Aston University

Student Name: Muhammad Aizaz Afzal

Student Number: 200261713

Module: EE4108 – Machine Learning

Program: MSc SmartNet (EMJMD)

Assignment # 2

Ans 1)

We first calculate its mean:

$$\mathbb{E}[x] = \int_0^{+\infty} x \frac{1}{\Gamma(a)} b^a x^{a-1} \exp(-bx) dx = \frac{b^a}{\Gamma(a)} \int_0^{+\infty} x^a \exp(-bx) dx$$

$$= \frac{b^a}{\Gamma(a)} \int_0^{+\infty} \left(\frac{u}{b}\right)^a \exp(-u) \frac{1}{b} du$$

$$= \frac{1}{\Gamma(a) \cdot b} \int_0^{+\infty} u^a \exp(-u) du$$

$$\boldsymbol{= \frac{1}{\Gamma(a) \cdot b} \cdot \Gamma(a+1) = \frac{a}{b}}$$

Where we have taken advantage of the property $\Gamma(a+1) = a\Gamma(a)$, Then we calculate $\mathbb{E}[\lambda^2]$

$$\int_0^{+\infty} x^2 \frac{1}{\Gamma(a)} b^a x^{a-1} \exp(-bx)\, dx = \frac{b^a}{\Gamma(a)} \int_0^{+\infty} x^{a+1} \exp(-bx)\, dx$$

$$= \frac{b^a}{\Gamma(a)} \int_0^{+\infty} \left(\frac{u}{b}\right)^{a+1} \exp(-u) \frac{1}{b} du$$

$$= \frac{1}{\Gamma(a) \cdot b^2} \int_0^{+\infty} u^{a+1} \exp(-u)\, du$$

$$= \frac{1}{\Gamma(a) \cdot b^2} \cdot \Gamma(a+2) = \frac{a(a+1)}{b^2}$$

Therefore, according to $\text{var}\,[\lambda] = \mathbb{E}[\lambda^2] - \mathbb{E}[\lambda]^2$, we can obtain:

$$\text{var}\,[\lambda] = \mathbb{E}[\lambda^2] - \mathbb{E}[\lambda]^2$$

$$\boldsymbol{\text{var}\,[\lambda] = \frac{a(a+1)}{b^2} - \left(\frac{a}{b}\right)^2 = \frac{a}{b^2}}$$

Ans 2)

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} M & -MBD^{-1} \\ -D^{-1}CM & D^{-1} + D^{-1}CMBD^{-1} \end{pmatrix}$$

$M = (A - BD^{-1}C)^{-1}$

We know that $XX^{-1} = I$

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}\begin{pmatrix} M & -MBD^{-1} \\ -D^{-1}CM & D^{-1} + D^{-1}CMBD^{-1} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$\begin{pmatrix} AM - BD^{-1}CM & -AMBD^{-1} + BD^{-1} + BD^{-1}CMBD^{-1} \\ CM - DD^{-1}CM & -MBCD^{-1} + DD^{-1} + DD^{-1}CMBD^{-1} \end{pmatrix} = \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix}$$

1st Entry:

$AM - BD^{-1}CM = I$

$M(A - BD^{-1}C) = I$

$M = (A - BD^{-1}C)^{-1}$ → SATISFIED

2nd Entry:

$CM - DD^{-1}CM = 0$

$CM - ICM => CM - CM = 0$ → SATISFIED

3rd Entry:

$-AMBD^{-1} + BD^{-1} + BD^{-1}CMBD^{-1} = 0$

$BD^{-1}(-AM + I + CMBD^{-1}) = 0$

$I = (AM - CMBD^{-1})$

$M(A - BD^{-1}C) = I$

$M = (A - BD^{-1}C)^{-1}$ → SATISFIED

4th Entry:

