In The Name OF God

Sharif University

Department of Electrical Engineering

Introduction to Machine Learning

Project Phase 2

Instructor:

Dr. Sajad Amini

Authors:

Sepehr Kazemi Ranjbar

99106599

Arman Lotfalikhani

99109166

2 Implementation

Step 1: Data preparation

Simulation Question 1.
As is requested in the project description in Telegram, we have included our report for this part in the jupyter notebook file

Step 2: Denoising algorithm

Simulation Question 2.
Same as simulation question 1. We have included our report for this part in the jupyter notebook file

Theory Question 1. if we have normal prior distribution $p_Z(z) = \mathcal{N}(\mu_z, \Sigma_z)$ and normal likelihood distribution $p_{Y|Z}(y|z) = \mathcal{N}(Wz + b, \Sigma_{Y|Z})$, show that joint distribution will be Normal and find its parameters $(\mu_{Y,Z}, \Sigma_{Y,Z})$.

$$p_{Z,Y}(z, y) = \mathcal{N}(\mu_{Z,Y}, \Sigma_{Z,Y})$$
$$\mu_{Z,Y} = \begin{bmatrix} \mu_Z \\ W\mu_Z + b \end{bmatrix}$$
$$\Sigma_{Z,Y} = \begin{bmatrix} \Sigma_Z & \Sigma_Z W^T \\ W\Sigma_Z & \Sigma_{Y|Z} + W\Sigma_Z W^T \end{bmatrix}$$

Show that the posterior is as follows:

$$p_{Z,Y}(z, y) = \mathcal{N}(\mu_{Z|Y}, \Sigma_{Z|Y})$$
$$\mu_{Z|Y} = \Sigma_{Z|Y}(W^T\Sigma_{Y|Z}^{-1}(y - b) + \Sigma_Z^{-1}\mu_Z)$$
$$\Sigma_{Z|Y}^{-1} = \Sigma_Z^{-1} + W^T\Sigma_{Y|Z}^{-1}W$$

For the first part, assuming y has dimension D_y and z has dimension D_x, we have:

$$\begin{aligned} P_{Z,Y}(z, y) &= P_{Y|Z}(y|z)P_Z(z) \\ &= \frac{1}{\sqrt{(2\pi)^{D_y} \det \Sigma_{Y|Z}}} \exp\Big(-\frac{1}{2}(y - Wz - b)^T\Sigma_{Y|Z}^{-1}(y - Wz - b)\Big) \\ &\times \frac{1}{\sqrt{(2\pi)^{D_z} \det \Sigma_Z}} \exp\Big(-\frac{1}{2}(z - \mu_Z)^T\Sigma_Z^{-1}(z - \mu_Z)\Big) \end{aligned}$$

For convenience and simplification of the expressions, we define $u := z - \mu_Z$ and $v := y - W\mu_Z - b$. As a result, we have $y - (Wz + b) = y - (W(z - \mu_Z + \mu_Z) + b) = v - Wu$.

$$P_{Z,Y}(z,y) = \frac{\exp\Big(-\frac{1}{2}[u^T\Sigma_Z^{-1}u + (v-Wu)^T\Sigma_{Y|Z}^{-1}(v-Wu)]\Big)}{\sqrt{(2\pi)^{D_y+D_z}\det\Sigma_{Y|Z}\det\Sigma_Z}} \tag{1}$$

Next, we have to simplify the exponent:

$$\begin{aligned} u^T\Sigma_Z^{-1}u + (v-Wu)^T\Sigma_{Y|Z}^{-1}(v-Wu) &= u^T(\Sigma_Z^{-1} + W^T\Sigma_{Y|Z}^{-1}W)u + v^T\Sigma_{Y|Z}^{-1}v \\ &\quad - v^T\Sigma_{Y|Z}^{-1}Wu - u^TW^T\Sigma_{Y|Z}^{-1}v \\ &= \begin{bmatrix} u^T & v^T \end{bmatrix} \begin{bmatrix} \Sigma_Z^{-1} + W^T\Sigma_{Y|Z}^{-1}W & -W^T\Sigma_{Y|Z}^{-1} \\ -\Sigma_{Y|Z}^{-1}W & \Sigma_{Y|Z}^{-1} \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \\ &= \begin{bmatrix} u^T & v^T \end{bmatrix} A \begin{bmatrix} u \\ v \end{bmatrix} \end{aligned} \tag{2}$$

