# Intruduction to Machine Learning
## Dr S.Amini

department: Electrical Engineering

Amirreza Velae         400102222
Mohammadreza Hajibabaei   99101416

Project Phase 2

July 1, 2023

# Intruduction to Machine Learning

Project Phase 2

Amirreza Velae        400102222
Mohammadreza Hajibabaei     99101416

## ▬▬ Theory questions

### Question 1

We want to find joint distribution of Z and Y:

$$p_Z(z) = N(\mu_z, \Sigma_z), p_{Y|Z} = N(Wz + b, \Sigma_{y|z})$$

First because of linearity of expectation we have:

$$\mu_{y|z} = E[y|z] = E[Wz + b] = WE[z] + b = W\mu_z + b$$

We are calculating this:

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

$$ln(p_{Z,Y}(z, y)) \propto -\frac{1}{2}(z - \mu_z)^T\Sigma_z^{-1}(z - \mu_z) - \frac{1}{2}(y - Wz - b)^T\Sigma_{y|z}^{-1}(y - Wz - b)$$

We want to find covariance matrix and we do not use linear terms:

$$ln(p_{Z,Y}(z, y)) \propto -\frac{1}{2}(z)^T\Sigma_z^{-1}(z) - \frac{1}{2}(y)^T\Sigma_{y|z}^{-1}(y) - \frac{1}{2}(Wz)^T\Sigma_y^{-1}(Wz) + y^T\Sigma_y^{-1}Wz$$

The previous equation in matrix form:

$$ln(p_{Z,Y}(z, y)) \propto -\frac{1}{2}\begin{bmatrix} z \\ y \end{bmatrix}^T \begin{bmatrix} \Sigma_z^{-1} + W^T\Sigma_y^{-1}W & -W^T\Sigma_y^{-1} \\ -\Sigma_y^{-1}W & \Sigma_y^{-1} \end{bmatrix} \begin{bmatrix} z \\ y \end{bmatrix}$$

So the precision matrix is:

$$\Sigma^{-1} = \begin{bmatrix} \Sigma_z^{-1} + W^T\Sigma_y^{-1}W & -W^T\Sigma_y^{-1} \\ -\Sigma_y^{-1}W & \Sigma_y^{-1} \end{bmatrix} \triangleq \Lambda = \begin{bmatrix} \Lambda_{zz} & \Lambda_{zy} \\ \Lambda_{yz} & \Lambda_{yy} \end{bmatrix}$$

We want to find the inverse of the above matrix so we use complement Schur.Consider a block matrix M:

$$\begin{bmatrix} E & F \\ G & H \end{bmatrix}$$

Then the inverse of M is:

$$M^{-1} = \begin{bmatrix} (M/H)^{-1} & -(M/H)^{-1}FH^{-1} \\ -H^{-1}G(M/H)^{-1} & H^{-1} + H^{-1}G(M/H)^{-1}FH^{-1} \end{bmatrix}$$

And Schur complement of M w.r.t H is:

$$(M/H)^{-1} = E - FH^{-1}G$$

proof:First, we make M diagonal:

$$\overbrace{\begin{bmatrix} I & -FH^{-1} \\ 0 & I \end{bmatrix}}^{X} \overbrace{\begin{bmatrix} E & F \\ G & H \end{bmatrix}}^{M} \overbrace{\begin{bmatrix} I & 0 \\ -H^{-1}G & I \end{bmatrix}}^{Z} = \overbrace{\begin{bmatrix} E - FH^{-1}G & 0 \\ 0 & H \end{bmatrix}}^{W}$$

Inverse from both side:

$$Z^{-1}M^{-1}X^{-1} = W^{-1} \rightarrow M^{-1} = ZW^{-1}X$$

$$\begin{bmatrix} E & F \\ G & H \end{bmatrix}^{-1} = \begin{bmatrix} I & 0 \\ -H^{-1}G & I \end{bmatrix} \begin{bmatrix} (M/H)^{-1} & 0 \\ 0 & H^{-1} \end{bmatrix} \begin{bmatrix} I & -FH^{-1} \\ 0 & I \end{bmatrix} =$$

$$\begin{bmatrix} (M/H)^{-1} & -(M/H)^{-1}FH^{-1} \\ -H^{-1}G(M/H)^{-1} & H^{-1} + H^{-1}G(M/H)^{-1}FH^{-1} \end{bmatrix}$$

Now we calculate the inverse of precision matrix with the previous way:

$$E = \Sigma_z^{-1} + W^T\Sigma_{y|z}^{-1}W, F = -W^T\Sigma_{y|z}^{-1}, G = -\Sigma_{y|z}^{-1}W, H = \Sigma_{y|z}^{-1}$$

So:

$$\Sigma = \begin{bmatrix} \Sigma_z & \Sigma_z W^T \\ W\Sigma_z & \Sigma_{y|z} + W\Sigma_z W^T \end{bmatrix}$$

And the mean matrix is:

$$\mu = \begin{bmatrix} \mu_z \\ W\mu_z + b \end{bmatrix}$$

For calculating posterior we use MVN.Assume $y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$ and parameters of joint distribution is:

