CS4641 Machine Learning - Practice Problems

Instructor: Bo Dai

March 5, 2025

Name: ____________________ **GT-ID:** ____________________

Please read the following instructions carefully.

- The exam consists of five problems, each worth 25 points. You are required to **complete four out of the five problems**. If you answer all five, only the top four highest-scoring problems will be counted.
- This is a **closed-book exam. No notes, external resources, or communication with others is allowed**.
- By submitting this exam, you confirm that you have upheld the Georgia Tech Honor Code.

Name: ____________________ **GT-ID:** ____________________

1 GMM M-Step Covariance Derivation

Notation:

- μ_l, Σ_l: Mean and covariance of component l
- y_j^l: Responsibility of component l for data point x_j

Given:

- Data points: $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, $\mathbf{x}_2 = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$
- Responsibilities: $y_1^1 = 0.7$, $y_2^1 = 0.3$
- Current mean: $\boldsymbol{\mu}_1 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$

The covariance update formula is:

$$\boldsymbol{\Sigma}_k = \frac{\sum_{i=1}^{N} y_i^k (\mathbf{x}_i - \boldsymbol{\mu}_k)(\mathbf{x}_i - \boldsymbol{\mu}_k)^\top}{\sum_{i=1}^{N} y_i^k}$$

(a) Please calculate the updated covariance matrix $\boldsymbol{\Sigma}_1$.

(b) True/False Assuming two components, $y_1^1 + y_1^2$ sums the responsibilities of different components for the same data point $\mathbf{x}_1$. By definition of mixture-model responsibilities, for each data point, the responsibilities across all components must sum to 1.

(c) Which of the following statements about the GMM M-step is correct?

A. In the M-step, we fix the mean and covariance and update only the prior mixing coefficients.

B. In the M-step, we update the mean and covariance of each component to maximize the likelihood given the current responsibilities.

C. In the M-step, responsibilities are recalculated based on the current parameters.

D. In the M-step, we do not change any parameters but only check convergence criteria.

(d) In your own words, describe how you might decide to use GMMs instead of K-Means in a real-world application.

2 K-Means Within-Cluster Sum of Squares Calculation

WCSS Definition

The **Within-Cluster Sum of Squares (WCSS)**, often used to measure the cohesion of clusters in a clustering algorithm is defined as:

$$\text{WCSS} = \sum_{i=1}^{k} \sum_{x \in C_i} \|x - \mu_i\|^2$$

where:

- C_i: Set of points in cluster i
- μ_i: Centroid of cluster i
- $\|x - \mu_i\|^2$: Squared Euclidean distance between point x and centroid μ_i

Initial State

- Points: $A(1, 1)$, $B(2, 2)$, $C(5, 5)$, $D(6, 6)$
- Initial centroids: $\boldsymbol{\mu}_1^{(0)} = (1.5, 1.5)$, $\boldsymbol{\mu}_2^{(0)} = (4, 4)$
- Initial assignments: $\{A, B\} \rightarrow$ Cluster 1, $\{C, D\} \rightarrow$ Cluster 2

Part 1: Compute Initial WCSS

Given the initial state, what is the value of the Within-Cluster Sum of Squares metric?

Part 2: Compute WCSS after Updating Centroids and Re-assigning Clusters

Following K-Means algorithm, we update the centroids, and re-assign points to clusters.

$$\boldsymbol{\mu}_1^{(1)} = (1.5, 1.5), \quad \boldsymbol{\mu}_2^{(1)} = (5.5, 5.5)$$

Clearly, assignments remain unchanged: $\{A, B\} \rightarrow$ Cluster 1, $\{C, D\} \rightarrow$ Cluster 2.

Given the new state after performing the updates and assignments, what is the value of the Within-Cluster Sum of Squares metric?

Name: ________________ **GT-ID:** ________________

Solutions

Problem 1: GMM M-Step Covariance Derivation

Notation:

- μ_l, Σ_l: Mean and covariance of component l
- y_j^l: Responsibility of component l for data point x_j

Given:

- Data points: $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, $\mathbf{x}_2 = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$
- Responsibilities: $y_1^1 = 0.7$, $y_2^1 = 0.3$
- Current mean: $\boldsymbol{\mu}_1 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$

Please derive the updated covariance matrix $\boldsymbol{\Sigma}_1$.