$-MBCD^{-1} + DD^{-1} + DD^{-1}CMBD^{-1} = I$

$-MBCD^{-1} + I + MBCD^{-1} = I$

$I = I$ → SATISFIED

Ans 3)

$$\Lambda^{-1} = (\Lambda + A^{\top}LA - (-A^{\top}L)(L)^{-1}(-LA))^{-1}$$
$$= \left(\Lambda + A^{\top}AL - (A^{\top}ALL^{-1}L)\right)^{-1} \qquad M_{11}$$
$$= (\Lambda + A^{\top}AL - A^{\top}AL)^{-1}$$
$$\boldsymbol{\Lambda^{-1} = \Lambda^{-1}}$$

$$\Lambda^{-1}A^{\top} = -\left(\Lambda^{-1}(-LA^{\top})(L^{-1})\right)$$
$$\Lambda^{-1}A^{\top} = \Lambda^{-1}A^{\top}LL^{-1} \qquad M_{12}$$
$$\boldsymbol{\Lambda^{-1}A^{\top} = \Lambda^{-1}A^{\top}}$$

$$A\Lambda^{-1} = -L^{-1}(-LA)\Lambda^{-1} = A\Lambda^{-1}$$
$$A\Lambda^{-1} = L^{-1}LA\Lambda^{-1} \qquad M_{21}$$
$$\boldsymbol{A\Lambda^{-1} = A\Lambda^{-1}}$$

$$L^{-1} + A\Lambda^{-1}A^{\top} = L^{-1} + L^{-1}(-LA)\Lambda^{-1}(-A^{\top}L)L^{-1}$$
$$L^{-1} + A\Lambda^{-1}A^{\top} = L^{-1} - L^{-1}LA\Lambda^{-1}(-A^{\top}LL^{-1}) \qquad M_{22}$$
$$L^{-1} + A\Lambda^{-1}A^{\top} = L^{-1} + L^{-1}LA\Lambda^{-1}(A^{\top}I)$$
$$\boldsymbol{L^{-1} + A\Lambda^{-1}A^{\top} = L^{-1} + A\Lambda^{-1}A^{\top}}$$

Ans 4)

Given that:

$$\boldsymbol{p(x) = \mathcal{N}(x \mid \mu_x, \Sigma_x)}\ , \quad \boldsymbol{p(z) = \mathcal{N}(z \mid \mu_z, \Sigma_z)}$$

$$\boldsymbol{y = x + z}$$

For $p(y|x)$ we assume gaussian distribution same as above cases:

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

To find the expression for $p(y|x)$ we need $\mu_{y|x}, \Sigma_{y|x}$

We know from PRML book,

$$\boldsymbol{E[x] = x}$$

$$\boldsymbol{cov[x] = E[x^2] - (E[x])^2 = 0}$$

Therefore, we can write:

$\mu_{y|x} = E[x] + E[z] = x + \mu_z$

$\Sigma_{y|x} = cov[x] + cov[z] = \Sigma_z$

$p(y|x) = \mathcal{N}(y \mid x + \mu_z, \Sigma_z)$

Comparinig with the formula and some formula from PRML book with the expressions we have obtained:

$p(\mathrm{x}) = \mathcal{N}(\mathrm{x} \mid \mu, \Lambda^{-1})$

$p(\mathrm{y} \mid \mathrm{x}) = \mathcal{N}(\mathrm{y} \mid \mathrm{Ax} + \mathrm{b}, \mathrm{L}^{-1})$

$p(\mathrm{y}) = \mathcal{N}(\mathrm{y} \mid \mathrm{A}\mu + \mathrm{b}, \mathrm{L}^{-1} + \mathrm{A}\Lambda^{-1}\mathrm{A}^{\mathrm{T}})$

$\mu = \mu_x, \quad x + \mu_z = Ax + b, \quad A = I, \quad b = \mu_z, \quad \Lambda^{-1} = \Sigma_x, \quad \mathrm{L}^{-1} = \Sigma_z$

Using these values and substituting them in original equation of marginal distribution of P(y):

$p(\mathrm{y}) = \mathcal{N}(\mathrm{y} \mid (\mathrm{Ax} + \mathrm{b}), \mathrm{L}^{-1} + \mathrm{A}\Lambda^{-1}\mathrm{A}^{\mathrm{T}})$

