Aston University

Student Name: Muhammad Aizaz Afzal

Student Number: 200261713

Module: EE4108 – Machine Learning

Program: MSc SmartNet (EMJMD)

Assignment # 1

Ans 1)

$$P(Head) = P(Tail) = \frac{1}{2}$$

- A coin is tossed, if head comes then place white ball in the box (H→W)
- Coin is tossed again, if tail comes then place red ball in the box (T→R)
- 3 balls are drawn from the box one by one (with replacement)

$$P(2\ Red\ balls\ placed\ in\ the\ box) = P(2\ White\ balls\ placed\ in\ the\ box) = \frac{1}{2} * \frac{1}{2} = \frac{1}{4}$$

$$P(RRR|RR) = 1, as\ there\ are\ only\ red\ balls\ in\ box$$

$$P(RRR|RW) = \frac{1}{2} * \frac{1}{2} * \frac{1}{2} = \frac{1}{8} = 0.125$$

$$P(RRR|WR) = \frac{1}{2} * \frac{1}{2} * \frac{1}{2} = \frac{1}{8} = 0.125$$

$$P(RRR|WW) = 0, as\ there\ will\ be\ no\ red\ balls\ in\ the\ box$$

$$P(2\ Red\ balls\ in\ box|3\ times\ a\ red\ ball\ is\ drawn\ from\ box) =$$

$$P(3\ times\ a\ red\ ball\ is\ drawn\ from\ box|2\ Red\ balls\ in\ box)P(2\ Red\ balls\ in\ box)/P(3\ times\ a\ red\ ball\ is\ drawn)$$

OR

$$P(RR|RRR) = \frac{P(RRR|RR)}{P(RRR)} * P(RR)$$

P(RRR) = P(RR)*P(RRR|RR) + P(RW)*P(RRR|RW) + P(WR)*P(RRR|WR) + P(WW)*P(RRR|WW)

P(RRR) = (0.25 x 1) + (0.25 x 0.125) + (0.25 x 0.125) + 0

P(RRR) = 5/16

$$P(RR|RRR) = \frac{1 * \left(\frac{1}{4}\right)}{\frac{5}{16}}$$

$$\boldsymbol{P(RR|RRR) = \frac{4}{5}}$$

Ans 2)

For maximum likelihood mean μ:

$$N(x|\mu,\sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right) * e^{-\left(\frac{1}{2\sigma^2}\right)(x-\mu)^2}$$

$$p(x|\mu,\sigma^2) = \prod_{n=1}^{N}\left\{\left(\frac{1}{2\pi\sigma^2}\right) * e^{-\left(\frac{1}{2\sigma^2}\right)(x-\mu)^2}\right\}$$

$$\ln p(x|\mu,\sigma^2) = \ln\left\{\prod_{n=1}^{N}\left(\frac{1}{2\pi\sigma^2}\right) * e^{-\left(\frac{1}{2\sigma^2}\right)(x-\mu)^2}\right\}$$

$$\ln p(x|\mu,\sigma^2) = -N\ln\left(\sqrt{2\pi}\right) - N\ln(\sigma) - \sum_{n=1}^{N}\frac{(x_n-\mu)^2}{2\sigma^2}$$

$$f(x_1,\dots x_n|\mu,\sigma) = \ln p(x|\mu,\sigma^2)$$

$$\frac{\delta f(x_1,\dots x_n|\mu,\sigma)}{\delta\mu} = -\frac{1}{2\sigma^2} * \sum_{n=1}^{N}\frac{\delta(x_n-\mu)^2}{\delta\mu}$$

$$\frac{\delta f(x_1,\dots x_n|\mu,\sigma)}{\delta\mu} = 0$$

$$-\frac{1}{2\sigma^2} * \sum_{n=1}^{N}\frac{\delta(x_n-\mu)^2}{\delta\mu} = 0$$

$$\sum_{n=1}^{N}\frac{\delta(x_n-\mu)^2}{\delta\mu} = 0$$

$$\sum_{n=1}^{N} -2(x_n-\mu) = 0$$

$$\sum_{n=1}^{N}(x_n-\mu) = 0$$

$$\sum_{n=1}^{N} x_n - N\mu = 0$$

$$\sum_{n=1}^{N} x_n = N\mu$$

$$\boldsymbol{\mu_{ML} = \frac{1}{N}\sum_{n=1}^{N} x_n}$$

For maximum likelihood variance σ^2:

$$\frac{\delta f(x_1, \dots x_n|\mu, \sigma)}{\delta\sigma} = -\frac{n}{\sigma} + \sum_{n=1}^{N} \frac{x_n - \mu}{\sigma^3}$$