Solution:

(a) The covariance update formula is:

$$\boldsymbol{\Sigma}_k = \frac{\sum_{i=1}^{N} y_i^k (\mathbf{x}_i - \boldsymbol{\mu}_k)(\mathbf{x}_i - \boldsymbol{\mu}_k)^\top}{\sum_{i=1}^{N} y_i^k}$$

1. **Compute deviations:**

$$\mathbf{x}_1 - \boldsymbol{\mu}_1 = \begin{bmatrix} -1 \\ -1 \end{bmatrix}, \quad \mathbf{x}_2 - \boldsymbol{\mu}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

2. **Compute outer products:**

$$(\mathbf{x}_1 - \boldsymbol{\mu}_1)(\mathbf{x}_1 - \boldsymbol{\mu}_1)^\top = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad (\mathbf{x}_2 - \boldsymbol{\mu}_1)(\mathbf{x}_2 - \boldsymbol{\mu}_1)^\top = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

3. **Weight and sum:**

$$\boldsymbol{\Sigma}_1 = \frac{0.7 \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} + 0.3 \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}}{0.7 + 0.3} = \boxed{\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}}$$

(b) $y_1^1 + y_1^2$ sums the responsibilities of different components for the same data point $\mathbf{x}_1$. By definition of mixture-model responsibilities, for each single data point, the responsibilities across all components must sum to 1. So yes, $y_1^1 + y_1^2 = 1$.

(c) The correct statement is B.

(d) If you suspect clusters have different shapes or if you want soft assignments (e.g., real-world scenarios where a data point might reasonably belong to multiple categories), GMMs provide more nuanced modeling. K-Means can be simpler and faster but is less flexible.

Name: ____________________ **GT-ID:** ____________________

Problem 2: K-Means Within-Cluster Sum of Squares Calculation

WCSS Definition

The **Within-Cluster Sum of Squares (WCSS)**, often used to measure the cohesion of clusters in a clustering algorithm is defined as:

$$\text{WCSS} = \sum_{i=1}^{k} \sum_{x \in C_i} \|x - \mu_i\|^2$$

where:

- C_i: Set of points in cluster i
- μ_i: Centroid of cluster i
- $\|x - \mu_i\|^2$: Squared Euclidean distance between point x and centroid μ_i

Initial State

- Points: $A(1,1)$, $B(2,2)$, $C(5,5)$, $D(6,6)$
- Initial centroids: $\boldsymbol{\mu}_1^{(0)} = (1.5, 1.5)$, $\boldsymbol{\mu}_2^{(0)} = (4, 4)$
- Initial assignments: $\{A, B\} \to$ Cluster 1, $\{C, D\} \to$ Cluster 2

Task 1: Compute Initial WCSS

Given the initial state, what is the value of the Within-Cluster Sum of Squares metric?

Solution:

$$\begin{aligned}
\text{WCSS}^{(0)} &= \sum_{\text{Cluster 1}} \|\mathbf{x} - \boldsymbol{\mu}_1^{(0)}\|^2 + \sum_{\text{Cluster 2}} \|\mathbf{x} - \boldsymbol{\mu}_2^{(0)}\|^2 \\
&= \left[(1-1.5)^2 + (1-1.5)^2 + (2-1.5)^2 + (2-1.5)^2\right] \\
&\quad + \left[(5-4)^2 + (5-4)^2 + (6-4)^2 + (6-4)^2\right] \\
&= (0.25 + 0.25 + 0.25 + 0.25) + (1 + 1 + 4 + 4) = \boxed{10}
\end{aligned}$$

Task 2: Compute WCSS after Updating Centroids and Re-assigning Clusters

Following K-Means algorithm, we update the centroids, and re-assign points to clusters.

$$\boldsymbol{\mu}_1^{(1)} = (1.5, 1.5), \quad \boldsymbol{\mu}_2^{(1)} = (5.5, 5.5)$$

Clearly, assignments remain unchanged: $\{A, B\} \rightarrow$ Cluster 1, $\{C, D\} \rightarrow$ Cluster 2.

Given the new state after performing the updates and assignments, what is the value of the Within-Cluster Sum of Squares metric?

Solution:

$$\begin{aligned}
\text{WCSS}^{(1)} &= \sum_{\text{Cluster 1}} \|\mathbf{x} - \boldsymbol{\mu}_1^{(1)}\|^2 + \sum_{\text{Cluster 2}} \|\mathbf{x} - \boldsymbol{\mu}_2^{(1)}\|^2 \\
&= \left[(1-1.5)^2 + (1-1.5)^2 + (2-1.5)^2 + (2-1.5)^2\right] \\
&\quad + \left[(5-5.5)^2 + (5-5.5)^2 + (6-5.5)^2 + (6-5.5)^2\right] \\
&= (0.25 + 0.25 + 0.25 + 0.25) + (0.25 + 0.25 + 0.25 + 0.25) = \boxed{2}
\end{aligned}$$