به نام خدا

پروژه درس آشنایی با ماشین لرنینگ

دکتر سجاد امینی

فاز دوم

مبین خطیب

۹۹۱۰۶۱۱۴

محمد علی هاشمی فر

۹۹۱۰۷۶۵۸

Theory Question ۱

$$\begin{bmatrix} y_۱ \\ y_۲ \end{bmatrix} \sim \mathcal{N}\Big(\begin{bmatrix} \mu_۱ \\ \mu_۲ \end{bmatrix}, \begin{bmatrix} \Sigma_{۱۱} & \Sigma_{۱۲} \\ \Sigma_{۲۱} & \Sigma_{۲۲} \end{bmatrix} \Big)$$

Then we have:

$$p(y_۱|y_۲) = \mathcal{N}(y_۱|\mu_{۱|۲}, \Sigma_{۱|۲})$$

$$\mu_{۱|۲} = \mu_۱ + \Sigma_{۱۲}\Sigma_{۲۲}^{-۱}(y_۲ - \mu_۲) \quad , \quad \Sigma_{۱|۲} = \Sigma_{۱۱} - \Sigma_{۱۲}\Sigma_{۲۲}^{-۱}\Sigma_{۲۱} \tag{۱}$$

Also we know that prior Gaussian is a prior conjugate for likelihood Gaussian and postrier distrubution is also gaussian so with above equations we can say:
($(y_۱ = y, y_۲ = z)$

$$p(y|z) = \mathcal{N}(y|Wz + b, \Sigma_{y|z}) \implies$$

$$\mu_{y|z} = Wz + b = \mu_y + \Sigma_{yz}\Sigma_z^{-۱}(z - \mu_z) \implies \begin{cases} W = \Sigma_{yz}\Sigma_z^{-۱} \implies \Sigma_{yz} = W\Sigma_z \\ b = \mu_y - \Sigma_{yz}\Sigma_z^{-۱}\mu_z = \mu_y - W\mu_z \end{cases}$$
$$\implies \mu_y = b + W\mu_z$$

$$\Sigma_{y|z} = \Sigma_y - \Sigma_{yz}\Sigma_z^{-۱}\Sigma_{zy} = \Sigma_y - W\Sigma_z\Sigma_z^{-۱}(W\Sigma_z)^T = \Sigma_y - W\Sigma_z W^T$$
$$\implies \Sigma_y = \Sigma_{y|z} + W\Sigma_z W^T$$

And so for joint distrubution we can say:

$$\begin{bmatrix} y \\ z \end{bmatrix} \sim \mathcal{N}\Big(\begin{bmatrix} \mu_y \\ \mu_z \end{bmatrix}, \begin{bmatrix} \Sigma_y & \Sigma_{yz} \\ \Sigma_{zy} & \Sigma_z \end{bmatrix} \Big)$$

The relations obtained for the parameters:

$$\Sigma_{yz} = W\Sigma_z \ , \ \mu_y = b + W\mu_z \ , \ \Sigma_y = \Sigma_{y|z} + W\Sigma_z W^T \tag{۲}$$
$$\implies \begin{bmatrix} y \\ z \end{bmatrix} \sim \mathcal{N}\Big(\begin{bmatrix} b + W\mu_z \\ \mu_z \end{bmatrix}, \begin{bmatrix} \Sigma_{y|z} + W\Sigma_z W^T & W\Sigma_z \\ \Sigma_z W^T & \boldsymbol{\Sigma_z} \end{bmatrix} \Big)$$

The log of the joint distribution is as follows (dropping irrelevant constants):

$$\log p(z; y) = -\frac{۱}{۲}(z - \mu z)^T\Sigma_z^{-۱}(z - \mu z) - \frac{۱}{۲}(y - Wz - b)^T\Sigma_y^{-۱}(y - Wz - b) \tag{۳}$$

$$Q = -\frac{۱}{۲}z^T\Sigma_z^{-۱}z - \frac{۱}{۲}y^T\Sigma_y^{-۱}y - \frac{۱}{۲}(Wz)^T\Sigma_y^{-۱}y - \frac{۱}{۲}y^T\Sigma_y^{-۱}(Wz) \tag{۴}$$

Simplifying the expression, we obtain:

$$Q = -\frac{۱}{۲}\begin{pmatrix} z \\ y \end{pmatrix}^T \begin{pmatrix} \Sigma_z^{-۱} + W^T\Sigma_y^{-۱}W & -W^T\Sigma_y^{-۱} \\ -\Sigma_z^{-۱}W & \Sigma_y^{-۱} \end{pmatrix} \begin{pmatrix} z \\ y \end{pmatrix} \tag{۵}$$

The precision matrix of the joint distribution is defined as:

$$\Sigma^{-1} = \begin{pmatrix} \Sigma_z^{-1} + W^T \Sigma_y^{-1} W & -W^T \Sigma_y^{-1} \\ -\Sigma_z^{-1} W & \Sigma_y^{-1} \end{pmatrix} = \Lambda \quad = \begin{pmatrix} \Lambda_{xx} & \Lambda_{xy} \\ \Lambda_{yx} & \Lambda_{yy} \end{pmatrix} \tag{۶}$$