$$\frac{\delta f(x_1, \dots x_n|\mu, \sigma)}{\delta\sigma} = 0$$

$$-\frac{n}{\sigma} + \sum_{n=1}^{N} \frac{x_n - \mu}{\sigma^3} = 0$$

$$\sum_{n=1}^{N} \frac{x_n - \mu}{\sigma^3} = \frac{n}{\sigma}$$

$$\sigma_{ML}^2 = \sum_{n=1}^{N} \frac{x_n - \mu_{ML}}{\sigma^2}$$

Ans 3)

$$\mathbb{E}[\sigma_{ML}^2] = \mathbb{E}\left[\frac{1}{N}\sum_{n=1}^{N} (x_n - \mu_{ML})^2\right]$$

$$= \frac{1}{N}\mathbb{E}\left[\sum_{n=1}^{N} (x_n - \mu_{ML})^2\right]$$

$$= \frac{1}{N}\mathbb{E}\left[\sum_{n=1}^{N} (x_n^2 - 2x_n\mu_{ML} + \mu_{ML}^2)\right]$$

$$= \frac{1}{N}\mathbb{E}\left[\sum_{n=1}^{N} x_n^2\right] - \frac{1}{N}\mathbb{E}\left[\sum_{n=1}^{N} 2x_n\mu_{ML}\right] + \frac{1}{N}\mathbb{E}\left[\sum_{n=1}^{N} \mu_{ML}^2\right]$$

$$= \mu^2 + \sigma^2 - \frac{2}{N}\mathbb{E}\left[\sum_{n=1}^{N} x_n\left(\frac{1}{N}\sum_{n=1}^{N} x_n\right)\right] + \mathbb{E}[\mu_{ML}^2]$$

$$= \mu^2 + \sigma^2 - \frac{2}{N^2}\mathbb{E}\left[\sum_{n=1}^{N} x_n\left(\sum_{n=1}^{N} x_n\right)\right] + \mathbb{E}\left[\left(\frac{1}{N}\sum_{n=1}^{N} x_n\right)^2\right]$$

$$= \mu^2 + \sigma^2 - \frac{2}{N^2}\mathbb{E}\left[\left(\sum_{n=1}^{N} x_n\right)^2\right] + \frac{1}{N^2}\mathbb{E}\left(\sum_{n=1}^{N} x_n\right)^2\Bigg]$$

$$= \mu^2 + \sigma^2 - \frac{1}{N^2}\mathbb{E}\left(\sum_{n=1}^{N} x_n\right)^2\Bigg]$$

$$= \frac{\big(N\sigma^2(N-1)\big)}{N^2}$$

$$Since, \mathbb{E}\left\{\left(\sum_{n=1}^{N} x_n\right)^2\right\} = N\sigma^2 + N^2\mu^2$$

$$\mathbb{E}[\sigma_{ML}^2] = \mu^2 + \sigma^2 - \frac{1}{N^2}[N(N\mu^2 + \sigma^2)]$$

$$\mathbb{E}[\sigma_{ML}^2] = \frac{N^2\mu^2 + N^2\sigma^2 - N^2\mu^2 - N\sigma^2}{N^2}$$

$$\boldsymbol{\mathbb{E}[\sigma_{ML}^2] = \left(\frac{N-1}{N}\right)\sigma^2}$$

Ans 4)

$$P(x) = \begin{cases} \lambda \mathrm{e}^{(-\lambda x)}, x \geq 0 \\ 0\,, x\ < 0 \end{cases}$$

$$P(\lambda | x_1, x_2, \dots \dots x_n) = \prod_{n=1}^{N} \lambda e^{-\lambda x_n}$$

$$P(\lambda | x_1, x_2, \dots \dots x_n) = \lambda^N e^{-\lambda \sum_{n=1}^{N} x_n}$$

$$\ln\big(P(\lambda | x_1, x_2, \dots \dots x_n)\big) = N ln(\lambda) - \lambda \sum_{n=1}^{N} x_n$$

Differentiate with respect to λ and equate to 0:

$$\frac{\ln\big(P(\lambda | x_1, x_2, \dots \dots x_n)\big)}{\delta\lambda} = N\left(\frac{1}{\lambda}\right) - \sum_{n=1}^{N} x_n$$

$$0 = \frac{N}{\lambda} - \sum_{n=1}^{N} x_n$$

$$\sum_{n=1}^{N} x_n = \frac{N}{\lambda}$$

$$\boldsymbol{\lambda_{ML} = \frac{N}{\sum_{n=1}^{N} x_n}}$$

Ans 5)

$$\boldsymbol{N(x|\mu,\Sigma) = \left(\frac{1}{(2\pi)^{\frac{D}{2}}\,|\Sigma|^{\frac{1}{2}}}\right) * e^{-\left(\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)}}$$

