CS/DSC/AI 391L: Machine Learning

Homework 5 - Theory

Lecture: Prof. Qiang Liu

Problem 1: Gaussian Graphical Models

(a) Are X_3 and X_4 correlated?

Answer: No, X_3 and X_4 are **uncorrelated**.

Explanation: In the covariance matrix Σ, the covariance between X_3 and X_4 is given by the entry Σ_{34}. Looking at Σ:

$$\Sigma_{34} = 0$$

Since $\Sigma_{34} = 0$, the covariance between X_3 and X_4 is zero, indicating that they are uncorrelated.

(b) Are X_3 and X_4 conditionally correlated given X_1 and X_2?

Answer: No, X_3 and X_4 are **conditionally uncorrelated** given X_1 and X_2; that is, $\text{cov}(X_3, X_4 \mid X_1, X_2) = 0$

Explanation: To determine conditional correlation, we need to: 1. Extract the relevant submatrix from Σ^{-1} 2. Check if the corresponding entry is zero

Looking at Σ^{-1}, the entry corresponding to X_3 and X_4 is zero, indicating conditional independence.

(c) What is the Markov blanket of X_2?

Answer: The Markov blanket of X_2 is $\{X_1\}$

Explanation: The Markov blanket consists of: 1. Parents 2. Children 3. Other parents of the children

Looking at Σ^{-1}, X_2 is only connected to X_1, making X_1 its only neighbor in the graph.

(d) Compute Σ_Y

Given:

$$\Sigma_Y = \begin{bmatrix} \Sigma_{11} + 0.2 & \Sigma_{12} - 0.2 \\ \Sigma_{12} - 0.2 & \Sigma_{22} + 0.2 \end{bmatrix}$$

Substitute the values from Σ:

$$\Sigma_Y = \begin{bmatrix} 0.71 + 0.2 & -0.43 - 0.2 \\ -0.43 - 0.2 & 0.46 + 0.2 \end{bmatrix} = \begin{bmatrix} 0.91 & -0.63 \\ -0.63 & 0.66 \end{bmatrix}$$

Problem 2: Expectation Maximization (EM)

(a) Next Iteration Values of μ_1 and μ_2

Given:

- Data points: $x^{(1)} = -1$ and $x^{(2)} = 1$
- Initial means: $\mu_1^{(0)} = -2$ and $\mu_2^{(0)} = 2$
- Gaussian components: $N(\mu_1, 1)$ and $N(\mu_2, 1)$

E-step: Computing responsibilities γ_{ij}, the probability that point $x^{(i)}$ belongs to component j.

For $x^{(1)} = -1$:

$$\begin{aligned}
\gamma_{11} &= \frac{N(x^{(1)}|\mu_1^{(0)}, 1)}{N(x^{(1)}|\mu_1^{(0)}, 1) + N(x^{(1)}|\mu_2^{(0)}, 1)} \\
&= \frac{e^{-\frac{(-1-(-2))^2}{2}}}{e^{-\frac{(-1-(-2))^2}{2}} + e^{-\frac{(-1-2)^2}{2}}} \\
&= \frac{e^{-0.5}}{e^{-0.5} + e^{-4.5}} \\
&= \frac{1}{1 + e^{-4}} \approx 0.98 \\
\gamma_{12} &= 1 - \gamma_{11} \approx 0.02
\end{aligned}$$

Similarly, for $x^{(2)} = 1$:

$$\begin{aligned}
\gamma_{21} &= \frac{N(x^{(2)}|\mu_1^{(0)}, 1)}{N(x^{(2)}|\mu_1^{(0)}, 1) + N(x^{(2)}|\mu_2^{(0)}, 1)} \\
&= \frac{e^{-\frac{(1-(-2))^2}{2}}}{e^{-\frac{(1-(-2))^2}{2}} + e^{-\frac{(1-2)^2}{2}}} \\
&= \frac{e^{-4.5}}{e^{-4.5} + e^{-0.5}} \\
&= \frac{1}{1 + e^{4}} \approx 0.02 \\
\gamma_{22} &= 1 - \gamma_{21} \approx 0.98
\end{aligned}$$

M-step: Update means using weighted averages:

$$\begin{aligned}\mu_1^{(1)} &= \frac{\gamma_{11}x^{(1)} + \gamma_{21}x^{(2)}}{\gamma_{11} + \gamma_{21}} \\ &= \frac{0.98(-1) + 0.02(1)}{0.98 + 0.02} \\ &= -0.96\end{aligned}$$

$$\begin{aligned}\mu_2^{(1)} &= \frac{\gamma_{12}x^{(1)} + \gamma_{22}x^{(2)}}{\gamma_{12} + \gamma_{22}} \\ &= \frac{0.02(-1) + 0.98(1)}{0.02 + 0.98} \\ &= 0.96\end{aligned}$$

Answer: Next iteration yields $\mu_1 = -0.96$, $\mu_2 = 0.96$

(b) Convergence Analysis

The EM algorithm will not converge to $\mu_1 = -1$ and $\mu_2 = 1$ because:

- The responsibilities are always between 0 and 1
- Each new mean is a weighted average of both data points
- This causes the means to move closer together in each iteration
- Eventually, both means will converge to 0 (the average of the data points)

(c) Fixed Point Analysis

Starting from $\mu_1 = \mu_2 = 2$:

- When both means are equal, each point has equal responsibility (0.5) for each component
- In the M-step, both means will update to the average of all points:

$$\mu_1 = \mu_2 = \frac{0.5(-1) + 0.5(1)}{0.5 + 0.5} = 0$$

- Once the means are both 0, they will remain there in subsequent iterations

Answer: The fixed point is $\mu_1 = \mu_2 = 0$

(d) K-means Fixed Point

K-means Steps:

1. **Assignment Step:**
   - $x^{(1)} = -1$ is closer to $\mu_1 = -2$
   - $x^{(2)} = 1$ is closer to $\mu_2 = 2$
2. **Update Step:**
   - $\mu_1 =$ mean of points assigned to cluster $1 = -1$
   - $\mu_2 =$ mean of points assigned to cluster $2 = 1$

Convergence:

- The assignments remain the same in subsequent iterations
- The means do not change after this point

Answer: The fixed point of K-means is $\mu_1 = -1$ and $\mu_2 = 1$