MACHINE LEARNING
PROJECT PHASE 2

DR SAJJAD AMINI
ELECTRICAL ENGINEERING
SUT
July 4, 2023
Shagheyegh Ataei 98102002
Naghmeh Adibi 99101076

Theory question 1 .

Given a normal prior distribution $p_Z(z) = N(\mu_z, \Sigma_z)$ and a normal likelihood distribution $p_{Y|Z}(y|z) = N(W_z + b, \Sigma_{Y|Z})$, we can find the joint distribution $p_{Z,Y}(z, y)$ as follows:

$$
\begin{aligned}
p_{Z,Y}(z,y) &= p_Z(z) p_{Y|Z}(y|z) \\
&= \frac{1}{\sqrt{(2\pi)^{d_z} |\Sigma_z|}} \exp\left(-\frac{1}{2}(z-\mu_z)^T \Sigma_z^{-1} (z-\mu_z)\right) \\
&\quad \cdot \frac{1}{\sqrt{(2\pi)^{d_y} |\Sigma_{Y|Z}|}} \exp\left(-\frac{1}{2}(y - W_z - b)^T \Sigma_{Y|Z}^{-1} (y - W_z - b)\right) \\
&= \frac{1}{\sqrt{(2\pi)^{d_z+d_y} |\Sigma_z| |\Sigma_{Y|Z}|}} \\
&\quad \cdot \exp\left(-\frac{1}{2} \begin{bmatrix} z-\mu_z \\ y - W_z - b \end{bmatrix}^T \begin{bmatrix} \Sigma_z^{-1} & -\Sigma_z^{-1} W \\ -W^T \Sigma_z^{-1} & \Sigma_{Y|Z}^{-1} + W^T \Sigma_z^{-1} W \end{bmatrix} \begin{bmatrix} z-\mu_z \\ y - W_z - b \end{bmatrix}\right) \\
&= N\left(\begin{bmatrix} \mu_z \\ W\mu_z + b \end{bmatrix}, \begin{bmatrix} \Sigma_z & \Sigma_z W^T \\ W\Sigma_z & \Sigma_{Y|Z} + W \Sigma_z W^T \end{bmatrix}\right)
\end{aligned}
$$

Therefore, the joint distribution $p_{Z,Y}(z, y)$ is a normal distribution with mean $\begin{bmatrix} \mu_z \\ W\mu_z + b \end{bmatrix}$ and covariance matrix $\begin{bmatrix} \Sigma_z & \Sigma_z W^T \\ W\Sigma_z & \Sigma_{Y|Z} + W \Sigma_z W^T \end{bmatrix}$.

Given the joint distribution $p_{Z,Y}(z, y)$ derived in part 1, we can find the posterior distribution $p_{Z|Y}(z|y)$ as follows:
we need to show that the posterior distribution is also normal and find its parameters.

The posterior distribution is given by:

$p_{Z|Y}(z|y) = N(\mu_{Z|Y}, \Sigma_{Z|Y})$

To find the parameters of the posterior distribution, we can use Bayes' theorem:

$p_{Z|Y}(z|y) = p_{Y,Z}(y,z)/p_Y(y)$

where $p_Y(y)$ is the marginal distribution of y, which can be obtained by integrating out z from the joint distribution:

$p_Y(y) = \int p_{Y,Z}(y,z)dz$

Now let's calculate the parameters of the posterior distribution.

First, let's calculate $p_Y(y)$:

$p_Y(y) = \int p_{Y,Z}(y,z)dz$
$= \int N([\mu_Z; W\mu_Z + b], [\Sigma_Z, \Sigma_Z W^T; W\Sigma_Z\Sigma_{Y|Z} + W\Sigma_Z W^T])dz$

Since we have a multivariate normal distribution, we can integrate it using properties of multivariate normal distributions. The result will be a normalizing constant that does not depend on z or y.

