CX4240 Computing for Data Analysis - Practice Midterm Exam

Name:
GTID:

Instructor: Bo Dai

- The exam consists of four problems, **each worth 25 points**. **In this practice exam, we only provide two sample problems; but in midterm exam, there will be four problems in total.**
- This is a **closed-book exam**. No external resources or communication with others is allowed. You are allowed to bring **a double-sided US-letter-size cheatsheet**.
- By submitting this exam, you confirm that you have upheld the Georgia Tech Honor Code.

1 Gaussian Mixture Models and EM Algorithm [25 pts]

Notation:

- $\pi_k, \boldsymbol{\mu}_k, \Sigma_k$: Mixing coefficient of the mixtures (prior), mean, and covariance of component k.
- γ_{nk}: Responsibility (posterior probability) of component k for data point $\boldsymbol{x}_n$.

Given: You are running the Expectation-Maximization (EM) algorithm for a Gaussian Mixture Model. In the current iteration, you have three data points:

- Data points: $\boldsymbol{x}_1 = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$, $\boldsymbol{x}_2 = \begin{pmatrix} 4 \\ 4 \end{pmatrix}$, $\boldsymbol{x}_3 = \begin{pmatrix} 6 \\ 2 \end{pmatrix}$.

After the E-step, the calculated responsibilities for component $k = 2$ are:

- Responsibilities: $\gamma_{12} = 0.2, \gamma_{22} = 0.6, \gamma_{32} = 0.2$.

The mean update formula in the M-step is:

$$\boldsymbol{\mu}_k = \frac{\sum_{n=1}^{N} \gamma_{nk} \boldsymbol{x}_n}{\sum_{n=1}^{N} \gamma_{nk}}. \tag{1}$$

The covariance update formula in the M-step is:

$$\Sigma_k = \frac{\sum_{n=1}^{N} \gamma_{nk} (\boldsymbol{x}_n - \boldsymbol{\mu}_k)(\boldsymbol{x}_n - \boldsymbol{\mu}_k)^\top}{\sum_{n=1}^{N} \gamma_{nk}}. \tag{2}$$

Questions:

1. For GMM, write down the likelihood (log-probability) $\log p(\boldsymbol{x})$ of a given data $\boldsymbol{x}$ using the parameters. You can assume there are K components in total.
2. Select all statements about GMM that are correct?

   A. In E-step, we fix the responsibilities and update the means, covariances, and mixing weights to maximize the log-likelihood.

   B. In M-step, we fix the means, covariances and mixing weights to calculate the responsibilities.

   C. The summation of the responsibilities over the samples $\sum_{n=1}^{N} \gamma_{nk}$ should equals to 1.

   D. The mixing coefficients (priors) π_k updated in the M-step must sum to 1 across all K components.

   E. All above statements are incorrect.
3. Please calculate the updated mean and covariance matrix for component $k = 2$.

Solution:

1. $$\log p(\boldsymbol{x}|\{\pi_k\}_{k=1}^K, \{\boldsymbol{\mu}_k\}_{k=1}^K, \{\Sigma_k\}_{k=1}^K) = \log\left(\sum_{k=1}^K \pi_k \mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}_k, \Sigma_k)\right) \tag{3}$$

2. The correct statement is D.

   A. Incorrect: In the E-step, we fix the parameters and update the responsibilities γ_{nk}.

   B. Incorrect: In the M-step, we update the parameters using the fixed responsibilities calculated in the E-step.

   C. Incorrect: The sum of responsibilities for a single data point across all components equals 1 ($\sum_k \gamma_{nk} = 1$). However, The sum $\sum_n \gamma_{nk}$ represents the total "effective number of points" assigned to cluster k, and it does not have to be 1.

   D. Correct: The mixing coefficients π_k represent a probability distribution (priors) and must sum to 1 over all K components.

3. First let's calculate the effective number of points N_2:

   $$N_2 = \sum_{n=1}^{3} \gamma_{n2} = 0.2 + 0.6 + 0.2 = 1.0 \tag{4}$$

   Update the mean vector $\boldsymbol{\mu}_2$:

   $$\boldsymbol{\mu}_2 = \frac{1}{N_2}\sum_{n=1}^{3} \gamma_{n2}\mathbf{x}_n \tag{5}$$

   $$\boldsymbol{\mu}_2 = \frac{1}{1.0}\left[0.2\begin{pmatrix}2\\2\end{pmatrix} + 0.6\begin{pmatrix}4\\4\end{pmatrix} + 0.2\begin{pmatrix}6\\2\end{pmatrix}\right] \tag{6}$$

   $$\boldsymbol{\mu}_2 = \begin{pmatrix}0.4+2.4+1.2\\0.4+2.4+0.4\end{pmatrix} = \begin{pmatrix}4.0\\3.2\end{pmatrix} \tag{7}$$