$p(\mathrm{y}) = \mathcal{N}(\mathrm{y} \mid \mu_x + \mu_z, \Sigma_z + \mathrm{A}\Sigma_x\mathrm{A}^{\mathrm{T}})$

$\mathbf{p(y) = \mathcal{N}(y \mid \mu_x + \mu_z, \Sigma_z + \Sigma_x)}$

Ans 5)

First we express the marginal and conditional distribution as:

$$\boldsymbol{p}(\mathbf{x}) = \boldsymbol{\mathcal{N}}(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Lambda}^{-1})$$
$$\boldsymbol{p}(\mathbf{y} \mid \mathbf{x}) = \boldsymbol{\mathcal{N}}(\mathbf{y} \mid \mathbf{Ax} + \mathbf{b}, \mathbf{L}^{-1})$$

$$\boldsymbol{z} = \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{pmatrix}$$

Natural log of joint distribution $p(\mathrm{x}, \mathrm{y})$ is expressed as:

$$\ln\, p(\mathrm{z}) = \ln\, p(\mathrm{x}) + \ln\, p(\mathrm{y} \mid \mathrm{x})$$

$$= -\frac{1}{2}(\mathrm{x} - \mu)^{\mathrm{T}}\Lambda(\mathrm{x} - \mu)\ -\frac{1}{2}(\mathrm{y} - \mathrm{Ax} - \mathrm{b})^{\mathrm{T}}\mathrm{L}(\mathrm{y} - \mathrm{Ax} - \mathrm{b}) + constant$$

$constant\ are\ independent\ of x\ and\ y$

$$= -\frac{1}{2}[x^T\Lambda\mathrm{x} - \mathrm{x}^{\mathrm{T}}\Lambda\mu - \mu^T\Lambda\mathrm{x} +\ \mu^T\Lambda\mu] - \frac{1}{2}[y^TLy - y^TLAx - yb - x^TA^TLy + x^TA^TLAx + x^TA^TLb - b^TLy + b^TLAx + b^TLb]$$

We separate terms with x:

$$= -\frac{1}{2}x^T[\Lambda + \mathrm{A^TLA}]\mathrm{x} - \frac{1}{2}x^T[\Lambda\mu] + \frac{1}{2}x^T[A^TLy] - \frac{1}{2}x^T[A^TLb]$$

$$= -\frac{1}{2}x^T[\Lambda + \mathrm{A^TLA}]\mathrm{x} + x^T[\Lambda\mu + A^TL(y - b)]$$

We can define:

$n = [\Lambda + \mathrm{A^TLA}], m = [\Lambda\mu + A^TL(y - b)]$, hence we can re-write previous equation as:

$$= -\frac{1}{2}x^Tn\mathrm{x} + x^Tm$$

Using completing the square method we can write:

$$= -\frac{1}{2}n^T[x^Tx - 2x^Tmn^{-1} + m^T(n^T)^{-1}n^{-1}m - m^T(n^T)^{-1}n^{-1}m\,]$$

$$= -\frac{1}{2}n^T[x^Tx - 2x^Tmn^{-1} + m^Tm(n^T)^{-1}n^{-1}] + \frac{1}{2}m^Tn^{-1}\,m$$

$$= -\frac{1}{2}n^T[x - mn^{-1}]^T[x - mn^{-1}] + \frac{1}{2}m^Tn^{-1}\,m$$

substitute $mn^{-1} = z$, we can re-write the equation as:

$$= -\frac{1}{2}[x - z]^Tn[x - z] + \frac{1}{2}z^Tn\,z$$

Subtitute the varaiables to get :

$=-\frac{1}{2}(x-z)^T(\Lambda+A^TLA)(x-z)+\frac{1}{2}z^T(\Lambda+A^TLA)z$

Where,

$$z=(\Lambda+A^TLA)^{-1}[\Lambda\mu+A^TL(y-b)]$$

Marginal distribution $p(\mathrm{y})$ is evaluated from :

$$p(\mathrm{y})=\int p(\mathrm{x,y})\mathrm{dx}$$

We take the integral according to x considering y as a constant, accordingly the integral

$$\int_{-\infty}^{\infty} e^{-(x-c)^2}dx=\sqrt{\pi}$$

The first term of the equation will be constant and we can find the y terms as :

