hw5-worked-out

October 29, 2024

1 Problem 1: Gaussian Multivariate

1.0.1 (a)

Are X_3 and X_4 correlated?

Solution Since the entry of $\sum_{34} = 0$, we can conclude that X_3 and X_4 have 0 covariance, and are thus **not correlated**.

$$\Sigma = \begin{bmatrix} 0.71 & -0.43 & 0.43 & 0 \\ -0.43 & 0.46 & -0.26 & 0 \\ 0.43 & -0.26 & 0.46 & 0 \\ 0 & 0 & 0 & 0.2 \end{bmatrix}$$

1.0.2 (b)

Solution Using the precision matrix Q to verify the conditional dependence of X_1 and X_2, we can see that $Q_{34} = 0$. This indicates that X_1 and X_2 are **conditionally independent** given X_1 and X_2.

Thus, we can conclude that $Cov(X_3, X_4 \mid X_1, X_2) = 0$.

$$Q = \begin{bmatrix} 5 & 3 & -3 & 0 \\ 3 & 5 & 0 & 0 \\ -3 & 0 & 5 & 0 \\ 0 & 0 & 0 & 5 \end{bmatrix}$$

1.0.3 (c)

Please find the Markov blanket of X_2. Recall that the Markov blanket of X_i is the set of variables (denoted by X_{M_i}), such that

$$X_i \perp X_{\neg\{i\} \cup M_i} | X_{M_i}$$

where

$$\neg\{i\} \cup M_i$$

denotes all the variables outside of $\{i\} \cup M_i$.

Solution Using the precision matrix, Q, we can find the Markov blanket by finding the minimal set of variables needed to make X_2 conditionally indpendent of the other variables.

Given

$$Q = \begin{bmatrix} 5 & 3 & -3 & 0 \\ 3 & 5 & 0 & 0 \\ -3 & 0 & 5 & 0 \\ 0 & 0 & 0 & 5 \end{bmatrix}$$

Since we are interested in X_2, we will evaluate the second row of Q, $[3, 5, 0, 0]$.

- $Q_{21} = 3$, so X_2 is conditionally dependent on X_1.
- Q_{23} and Q_{24} both equal 0, so X_2 is conditionally independent of X_3 and X_4.

Thus, **the Markov blanket for X_2 is:**

$$X_{M_2} = \{X_1\}$$

1.0.4 (d)

Assume that $Y = [Y_1, Y_2]^\top$ is defined by

$$Y_1 = X_1 + X_4$$
$$Y_2 = X_2 - X_4$$

Please calculate the covariance matrix of Y

Solution Let us fix some matrix A, which is the transformation matrix that maps X to Y:

$$Y = AX$$

Deriving A:

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} X_1 + X_4 \\ X_2 - X_4 \end{bmatrix}$$

$$Y = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix}$$

$$A = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 \end{bmatrix}$$

Now that we have A, the transformation matrix, we can use the relationship between Σ, X, and Y to find $Cov(Y)$. Since Y is a linear transformation of X by $Y = AX$, then:

$$Cov(Y) = A\Sigma A^\top$$

We will need to calculate $A\Sigma A^\top$:

$$A\Sigma A^\top = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} 0.71 & -0.43 & 0.43 & 0 \\ -0.43 & 0.46 & -0.26 & 0 \\ 0.43 & -0.26 & 0.46 & 0 \\ 0 & 0 & 0 & 0.2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 1 & -1 \end{bmatrix}$$

Which comes out to:

$$\text{Cov}(Y) = A\Sigma A^\top = \begin{bmatrix} 0.91 & -0.63 \\ -0.63 & 0.66 \end{bmatrix}$$

2 Problem 2: Expectation Maximization

2.0.1 (a)

Assume we run EM starting from an initialization of $\mu_1 = -2$ and $\mu_2 = 2$. Please decide the value of μ_1 and μ_2 at the next iteration of EM algorithm. (You may find it handy to know that $\frac{1}{1+exp(-4)} \approx 0.98$)

Solution Given: $x^1 = -1$, $x^2 = 1$, $\frac{1}{1+exp(-4)} \approx 0.98$), $\mu_1 = -2$, and $\mu_2 = 2$.

Step 0: Initialize the unknown parameters

Already done. $\mu_1 = -2$ and $\mu_2 = 2$ is given.

Step 1: Calculate the posterior distribution

Let's fix $\gamma_{ik}^t = Pr(z_i = k \mid x_i, \theta_t)$, the posterior distribution at iteration t.