Using the fact that $\mu y = W \mu z + b$, we have

$$p(z|y) = \mathcal{N}(\mu_{z|y}; \Sigma_{z|y}) \tag{۷}$$

where the precision matrix of $p(z|y)$ is given by

$$p(z|y) = \mathcal{N}(\mu_{z|y}; \Sigma_{z|y}) \tag{۸}$$

where

$$\Sigma_{z|y} = \Lambda_{xx}^{-1} = (\Sigma_z^{-1} + W^T \Sigma_y^{-1} W)^{-1} \tag{۹}$$

The mean of $p(z|y)$ is given by

$$\mu z|y = \Sigma_{z|y}(\Lambda_{xx}^{-1} \mu z - \Lambda_{xy}(y - \mu_y)) \tag{۱۰}$$

$$= \Sigma_{z|y} \left(\Sigma_z^{-1} \mu_z + W^T \Sigma_y^{-1} W \mu_z + W^T \Sigma_y^{-1} (y - \mu_y) \right) \tag{۱۱}$$

$$= \Sigma_{z|y} \left(\Sigma_z^{-1} \mu z + W^T \Sigma_y^{-1} (W \mu_z + y - \mu_y) \right) \tag{۱۲}$$

$$= \Sigma_{z|y} \left(\Sigma_z^{-1} \mu z + W^T \Sigma_y^{-1} (y - b) \right) \tag{۱۳}$$

Theory Question ۲

To prove that the posterior distribution p(Z|Y), given a GMM prior distribution p(Z) and a normal conditional distribution p(Y|Z), is also a GMM, we can use the properties of Bayesian inference and the fact that the Gaussian distribution is a conjugate prior for itself. Let's denote the prior GMM as p(Z) and the conditional distribution as$p(Y|Z)$. The joint distribution of the latent variable Z and the observed variable Y is given by: $p(Z, Y) = p(Y|Z) * p(Z)$ According to Bayes' theorem, the posterior distribution p(Z|Y) is obtained by normalizing the joint distribution: $p(Z|Y) = p(Y|Z) * p(Z)/p(Y)$ To compute$p(Z|Y)$, we need to calculate the numerator and denominator separately.

1. Numerator: The numerator of Bayes' theorem, p(Y|Z) * p(Z), represents the joint distribution of the observed data Y and the latent variable Z. Since $p(Y|Z)$ is a normal distribution, and p(Z) is a GMM, their product will result in a mixture of normals, i.e., a GMM.

2. Denominator: The denominator, p(Y), is the marginal distribution of the observed data Y. It is calculated by integrating the joint distribution over all possible values of Z: $p(Y) = \int p(Y|Z) * p(Z)dZ$ Integrating a GMM can be analytically intractable, but it can be approximated using techniques like Monte Carlo methods or variational inference. The denominator will depend on the observed data Y but may not have a simple closed-form expression. Given that the numerator is a GMM and the denominator may be approximated, the posterior distribution$p(Z|Y)$will be proportional to the numerator divided by the denominator. This means that $p(Z|Y)$ will also be a GMM, where the mixture weights, means, and covariance matrices of the components will depend on the specific form of $p(Y|Z)$, p(Z), and the observed data Y. While the exact parameters of the posterior GMM might not have a closed-form solution, techniques like the EM algorithm or variational inference can be used to estimate these parameters iteratively based on the observed data and the prior distribution. These methods can provide a practical approximation of the posterior GMM.

In summary, the posterior distribution $p(Z|Y)$ is a GMM due to the properties of Bayesian inference and the conjugacy relationship between the Gaussian distribution and itself. However, the exact form and parameters of the posterior GMM may require approximation methods depending on the specific distributions involved.

This is a mixture of normal distributions, with the same number of components as the prior distribution. The weights of the normal distributions are updated to reflect the Actually when prior and likelohood be normal then postrier will be new data point. normal too.Beause they are under conjugate prior .When the input is GMM every part of it is normal which we can achieve to its parameters with answers in quesdtion theory 1 and the method is every time we get prior from each part and gain parameters of postrier with previous problem.So outputs will be some normal distrubution wuth these things. Therefore, the posterior distribution is normal and a GMM.

$$p(Z|y) = \frac{p(y|Z)p(Z)}{p(y)} \tag{۱۴}$$

Theory Question ۳

1. Flexibility: GMM allows for capturing complex multimodal distributions. It represents the data as a mixture of multiple Gaussian components, each representing a distinct mode or cluster. This flexibility makes GMM well-suited for modeling data with underlying structure that exhibits multiple modes or clusters.

2. Uncertainty modeling: GMM provides a natural way to model uncertainty in the latent variable Z. Each Gaussian component in the mixture corresponds to a potential mode or cluster, and the mixture weights represent the probabilities of each component. This allows for capturing the uncertainty associated with the assignment of data points to different modes or clusters.

3. Generative modeling: GMM can be used as a generative model, meaning it can generate new samples that resemble the original data distribution. This property is useful in tasks such as data synthesis or anomaly detection, where new samples need to be generated based on the learned distribution.

4. Easy parameter estimation: GMM parameters (mixture weights, means, and covariance matrices) can be estimated using the expectation-maximization (EM) algorithm or other optimization techniques. These estimation methods are relatively straightforward and well-studied, making GMMs accessible for many applications.

5. Widely used and studied: GMMs have been extensively studied and used in various domains, including computer vision, natural language processing, and pattern recognition. Consequently, there is a wealth of research, tools, and libraries available for working with GMMs, making them a popular choice in practice.