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \Lambda = \Sigma^{-1} = \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{bmatrix}$$

Then the posterior distribution is:

$$p(y_1|y_2) = \mathcal{N}(y_1|\mu_{1|2}, \Sigma_{1|2})$$

$$\mu_{1|2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(y_2 - \mu_2) = \mu_1 - \Lambda_{11}^{-1}\Lambda_{12}(y_2 - \mu_2) = \Sigma_{1|2}(\Lambda_{11}\mu_1 - \Lambda_{12}(y_2 - \mu_2))$$

$$\Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

Proof with complement Schur:

$$p(y_1, y_2) \propto exp(-\frac{1}{2}\begin{bmatrix} y_1 - \mu_1 \\ y_2 - \mu_2 \end{bmatrix}^T \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}^{-1} \begin{bmatrix} y_1 - \mu_1 \\ y_2 - \mu_2 \end{bmatrix})$$

$$= exp(-\frac{1}{2}\begin{bmatrix} y_1 - \mu_1 \\ y_2 - \mu_2 \end{bmatrix}^T \begin{bmatrix} I & 0 \\ -\Sigma_{22}^{-1}\Sigma_{21} & I \end{bmatrix} \begin{bmatrix} (\Sigma/\Sigma_{22})^{-1} & 0 \\ 0 & \Sigma_{22}^{-1} \end{bmatrix} \begin{bmatrix} I & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} y_1 - \mu_1 \\ y_2 - \mu_2 \end{bmatrix})$$

then we expand this equation:

$$exp(-\frac{1}{2}(y_1-\mu_1-\Sigma_{12}\Sigma_{22}^{-1}(y_2-\mu_2))^T(\Sigma/\Sigma_{22})^{-1}(y_1-\mu_1-\Sigma_{12}\Sigma_{22}^{-1}(y_2-\mu_2)))exp(-\frac{1}{2}(y_2-\mu_2)^T\Sigma_{22}^{-1}(y_2-\mu_2))$$

The above equation is in the form of the 2nd exponential of y1 and y2 multiplied by the 2nd exponential of y1 so:

$$p(y_1, y_2) = p(y_1|y_2)p(y_2) = \mathcal{N}(y_1|\mu_{1|2}, \Sigma_{1|2})\mathcal{N}(y_2|\mu_2, \Sigma_{22})$$

$$\mu_{1|2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(y_2 - \mu_2), \Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

Now we calculate parameters of posterior distribution in this way:

$$p(z|y) = \mathcal{N}(z|\mu_{z|y}, \Sigma_{z|y})$$

$$\Sigma_{z|y} = \Lambda_{zz}^{-1} \rightarrow \Sigma_{z|y} = \Sigma_z^{-1} + W^T\Sigma_y^{-1}W$$

$$\mu_{z|y} = \Sigma_{z|y}(\Lambda_{11}^{-1}\mu_z - \Lambda_{zy}(y - \mu_y)) = \Sigma_{z|y}(\Sigma_z^{-1}\mu_z + W^T\Sigma_y^{-1}W\mu_z + W^T\Sigma_y^{-1}(y - \mu_y))$$

$$= \Sigma_{z|y}(\Sigma_z^{-1}\mu_z + W^T\Sigma_y^{-1}(y - b))$$

## Question 2

$$p_Z(z) = \sum_{k=1}^{K} \pi_k N(z|\mu_{z,k}, \Sigma_{z,k}), p_{Y|Z} = N(y|Wz + b, \Sigma_{y|z})$$

$$p_{Z|Y}(z|y) = \frac{p_{Y|Z}(y|z)p_Z(z)}{p_Y(y)} = \frac{\sum_{k=1}^{K} \pi_k N(z|\mu_{z,k}, \Sigma_{z,k})N(y|Wz + b, \Sigma_{y|z})}{p_Y(y)}$$

According to LGS:

$$\frac{N(z|\mu_z, \Sigma_z)N(y|Wz + b, \Sigma_{y|z})}{p_Y(y)} = N(z|\mu_{z|y}, \Sigma_{z|y})$$

$$p_{Z|Y}(z|y) = \frac{\sum_{k=1}^{K} \pi_k N(z|\mu_{z,k}, \Sigma_{z,k})N(y|Wz + b, \Sigma_{y|z})}{p_Y(y)} = \sum_{k=1}^{K} \pi_k N(z|\mu_{z,k|y}, \Sigma_{z,k|y})$$

So the posterior distribution is also GMM and we can calculate the parameters according to LGS:

$$\Sigma_{z,k|y}^{-1} = \Sigma_{z,k}^{-1} + W^T \Sigma_y^{-1} W$$

$$\mu_{z,k|y} = \Sigma_{z,k|y}(W^T \Sigma_y^{-1}(y - b) + \Sigma_{z,k}^{-1}\mu_{z,k})$$

## Question 3

According to part2 if the prior distribution of Z is GMM and the distribution of likelihood is Gaussian,then the posterior distribution is GMM with parameters that calculated in previous part so for this problem GMM distribution is suggested for random variable Z.

# End of Project Phase 2