We now use the inversion theorem for a block matrix, as stated in [**Murphy**]:

$$\begin{aligned} M = \begin{bmatrix} E & F \\ G & H \end{bmatrix} \quad S = E - FH^{-1}G &\Rightarrow 1: \det M = \det S \det H \\ 2: \ M^{-1} &= \begin{bmatrix} S^{-1} & -S^{-1}FH^{-1} \\ -H^{-1}GS^{-1} & H^{-1} + H^{-1}GS^{-1}FH^{-1} \end{bmatrix} \end{aligned} \tag{3}$$

For the above matrix A we have:

$$\begin{aligned} &S = \Sigma_Z^{-1} + W^T\Sigma_{Y|Z}^{-1}W - W^T\Sigma_{Y|Z}^{-1}\Sigma_{Y|Z}\Sigma_{Y|Z}^{-1}W = \Sigma_Z^{-1} \\ \Rightarrow &\boxed{A^{-1} = \begin{bmatrix} \Sigma_Z & \Sigma_Z W^T \\ W\Sigma_Z & \Sigma_{Y|Z} + W\Sigma_Z W^T \end{bmatrix}} \quad \det A = \det\Sigma_Z^{-1}\det\Sigma_{Y|Z}^{-1} \\ &\Rightarrow \det A^{-1} = \det\Sigma_Z \det\Sigma_{Y|Z} \end{aligned} \tag{4}$$

Using the above results, we can write the joint distribution in a simple form:

$$\begin{aligned} P_{Z,Y}(z,y) &= \frac{1}{\sqrt{(2\pi)^{D_y+D_z}\det A^{-1}}}\exp\Big(-\frac{1}{2}\begin{bmatrix} u^T & v^T \end{bmatrix} A \begin{bmatrix} u \\ v \end{bmatrix}\Big) \\ &= \boxed{\frac{1}{\sqrt{(2\pi)^{D_y+D_z}\det A^{-1}}}\exp\Big(-\frac{1}{2}\begin{bmatrix} z-\mu_Z \\ y - W\mu_Z - b \end{bmatrix}^T A \begin{bmatrix} z-\mu_Z \\ y - W\mu_Z - b \end{bmatrix}\Big)} \end{aligned} \tag{5}$$

We conclude that the final form of the distribution is indeed a normal distribution with the mean and variance as requested to prove. Note that A^{-1} is equal to the $\Sigma_{Z,Y}$ in the question.

For the second part, using the definitions of u and v , we have:

$$P_{Z|Y}(z|y) = \frac{P_{Y|Z}(y|z)P_Z(z)}{P_Y(y)} = \frac{P_{Y|Z}(y|z)P_Z(z)}{\int P_{Y|Z}(y|z)P_Z(z)dz} \tag{6}$$

Now we use equations (1) (2) .Defining matrices $B = \Sigma_Z^{-1} + W^T\Sigma_{Y|Z}^{-1}W$ and $C = \Sigma_{Y|Z}^{-1}W$ for further notational simplicity, we have :

$$\begin{aligned} P_{Z|Y}(z|y) &= \frac{\exp\Big(-\frac{1}{2}[u^T\Sigma_Z^{-1}u + (v-Wu)^T\Sigma_{Y|Z}^{-1}(v-Wu)]\Big)}{\int \exp\Big(-\frac{1}{2}[u^T\Sigma_Z^{-1}u + (v-Wu)^T\Sigma_{Y|Z}^{-1}(v-Wu)]\Big)du} \\ &= \frac{\exp\Big(-\frac{1}{2}[u^TBu + v^T\Sigma_{Y|Z}^{-1}v - v^TCu - u^TC^Tv]\Big)}{\int \exp\Big(-\frac{1}{2}[u^TBu + v^T\Sigma_{Y|Z}^{-1}v - v^TCu - u^TC^Tv]\Big)du} \\ &= \frac{\exp\Big(-\frac{1}{2}[u^TBu - v^TCu - u^TC^Tv]\Big)}{\int \exp\Big(-\frac{1}{2}[u^TBu - v^TCu - u^TC^Tv]\Big)du} \end{aligned} \tag{7}$$

Next, we factorize the exponents as follows:

$$u^TBu - v^TCu - u^TC^Tv = (u - B^{-1}C^Tv)^TB(u - B^{-1}C^Tv) - v^TCB^{-1}C^Tv \tag{8}$$

$$P_{Z|Y}(z|y) = \frac{\exp\Big(-\frac{1}{2}[(u - B^{-1}C^Tv)^TB(u - B^{-1}C^Tv)]\Big)}{\int \exp\Big(-\frac{1}{2}[(u - B^{-1}C^Tv)^TB(u - B^{-1}C^Tv)]\Big)du} \tag{9}$$

