Machine Learning Homework 11

Author: Jesus Arturo Sol Navarro

1 Problem 5

- Expeted value

$$E[\mathbf{x}] = \int \mathbf{x} \sum_k \pi_k \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) d\mathbf{x} = \sum_k \pi_k \int \mathbf{x} \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) d\mathbf{x}$$

  Expectation of a gaussian distribution is it's mean, thus

$$E[\mathbf{x}] = \sum_k \pi_k \mu_k$$

- Covariance

$$E[\mathbf{x}\mathbf{x}^T] = \int \mathbf{x}\mathbf{x}^T p(\mathbf{x})\, d\mathbf{x}. = \sum_k \pi_k \int \mathbf{x}\mathbf{x}^T \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) d\mathbf{x} = \sum_k \pi_k \left(\boldsymbol{\Sigma}_k + \boldsymbol{\mu}_k \boldsymbol{\mu}_k^T\right)$$

  From the identity provided and the results from the expected value:

$$Cov[x] = \sum_k \pi_k \left(\boldsymbol{\Sigma}_k + \boldsymbol{\mu}_k \boldsymbol{\mu}_k^T\right) - (\sum_k \pi_k \mu_k)(\sum_k \pi_k \mu_k)^T$$

2 Problem 6

- E Step

$$\gamma_t(z_{ik}) = \frac{\pi_k^{(t)} \mathcal{N}(x_i | \mu_k^{(t)}, \sigma^2 I)}{\sum_{j=1}^K \pi_j^{(t)} \mathcal{N}(x_i | \mu_j^{(t)}, \sigma^2 I)}.$$

- M Step

$$\pi_k^{(t+1)} = \frac{N_k}{N} \quad \text{where} \quad N_k = \sum_{i=1}^N \gamma_t(z_{ik}).$$

$$\mu_k^{(t+1)} = \frac{1}{N_k} \sum_{i=1}^N \gamma_t(z_{ik}) x_i$$

- If $\sigma^2 I$ is close to zero, then all the gaussian become peaky around the mean. Each point is asigned to the closest cluster with probability really close to 1, in other words $\gamma_t(z_{ik}) \approx 1$ if the squared euclidean distance between the data point and the nearest centroid is the smallest in comparison to other centroids.

$$\gamma_t(z_{ik}) = z_{ik} = \begin{cases} 1 & \text{if } k = \arg\min_j \|x_i - \mu_j\|^2, \\ 0 & \text{else.} \end{cases}$$

$$\mu_k^{(t+1)} = \frac{1}{N_k} \sum_{i=1}^N z_{ik} x_i \quad \text{where} \quad N_k = \sum_{i=1}^N z_{ik}.$$

3 Problem 7

- a)
  - 1) Draw a one-hot cluster indicator $z_x \sim \text{Cat}(\pi_k^x)$.
  - 2) Draw a one-hot cluster indicator $z_y \sim \text{Cat}(\pi_l^y)$.
  - 3) Sample $x \sim \mathcal{N}(\mu_k^x, \Sigma_k^x)$ and $y \sim \mathcal{N}(\mu_l^y, \Sigma_l^y)$.
  - 4) z = x + y

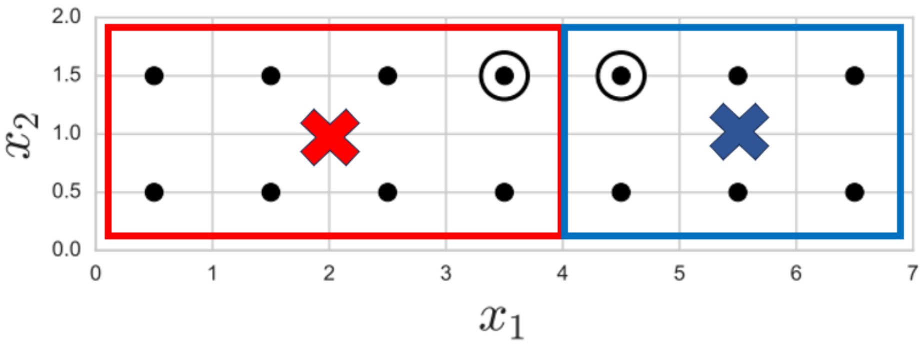
- b) The distribution $p(z \mid \theta^x, \theta^y)$ remains a mixture of Gaussians because the sum of two independent Gaussian random variables is still Gaussian.The resulting Gaussian Mixture Model has $K_x K_y$ individual gaussians distributions, where the mean and covariance of each new distribution is the sum of the respective means and covariances.

- c)

$$p(\mathbf{z} \mid \boldsymbol{\theta}^x, \boldsymbol{\theta}^y) = \sum_{k=1}^{K_x} \sum_{l=1}^{K_y} \pi_k^x \pi_l^y \mathcal{N}(\mathbf{z} \mid \boldsymbol{\mu}_k^x + \boldsymbol{\mu}_l^y, \boldsymbol{\Sigma}_k^x + \boldsymbol{\Sigma}_l^y). \tag{1}$$

4 Problem 8

- a) The K-Means algorithm converged in just one iteration to the centroids $\mu_1 = (2, 1)$ and $\mu_2 = (5.5, 1)$.



- b) If the centroids are initialized as follow $\mu_1 = (1.5, 0)$ and $\mu_2 = (6, 2)$. It takes two iterations to converge to the centroids found in the last section.