$=-\frac{1}{2}y^T[L-LA(\Lambda+A^TLA)^{-1}A^TL]y+y^T\{[L-LA(\Lambda+A^TLA)^{-1}A^TL]b+LA(\Lambda+A^TLA)^{-1}\Lambda\mu\}$

We can extract the covariance matrix $cov[y]$ by taking the inverse of y^Ty coeffeicent

$$L-LA(\Lambda+A^TLA)^{-1}A^TL$$

Using Woodbury inversion formula: $(\mathrm{X}+\mathrm{YZU})^{-1}=\mathrm{X}^{-1}-\mathrm{X}^{-1}\mathrm{Y}(\mathrm{Z}^{-1}+\mathrm{UX}^{-1}\mathrm{Y})^{-1}\mathrm{UX}^{-1}$

$$\mathrm{X}^{-1}=L,$$
$$Y=A,$$
$$Z^{-1}=\Lambda,$$
$$\mathrm{U}=\mathrm{A}^{\mathrm{T}}$$

We can write:

$$\mathbf{cov[y]=(L^{-1}+A\,\Lambda^{-1}A^T)}$$

y^T coeffecient must equal to $E[y](cov[y])^{-1}$:

$E[y](L^{-1}+A\,\Lambda^{-1}\mathrm{A}^{\mathrm{T}})^{-1}=\{[L-LA(\Lambda+A^TLA)^{-1}A^TL]b+LA(\Lambda+A^TLA)^{-1}\Lambda\mu\}$

$E[y]=(L^{-1}+A\,\Lambda^{-1}\mathrm{A}^{\mathrm{T}})\{[L-LA(\Lambda+A^TLA)^{-1}A^TL]b+LA(\Lambda+A^TLA)^{-1}\Lambda\mu\}$

$E[y]=(L^{-1}+A\,\Lambda^{-1}\mathrm{A}^{\mathrm{T}})\{(L^{-1}+A\,\Lambda^{-1}\mathrm{A}^{\mathrm{T}})^{-1}b+LA(\Lambda+A^TLA)^{-1}\Lambda\mu\}$

$E[y]=\{b+(L^{-1}+A\,\Lambda^{-1}\mathrm{A}^{\mathrm{T}})LA(\Lambda+A^TLA)^{-1}\Lambda\mu\}$

Using Woodbury inversion formula, we can write

$(\Lambda+A^TLA)^{-1}=\Lambda^{-1}-\Lambda^{-1}A^T(L^{-1}+A\Lambda^{-1}A^T)^{-1}A\Lambda^{-1}$

$(L^{-1}+A\,\Lambda^{-1}\mathrm{A}^{\mathrm{T}})\,LA(\Lambda^{-1}-\Lambda^{-1}A^T(L^{-1}+A\Lambda^{-1}A^T)^{-1}A\Lambda^{-1})$ =>

$(L^{-1}+A\,\Lambda^{-1}\mathrm{A}^{\mathrm{T}})\,LA\Lambda^{-1}-(L^{-1}+A\,\Lambda^{-1}\mathrm{A}^{\mathrm{T}})\,LA\Lambda^{-1}\mathrm{A}^{\mathrm{T}}(L^{-1}+A\Lambda^{-1}A^T)^{-1}A\Lambda^{-1}$ =>

$(L^{-1} + A\,\Lambda^{-1}\mathrm{A}^{\mathrm{T}})\,LA\Lambda^{-1} - LA\Lambda^{-1}\mathrm{A}^{\mathrm{T}}A\Lambda^{-1} =>$

$L^{-1}LA\Lambda^{-1} + \ A\,\Lambda^{-1}\mathrm{A}^{\mathrm{T}}LA\Lambda^{-1} - A\,\Lambda^{-1}\mathrm{A}^{\mathrm{T}}LA\Lambda^{-1} = A\Lambda^{-1}$

$E[y] = b + A\Lambda^{-1}\,\Lambda\mu = A\mu + b$

$$\boldsymbol{E[y] = A\mu + b}$$