$\int p_{Y,Z}(y,z)dz = \int N([\mu_Z; W\mu_Z + b], [\Sigma_Z, \Sigma_Z W^T; W\Sigma_Z\Sigma_{Y|Z} + W\Sigma_Z W^T])dz$
$= Normalizingconstant(let'scallitC) * \int N([\mu_Z; W\mu_Z + b], [\Sigma_Z, \Sigma_Z W^T; W\Sigma_Z\Sigma_{Y|Z} + W\Sigma_Z W^T])dz$
$= C * N([\mu_Y; b], [\Sigma_Y, \Sigma_Y W^T; W\Sigma_Y\Sigma_{Y|Z} + W\Sigma_Z W^T])$ The result of this integration is $p_Y(y)$, which is a normal distribution with parameters μ_Y and Σ_Y.

Now let's calculate the numerator of Bayes' theorem, which is $p_{Y,Z}(y,z)$:

$p_{Y,Z}(y,z) = N([\mu_Z; W\mu_Z + b], [\Sigma_Z, \Sigma_Z W^T; W\Sigma_Z\Sigma_{Y|Z} + W\Sigma_Z W^T])$

To expand the numerator of Bayes' theorem, we can write it as:

$$p_{Y,Z}(y,z) = (2\pi)^{-k/2}|\Sigma_Z|^{-1/2}exp[-0.5*(z-\mu_Z)^T\Sigma_Z^{-1}(z-\mu_Z)]*(2\pi)^{-m/2} |\Sigma_{Y|Z}+W\Sigma_Z W^T|^{-1/2}exp[-0.5*(y-Wz-b)^T(\Sigma_{Y|Z}+W\Sigma_Z W^T)^{-1}(y-Wz-b)]$$

Expanding the exponentials and simplifying, we get:

$$p_{Y,Z}(y,z) = (2\pi)^{-(k+m)/2}|\Sigma_Z|^{-1/2}|\Sigma_{Y|Z}+W\Sigma_Z W^T|^{-1/2}exp[-0.5*(z-\mu_Z)^T\Sigma_Z^{-1}(z-\mu_Z)] exp[-0.5*(y-Wz-b)^T(\Sigma_{Y|Z}+W\Sigma_Z W^T)^{-1}(y-Wz-b)]$$

Finally, we can substitute these results into Bayes' theorem to obtain the parameters of the posterior distribution:

$$\mu_{Z|Y} = \Sigma_{Z|Y}[W^T\Sigma_Y^{-1}(yb)+\Sigma_Z^{-1}\mu_Z]$$
$$\Sigma_{Z|Y}^{-1} = \Sigma_Z^{-1}+W^T\Sigma_Y^{-1}W$$

Therefore, the posterior distribution is also normal with parameters $\mu_{Z|Y}$ and $\Sigma_{Z|Y}$.

Theory question 2 .

If the prior distribution is a Gaussian mixture model (GMM), then the posterior distribution may not be normal or a GMM. The exact form of the posterior distribution will depend on the specific parameters of the prior and likelihood distributions.

However, if we assume that the posterior distribution is also a GMM, then we can use Bayesian inference to update the parameters of the prior distribution. The updated parameters will depend on both the prior and likelihood distributions.

The parameters of the GMM posterior distribution can be obtained using standard techniques for Bayesian inference, such as Markov chain Monte Carlo (MCMC) methods or variational inference.

Theory question 3 .

GMM is a preferred prior distribution for Z because it allows for more flexibility in modeling complex data distributions. GMM is a preferred prior distribution for Z because it can model complex distributions that may not be well approximated by a single Gaussian distribution. GMM allows for multiple modes and can capture non-Gaussian features of the data. Moreover, GMM is a flexible and computationally efficient method for modeling high-dimensional data. By using GMM as a prior distribution, we can incorporate prior knowledge about the structure of the data into our inference process and improve the accuracy of our estimates.

Simulation questions .

codes are completely explained on colab :)

https://colab.research.google.com/drive/1EsOS-cGtUDlFz_UqS1LIQmsT32qBxUuZ

https://colab.research.google.com/drive/124TmmgUqiUPvKuWszzKEOL9ZTqmLkmd3