

What are the clusters after K-Means converges? Draw your solution in the figure above, i.e. mark the location of the centroids with $\times$'s and show the clusters by drawing two bounding boxes around the points assigned to each cluster.

How many iterations did it take for K-Means to converge in the above problem?

b) Provide a different initialization, for which the algorithm will take **more** iterations to converge to the **same** solution. Make sure that your initialization does not lead to ties. Draw your initialization in the figure below.

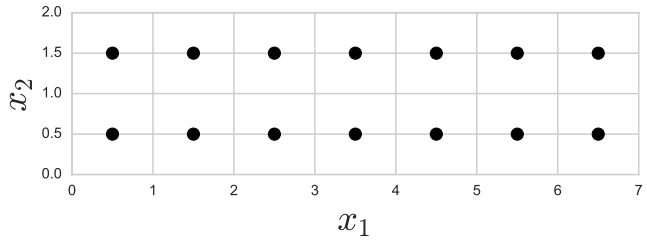


Figure 2: Provide your initialization