The final step is to use the basic property of a normal distribution, that is, its integral over the whole space equals to one:

$$\begin{aligned} &\int \frac{1}{\sqrt{(2\pi)^{D_x}\det\Sigma_X}}\exp\Big(-\frac{1}{2}(x-\mu_X)^T\Sigma_X^{-1}(x-\mu_X)\Big)dx = 1 \\ \Rightarrow &\int \exp\Big(-\frac{1}{2}(x-\mu_X)^T\Sigma_X^{-1}(x-\mu_X)\Big) = \sqrt{(2\pi)^{D_x}\det\Sigma_X} \end{aligned} \tag{10}$$

We can now write:

$$\boxed{P_{Z|Y}(z|y) = \frac{\exp\Big(-\frac{1}{2}[(u - B^{-1}C^Tv)^TB(u - B^{-1}C^Tv)]\Big)}{\sqrt{(2\pi)^{D_z}\det B^{-1}}}} \tag{11}$$

It is confirmed that the posterior is indeed a normal distribution with $\boxed{\Sigma_{Z|Y}^{-1} = B = \Sigma_Z^{-1} + W^T \Sigma_{Y|Z}^{-1} W}$, recalling that u and v are shifted versions of z and y. Finally, we compute the mean of the posterior:

$$\begin{aligned}
u - B^{-1}Cv &= z - \mu_Z - \Sigma_{Z|Y} W^T \Sigma_{Y|Z}^{-1}(y - W\mu_Z - b) \\
&= z - \Sigma_{Z|Y} W^T \Sigma_{Y|Z}^{-1}(y - b) + [-I + (\Sigma_Z^{-1} + W^T \Sigma_{Y|Z}^{-1} W)^{-1} W^T \Sigma_{Y|Z}^{-1} W]\mu_Z \\
&= z - \Sigma_{Z|Y} W^T \Sigma_{Y|Z}^{-1}(y - b) - (\Sigma_Z^{-1} + W^T \Sigma_{Y|Z}^{-1} W)^{-1} \Sigma_Z^{-1} \mu_z \\
&= z - \Sigma_{Z|Y}[W^T \Sigma_{Y|Z}^{-1}(y - b) + \Sigma_Z^{-1}\mu_z] = z - \mu_{Z|Y} \\
&\Rightarrow \boxed{\mu_{Z|Y} = \Sigma_{Z|Y}[W^T \Sigma_{Y|Z}^{-1}(y - b) + \Sigma_Z^{-1}\mu_z]}
\end{aligned} \tag{12}$$

Theory Question 2. Assume that prior distribution is GMM and $p_{Y|Z}$ is a normal distribtion. Is posterior distribution normal? Is it also a GMM? If it is GMM, what are parameters of this GMM if we know parameters of the aforementioned distributions?

We have to modify equation (1) and replace the prior distribution with a GMM. In this case, we have:

$$\begin{aligned}
P_{Z|Y}(z|y) = \frac{P_{Y|Z}(y|z)P_Z(z)}{P_Y(y)} &= \frac{P_{Y|Z}(y|z)P_Z(z)}{\int P_{Y|Z}(y|z)P_Z(z)dz} \\
&= \frac{\sum_{k=1}^K \pi_k \mathcal{N}(\mu_{Y|Z}, \Sigma_{Y|Z})\mathcal{N}(\mu_k, \Sigma_k)}{\int \sum_{k=1}^K \pi_k \mathcal{N}(\mu_{Y|Z}, \Sigma_{Y|Z})\mathcal{N}(\mu_k, \Sigma_k)dz}
\end{aligned} \tag{13}$$

In the next step, we use u, v, C, B from the previous section, but we have to label them by the index k. Using a modified version of (6) and (7), we can write:

$$\begin{aligned}
P_{Z|Y}(z|y) &= \frac{\sum_{k=1}^K \frac{\pi_k}{\sqrt{(2\pi)^{D_y+D_z}\det\Sigma_{Y|Z}\det\Sigma_k}} \exp\Big(-\frac{1}{2}[u_k^T B_k u_k + v_k^T \Sigma_{Y|Z}^{-1} v_k - v_k^T C u_k - u_k^T C^T v_k]\Big)}{\int \sum_{k=1}^K \frac{dz\pi_k}{\sqrt{(2\pi)^{D_y+D_z}\det\Sigma_{Y|Z}\det\Sigma_k}} \exp\Big(-\frac{1}{2}[u_k^T B_k u_k + v_k^T \Sigma_{Y|Z}^{-1} v_k - v_k^T C u_k - u_k^T C^T v_k]\Big)} \\
&= \frac{\sum_{k=1}^K \frac{\pi_k}{\sqrt{(2\pi)^{D_z}\det\Sigma_k}} \exp\Big(-\frac{1}{2}[u_k^T B_k u_k + v_k^T \Sigma_{Y|Z}^{-1} v_k - v_k^T C u_k - u_k^T C^T v_k]\Big)}{\sum_{k=1}^K \int \frac{\pi_k}{\sqrt{(2\pi)^{D_z}\det\Sigma_k}} \exp\Big(-\frac{1}{2}[u_k^T B_k u_k + v_k^T \Sigma_{Y|Z}^{-1} v_k - v_k^T C u_k - u_k^T C^T v_k]\Big)du_k}
\end{aligned} \tag{14}$$

Modifying the previous factorization is as follows:

$$\begin{aligned}
u_k^T B_k u_k + v_k^T \Sigma_{Y|Z}^{-1} v_k - v_k^T C u_k - u_k^T C^T v_k &= (u_k - B_k^{-1} C^T v_k)^T B_k (u_k - B_k^{-1} C^T v_k) \\
&\quad + v_k^T (\Sigma_{Y|Z}^{-1} - C B_k^{-1} C^T) v_k
\end{aligned} \tag{15}$$

Again, we use the normalization property of each gaussian distribution to evaluate the integrals:

$$\boxed{P_{Z|Y}(z|y) = \frac{\sum_{k=1}^{K} \frac{\pi_k \sqrt{\det B_k^{-1}}}{\sqrt{\det \Sigma_k}} \exp\Big(-\frac{1}{2} v_k^T (\Sigma_{Y|Z}^{-1} - C B_k^{-1} C^T) v_k \Big) \mathcal{N}(\mu_{Z|Yk}, B_K^{-1})}{\sum_{k=1}^{K} \frac{\pi_k \sqrt{\det B_k^{-1}}}{\sqrt{\det \Sigma_k}} \exp\Big(-\frac{1}{2} v_k^T (\Sigma_{Y|Z}^{-1} - C B_k^{-1} C^T) v_k \Big)}} \tag{16}$$

Where the parameters are:

$$\boxed{\begin{aligned} B_k = \Sigma_k^{-1} + W^T \Sigma_{Y|Z}^{-1} W \\ C = \Sigma_{Y|Z}^{-1} W \\ v_k = y - b - W\mu_k \end{aligned}} \tag{17}$$

We conclude that the posterior is again a GMM. The mean and variance of each component is the same as in Question 1, with Σ_k replacing Σ_Z and the new coefficients are:

$$\boxed{\begin{aligned} \pi_k' &= \frac{\frac{\pi_k \sqrt{\det B_k^{-1}}}{\sqrt{\det \Sigma_k}} \exp\Big(-\frac{1}{2} v_k^T (\Sigma_{Y|Z}^{-1} - C B_k^{-1} C^T) v_k \Big)}{\sum_{k=1}^{K} \frac{\pi_k \sqrt{\det B_k^{-1}}}{\sqrt{\det \Sigma_k}} \exp\Big(-\frac{1}{2} v_k^T (\Sigma_{Y|Z}^{-1} - C B_k^{-1} C^T) v_k \Big)} \\ &= \frac{\frac{\pi_k}{\sqrt{\det\big(I + \Sigma_k W^T \Sigma_{Y|Z}^{-1} W\big)}} \exp\Big(-\frac{1}{2} v_k^T (\Sigma_{Y|Z}^{-1} - C B_k^{-1} C^T) v_k \Big)}{\sum_{k=1}^{K} \frac{\pi_k}{\sqrt{\det\big(I + \Sigma_k W^T \Sigma_{Y|Z}^{-1} W\big)}} \exp\Big(-\frac{1}{2} v_k^T (\Sigma_{Y|Z}^{-1} - C B_k^{-1} C^T) v_k \Big)} \end{aligned}} \tag{18}$$

Theory Question 3. Why GMM is prefered prior distribution for Z?
Because a GMM, if we use the right number of components (which in the project, will be determined with cross validation) we can approximate any desired distribution. Also, the posterior will be a GMM as well, with closed-form coefficients, means and variances. This enables us to get a MAP estimate easily.

Simulation Question 3.
Report in the jupyter notebook file

Simulation Question 4.
Report in the jupyter notebook file

References

[1] Kevin P.Murphy (2012) *Machine Learning-A Probabilistic Perspective*, MIT Press.