$$x\epsilon R^2, \Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

$$\det(\Sigma) = \sigma_1^2\sigma_2^2 - [\rho^2\sigma_1^2\sigma_2^2] = \sigma_1^2\sigma_2^2[1-\rho^2]$$

$$Adjoint(\Sigma) = \begin{bmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix}$$

$$(\Sigma\ )^{-1} = \frac{\begin{bmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix}}{\sigma_1^2\sigma_2^2[1-\rho^2]}$$

$$\boldsymbol{x} = \begin{bmatrix} \boldsymbol{x_1} \\ \boldsymbol{x_2} \end{bmatrix}, \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu_1} \\ \boldsymbol{\mu_2} \end{bmatrix}$$

$$x-\mu = \begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{bmatrix}$$

$$(x-\mu)^T = [x_1-\mu_1 \qquad x_2-\mu_2]$$

$$(x-\mu)^T\Sigma^{-1} = [x_1-\mu_1 \quad x_2-\mu_2]\frac{\begin{bmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix}}{\sigma_1^2\sigma_2^2[1-\rho^2]}$$

$$[x_1-\mu_1 \quad x_2-\mu_2]\frac{\begin{bmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix}}{\sigma_1^2\sigma_2^2[1-\rho^2]} \Rightarrow$$

$$\Rightarrow \frac{\frac{1}{2}[(x_1-\mu_1)^2\sigma_2^2 - 2(x_2-\mu_2)(x_1-\mu_1)\rho\sigma_1\sigma_2 + (x_2-\mu_2)^2\sigma_1^2]}{\sigma_1^2\sigma_2^2(1-\rho^2)}$$

$$\Rightarrow \frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2\sigma_2^2 - 2(x_2-\mu_2)(x_1-\mu_1)\rho\sigma_1\sigma_2 + (x_2-\mu_2)^2\sigma_1^2}{\sigma_1^2\sigma_2^2}\right]$$

$$\Rightarrow \frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} - \frac{2(x_2-\mu_2)(x_1-\mu_1)\rho}{\sigma_1^2\sigma_2^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]$$

$$P(x_1,x_2) = \frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2(1-\rho^2)}}e^{\left(\frac{-1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} - \frac{2(x_2-\mu_2)(x_1-\mu_1)\rho}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]\right)}$$

$$\boldsymbol{P(x_1,x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}e^{\left(\frac{-1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2} - \frac{2\rho(x_2-\mu_2)(x_1-\mu_1)}{\sigma_1\sigma_2}\right]\right)}}$$

Ans 6 i)

Since,

$$\boldsymbol{P(x_1,x_2) = \frac{e^{\left(\frac{-1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}-\frac{2\rho(x_2-\mu_2)(x_1-\mu_1)}{\sigma_1\sigma_2}\right]\right)}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}}$$

According to the bayes rule, $P(x_2|x_1) = \frac{P(x_1,\ x_2)}{P(x_1)}$

$$p(x_1) = \int\ p(x_1,x_2)\mathrm{d}x_2$$

We use completing the square inside $p(x_1,x_2)$ w.r.t. x_2:

$$2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}\cdot p(x_1,x_2) =>$$

$$\mathrm{e}^{\left(-\frac{1}{2(1-\rho^2)}\frac{(x_1-\mu_1)^2}{\sigma_1^2}\right)}.\,e^{\left(-\frac{1}{2(1-\rho^2)}\left(\frac{(x_2-\mu_2)^2}{\sigma_2^2}-\frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}+\frac{\rho^2(x_1-\mu_1)^2}{\sigma_1^2}\right)\right)}.\,e^{\left(\frac{1}{2(1-\rho^2)}\frac{\rho^2(x_1-\mu_1)^2}{\sigma_1^2}\right)}$$

$$= e^{\left(-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}\right)}.\,e^{\left(-\frac{1}{2\sigma_2^2(1-\rho^2)}\left((x_2-\mu_2)-\frac{\sigma_2\rho(x_1-\mu_1)}{\sigma_1}\right)^2\right)}$$

Now integrate it with respect to x_2:

$$\int\ p(x_1,x_2)\mathrm{d}x_2 \quad = \frac{\mathrm{e}^{\left(-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}\right)}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\cdot\int\ \mathrm{e}^{\left(-\frac{1}{2\sigma^2}(x_2-\mu)^2\right)}\mathrm{d}x_2$$

$$= \frac{\mathrm{e}^{\left(-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}\right)}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\cdot\sqrt{2\pi\sigma_2^2(1-\rho^2)}$$

$$P(x_2|x_1) = \frac{\left(\frac{e^{\left(\frac{-1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}-\frac{2\rho(x_2-\mu_2)(x_1-\mu_1)}{\sigma_1\sigma_2}\right]\right)}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\right)}{\frac{e^{\left(-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}\right)}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\cdot\sqrt{2\pi\sigma_2^2(1-\rho^2)}}$$