   Update the covariance matrix $\boldsymbol{\Sigma}_2$:

   $$\boldsymbol{\Sigma}_2 = \frac{1}{N_2}\sum_{n=1}^{3} \gamma_{n2}(\mathbf{x}_n - \boldsymbol{\mu}_2)(\mathbf{x}_n - \boldsymbol{\mu}_2)^\top \tag{8}$$

   Calculating each term $(\mathbf{x}_n - \boldsymbol{\mu}_2)(\mathbf{x}_n - \boldsymbol{\mu}_2)^\top$ weighted by γ_{n2}:

   - $n = 1 : 0.2\begin{pmatrix}2-4\\2-3.2\end{pmatrix}\begin{pmatrix}-2 & -1.2\end{pmatrix} = 0.2\begin{pmatrix}4 & 2.4\\2.4 & 1.44\end{pmatrix} = \begin{pmatrix}0.8 & 0.48\\0.48 & 0.288\end{pmatrix}$
   - $n = 2 : 0.6\begin{pmatrix}4-4\\4-3.2\end{pmatrix}\begin{pmatrix}0 & 0.8\end{pmatrix} = 0.6\begin{pmatrix}0 & 0\\0 & 0.64\end{pmatrix} = \begin{pmatrix}0 & 0\\0 & 0.384\end{pmatrix}$
   - $n = 3 : 0.2\begin{pmatrix}6-4\\2-3.2\end{pmatrix}\begin{pmatrix}2 & -1.2\end{pmatrix} = 0.2\begin{pmatrix}4 & -2.4\\-2.4 & 1.44\end{pmatrix} = \begin{pmatrix}0.8 & -0.48\\-0.48 & 0.288\end{pmatrix}$

Summing these terms:

$$\boldsymbol{\Sigma}_2 = \begin{pmatrix} 0.8 + 0 + 0.8 & 0.48 + 0 - 0.48 \\ 0.48 + 0 - 0.48 & 0.288 + 0.384 + 0.288 \end{pmatrix} \tag{9}$$

$$\boldsymbol{\Sigma}_2 = \begin{pmatrix} 1.6 & 0 \\ 0 & 0.96 \end{pmatrix} \tag{10}$$

Final Results:

$$\boldsymbol{\mu}_2 = \begin{pmatrix} 4 \\ 3.2 \end{pmatrix}, \quad \boldsymbol{\Sigma}_2 = \begin{pmatrix} 1.6 & 0 \\ 0 & 0.96 \end{pmatrix} \tag{11}$$

2 Linear Regression for Apartment Price Prediction [25 pts]

Suppose we want to predict the monthly rent of an apartment in Atlanta using a linear model. For each apartment, we consider two features:

$$x_1 = \text{living area (in hundreds of ft}^2\text{)}, \qquad x_2 = \text{number of bedrooms.}$$

Assume the rent follows the linear regression model

$$y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \varepsilon, \qquad \varepsilon \sim \mathcal{N}(0, \sigma^2),$$

where ε is a random noise term that captures variability in rent prices not explained by the linear model. We collect the following training dataset consisting of $n = 4$ samples:

x_1	x_2	y
1	1	2.5
2	1	3.5
1	2	3.0
2	2	4.0

Use *least squares regression* to estimate the parameter vector

$$\vec{\theta} = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{bmatrix}.$$

Hint I. For linear regression with an intercept term, we augment each feature vector with a leading 1. Stacking the augmented feature vectors forms the *design matrix*

$$X = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \\ 1 & 2 & 2 \end{pmatrix}.$$

Hint II. The least squares estimator minimizes the squared error

$$\min_{\theta} \|y - X\theta\|_2^2,$$

whose closed-form solution is given by the normal equation

$$\theta^* = (X^T X)^{-1} X^T y.$$

Solution:
The least squares estimator is given by

$$\theta^* = (X^T X)^{-1} X^T y.$$

First compute

$$X^T X = \begin{pmatrix} 4 & 6 & 6 \\ 6 & 10 & 9 \\ 6 & 9 & 10 \end{pmatrix}, \qquad X^T y = \begin{pmatrix} 13 \\ 20.5 \\ 20 \end{pmatrix}.$$

Thus,

$$\theta^* = \begin{pmatrix} 4 & 6 & 6 \\ 6 & 10 & 9 \\ 6 & 9 & 10 \end{pmatrix}^{-1} \begin{pmatrix} 13 \\ 20.5 \\ 20 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 0.5 \end{pmatrix}.$$

Final answer:

$$\vec{\theta} = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 0.5 \end{bmatrix}.$$