Problem 1

Tuesday, 29 January 2019 11:22

Problem 1: Consider a mixture of K Gaussians

PDF $$p(\boldsymbol{x}) = \sum_k \pi_k \mathcal{N}(\boldsymbol{x} \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \sum_{k=1}^{K} p(z=k)\, p(x \mid z=k)$$

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

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

$$\mathbb{E}[x] = \int x\, p(x)\, dx = \int x \sum_{k=1}^{K} \pi_k \mathcal{N}(x \mid \theta_k)\, dx =$$

$$\sum_{k=1}^{K} \pi_k \underbrace{\int x\, \mathcal{N}(x \mid \theta_k)\, dx}_{\mathbb{E}[x \mid z=k] = \mu_k} = \sum_{k=1}^{K} \pi_k \mu_k = \mu$$

Remark: $$\underbrace{\mathbb{E}[g(X)]}_{\int g(x)\, p(x)\, dx} = \underbrace{\mathbb{E}_z\left[\mathbb{E}_{X|z}[g(X) \mid z]\right]}_{\sum_k \pi_k \cdot \mathbb{E}[g(X) \mid z=k]}$$

$$\mathrm{Cov}(x) = \mathbb{E}[x x^T] - \underbrace{\mathbb{E}[x]}_{\mu} \cdot \underbrace{\mathbb{E}[x]^T}_{\mu^T}$$

$$\mathbb{E}_X[x x^T] = \mathbb{E}_z\left[\underbrace{\mathbb{E}_{X|z}[x x^T \mid z]}\right] = \sum_{k=1}^{K} \pi_k \left(\Sigma_k + \mu_k \mu_k^T\right)$$

$$\rightarrow \quad \Sigma_k = \mathbb{E}[x x^T \mid z=k] - \underbrace{\mathbb{E}[x \mid z=k]}_{\mu_k} \cdot \underbrace{\mathbb{E}[x \mid z=k]^T}_{\mu_k^T}$$

μ_k μ_k^T

$$\widehat{Cov}(X) = \sum_k \pi_k \left(\Sigma_k + \mu_k \mu_k^T\right) - \sum_k \underbrace{\sum_j \pi_k \pi_j \mu_k \mu_j^T}_{\mu \mu^T} =$$

$$= \sum_k \pi_k \; \Sigma_k - \sum_k \sum_j \pi_k \left(\pi_j - 1_{k=j}\right) \mu_k \mu_j^T$$

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.

Z-update step for k-means (Reminder)

$$z_{ik} = \begin{cases} 1 & \text{if } k = \arg\max_j \|x_i - \mu_j\|^2 \\ 0 & \text{else} \end{cases}$$

GMM, E-step:

μ_k, Σ_k $\quad \Sigma_k = \sigma^2 I$

$$\gamma(z_{ik}) = \frac{\pi_k \mathcal{N}(x_i \mid \theta_k)}{\mathcal{N}} = \frac{\pi_k \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left(\frac{-\|x_i - \mu_k\|^2}{2\sigma^2}\right)}{\sum_{j=1}^{K} \pi_j \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left(\frac{-\|x_i - \mu_j\|^2}{2\sigma^2}\right)}$$

$$= \frac{1}{1 + \sum_{j \neq k} \frac{\pi_j}{\pi_k} \exp\left(\underbrace{\frac{-\|x_i - \mu_j\|^2 + \|x_i - \mu_k\|^2}{2\sigma^2}}\right)}$$

Case 1: for all j $\|x_i - \mu_k\| \leq \|x_i - \mu_j\|$ $\Rightarrow$

$$0 \geq \frac{-\|x_i - \mu_j\|^2 + \|x_i - \mu_k\|^2}{2\sigma^2} \xrightarrow{\sigma \to 0} -\infty \Rightarrow$$

$$\exp(\ \_\_\_\_ \ ((\ \_\_\_\_\_\_ \ ) \xrightarrow{\sigma \to 0} 0$$

$$\Rightarrow \gamma(z_{ik}) \xrightarrow{\sigma \to 0} 1$$

Case 2: for some j $\|x_i - \mu_j\| < \|x_i - \mu_k\|$ $\Rightarrow$

$$0 < \frac{-\|x_i - \mu_j\|^2 + \|x_i - \mu_k\|^2}{2\sigma^2} \xrightarrow{\sigma \to 0} +\infty \Rightarrow$$