Then, for the **first data point**:

$$\gamma_{11} = Pr(z_i = 1 \mid x_i, \theta_t)$$

$$= \frac{\pi_1 \mathcal{N}(x^1 \mid \mu_1, 1)}{\pi_1 \mathcal{N}(x^1 \mid \mu_1, 1) + \pi_2 \mathcal{N}(x^1 \mid \mu_2, 1)}$$

Substituting in the given values:

$$= \frac{\pi_1 \frac{1}{\sqrt{2\pi}} exp(-\frac{(-1+2)^2}{2})}{\pi_1 \frac{1}{\sqrt{2\pi}} exp(-\frac{(-1+2)^2}{2}) + \frac{1}{\sqrt{2\pi}} exp(-\frac{(-1-2)^2}{2})}$$

Simplifying:

$$= \frac{exp(-\frac{1}{2})}{exp(-\frac{1}{2}) + exp(-\frac{9}{2})}$$

Using the given $\frac{1}{1+exp(-4)} \approx 0.98$):

$$\gamma_{11} \approx 0.98$$

Thus,

$$\gamma_{12} \approx 0.02$$

And for the **second data point**:

$$\gamma_{21} = Pr(z_i = 1 \mid x_i, \theta_t)$$

$$= \frac{\pi_1 \mathcal{N}(x^2 \mid \mu_1, 1)}{\pi_1 \mathcal{N}(x^2 \mid \mu_1, 1) + \pi_2 \mathcal{N}(x^2 \mid \mu_2, 1)}$$

Substituting in the given values:

$$= \frac{\pi_1 \frac{1}{\sqrt{2\pi}} exp(-\frac{(1+2)^2}{2})}{\pi_1 \frac{1}{\sqrt{2\pi}} exp(-\frac{(1+2)^2}{2}) + \frac{1}{\sqrt{2\pi}} exp(-\frac{(1-2)^2}{2})}$$

Simplifying:

$$= \frac{exp(-\frac{9}{2})}{exp(-\frac{9}{2}) + exp(-\frac{1}{2})}$$

Solving:

$$\gamma_{21} \approx 0.02$$

Thus,

$$\gamma_{22} \approx 0.98$$

Step 2: Maximize the Likelihood Function

Using the update rule for μ derived in Lecture 4.2.0:

$$\mu_k^{t+1} = \frac{\sum_{i=1}^n \gamma_{ik}^t x_i}{\sum_{i=1}^n \gamma_{ik}^t}$$

We have:

$$\mu_1^{t+1} = \frac{(.98 \cdot -1) + (.02 \cdot 1)}{.98 + .02} = frac{-}.98 + .021 = -.96$$
$$\mu_2^{t+1} = \frac{(.02 \cdot -1) + (.98 \cdot 1)}{.02 + .98} = frac{-}.02 + .981 = .96$$

2.0.2 (b)

Do you think EM (when initialized with $\mu_1 = -2$ and $\mu_2 = 2$) will eventually converge to $\mu_1 = -1$ and $\mu_2 = 1$(i.e., coinciding with the two data points). Please justify your answer using either your theoretical understanding or the result of an empirical simulation.

Solution I think the EM algorithm **will** eventually converge to $\mu_1 = -1$ and $\mu_2 = 1$, coinciding with the two data points. Here is my reasoning:

There are exactly two points and exactly two distributions, with the distributions having different means (μ). This means that the distributions are **not** identical. Thus, naturally each point will coincide precisely with a distribution. Furthermore, the symmetry of the model, including the means and the data points, around 0 suggests that the model will converge nicely.

2.0.3 (c)

Please decide the fixed point of EM when we initialize it from $\mu_1 = \mu_2 = 2$

Solution With $\mu_1 = \mu_2$ both initialized to 2, and $\gamma_{ik}^t = Pr(z_i = k \mid x_i, \theta_t)$. The steps of EM will be as follows:

Step 1: Calculate the posterior distribution

Since the means are equal, we can assume that:

$$\gamma_{i1} = \gamma_{i2} = .5$$

This is because the likelihood for each point will be the same, since the distributions are initialized to be identical.

Step 2: Maximize the Likelihood Function

Using the update rule for μ derived in Lecture 4.2.0:

$$\mu_k^{t+1} = \frac{\sum_{i=1}^n \gamma_{ik}^t x_i}{\sum_{i=1}^n \gamma_{ik}^t}$$

We have:

$$\mu_1 = \frac{.5 \cdot (-1) + .5 \cdot 1}{.5 + .5} = 0$$

and:

$$\mu_2 = \frac{.5 \cdot (-1) + .5 \cdot 1}{.5 + .5} = 0$$

And since they are again identical, this cycle will repeat. Thus, **the fixed point for μ_1 and μ_2 is 0.**

2.0.4 (d)

Please decide the fixed point of K-means when we initialize it from $\mu_1 = -2$ and $\mu_2 = 2$.

Solution Given that $x_1 = -1$, $x_2 = 1$, $\mu_1 = -2$ and $\mu_2 = 2$ for a K-means problem I have worked out the steps of the K-means algorithm below.

As we can see, after 1 iteration the algorithm converges and reaches **the fixed point of $\mu_1 = -1$ and $\mu_2 = 1$.**