$$\exp\left(\text{———} \;\| \text{———} \right) \xrightarrow{\sigma \to 0} +\infty \Rightarrow$$

$$\gamma(z_{ik}) \xrightarrow{\sigma \to 0} 0$$

M-step:

$$N_k = \sum_{i=1}^{N} \gamma(z_{ik}) \longrightarrow \# \text{ samples } i, \text{ s.t. } \gamma(z_{ik}) = 1$$

$$\mu_k^{new} = \frac{1}{N_k} \sum_{i=1}^{N} \gamma(z_{ik}) x_i \longrightarrow \text{mean of } \{x_i \mid \gamma(z_{ik}) = 1\}$$

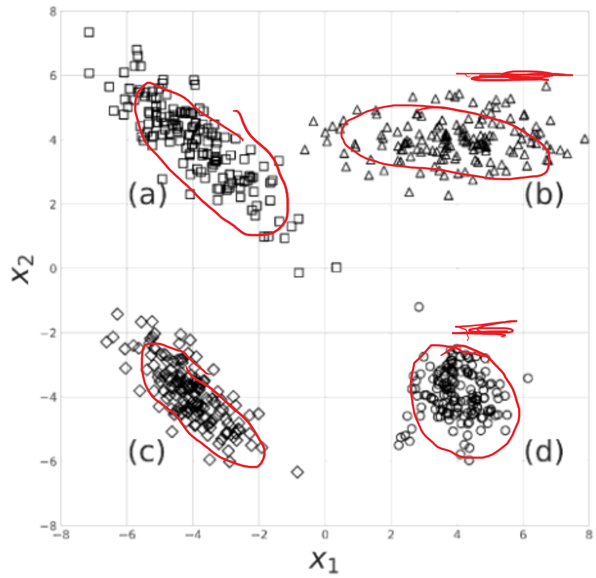
Problem 3

Tuesday, 29 January 2019 11:24

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**
$\begin{bmatrix} 2 & -1.7 \\ -1.7 & 2 \end{bmatrix}$	a
$\begin{bmatrix} 0.9 & -0.8 \\ -0.8 & 1.2 \end{bmatrix}$	c
$\begin{bmatrix} 3 & 0 \\ 0 & 0.5 \end{bmatrix}$	b
$\begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}$	d



Problem 4

Tuesday, 29 January 2019 11:24

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.

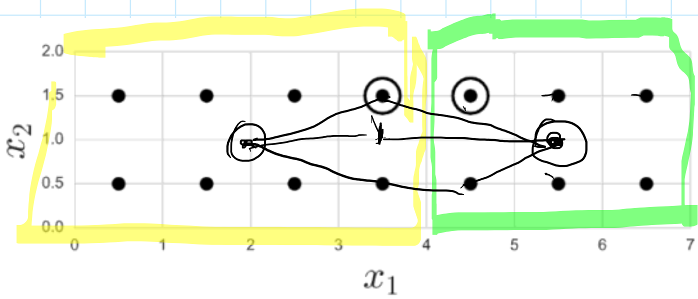


Figure 1: K-Means Dataset

What are the clusters after K-Means converges? Draw your solution in the figure above, i.e. mark the location of the centroids with ×'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.

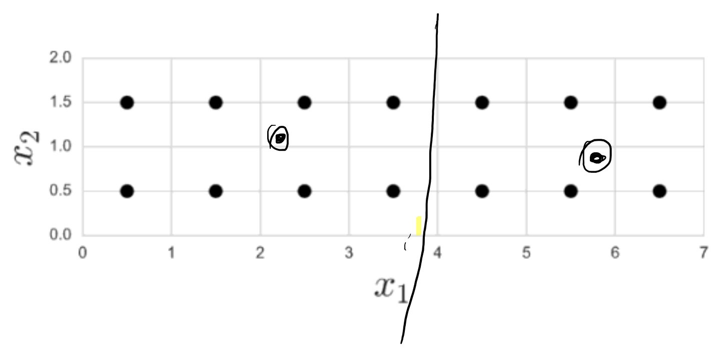


Figure 2: Provide your initialization