$$P(x_2|x_1) = \frac{e^{\left(\frac{(x_1-\mu_1)^2}{2\sigma_1^2}\right)}e^{\left(\frac{-1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}-\frac{2\rho(x_2-\mu_2)(x_1-\mu_1)}{\sigma_1\sigma_2}\right]\right)}}{\sqrt{2\pi\sigma_2^2(1-\rho^2)}}$$

$$\mathbf{P(x_2|x_1)} = \frac{e^{\left(\frac{-1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}-\frac{2\rho(x_2-\mu_2)(x_1-\mu_1)}{\sigma_1\sigma_2}\right]+\left(\frac{(x_1-\mu_1)^2}{2\sigma_1^2}\right)\right)}}{\sqrt{2\pi\sigma_2^2(1-\rho^2)}}$$

Ans 6 ii)

$$\sigma_1 = \sigma_2 = 1$$

$$P(x_2|x_1) = \frac{e^{\left(\frac{-1}{2(1-\rho^2)}[(x_1-\mu_1)^2 + (x_2-\mu_2)^2 - 2\rho(x_2-\mu_2)(x_1-\mu_1)]+(x_1-\mu_1)^2\right)}}{\sqrt{2\pi(1-\rho^2)}}$$

$$P(x_2|x_1) = \frac{e^{\left(\frac{-(x_1-\mu_1)^2}{2(1-\rho^2)}+\frac{(x_2-\mu_2)^2}{2(1-\rho^2)}-\frac{2\rho(x_2-\mu_2)(x_1-\mu_1)}{2(1-\rho^2)}+(x_1-\mu_1)^2\right)}}{\sqrt{2\pi(1-\rho^2)}}$$

$$P(x_2|x_1) = \frac{e^{\left(\frac{-(x_1-\mu_1)^2}{2(1-\rho^2)}+\frac{(x_2-\mu_2)^2}{2(1-\rho^2)}-\frac{2\rho(x_2-\mu_2)(x_1-\mu_1)}{2(1-\rho^2)}+\frac{2(1-\rho^2)(x_1-\mu_1)^2}{2(1-\rho^2)}\right)}}{\sqrt{2\pi(1-\rho^2)}}$$

$$P(x_2|x_1) = \frac{e^{\left(\frac{-(x_1-\mu_1)^2+2(1-\rho^2)(x_1-\mu_1)^2 +(x_2-\mu_2)^2-2\rho(x_2-\mu_2)(x_1-\mu_1)}{2(1-\rho^2)}\right)}}{\sqrt{2\pi(1-\rho^2)}}$$

$$P(x_2|x_1) = \frac{e^{\left(\frac{(x_1-\mu_1)^2(1-2\rho^2)+(x_2-\mu_2)^2-2\rho(x_2-\mu_2)(x_1-\mu_1)}{2(1-\rho^2)}\right)}}{\sqrt{2\pi(1-\rho^2)}}$$

$$\boldsymbol{p(x_2 \mid x_1) = \frac{e^{\left(-\frac{1}{2(1-\rho^2)}\left(\rho(x_1-\mu_1)-(x_2-\mu_2)\right)^2\right)}}{\sqrt{2\pi(1-\rho^2)}}}$$

Ans 7 a)

Mean of uniform distribution:

Let X have a uniform distribution in the interval (a, b). Then the density function of X is

$f(x) = \frac{1}{b-a}$ if $a \le x \le b$ and 0 $otherwise$

The mean is given by:

$$E[X] = \int_a^b \frac{x}{b-a} dx$$

$$= \frac{b^2 - a^2}{2(b-a)} = \frac{b+a}{2}$$

The variance is given by $\boldsymbol{E[X^2] - (E[X])^2}$

$$E[X^2] = \int_a^b \frac{x^2}{b-a} dx$$

$$= \frac{b^3 - a^3}{3(b-a)}$$

$$= \frac{b^2 + ba + a^2}{3}$$

$$Variance = \frac{b^2 + ba + a^2}{3} - \frac{(b+a)^2}{4}$$

$$\boldsymbol{Variance = \frac{(a-b)^2}{12}}$$

For binomial distribution:

$p(x) = \binom{n}{x} p^x q^{n-x}$

Then E(X) $= \sum_{x=0}^{n} x \binom{n}{x} p^x q^{n-x}$

$= \sum_{x=0}^{n} x \dfrac{n!}{(n-x)!\,x!} p^x q^{n-x}$

$= \sum_{x=1}^{n} \frac{n!}{(x-1)!(n-x)!} p^x q^{n-x}$, (Since the $x = 0$ term vanishes)

E(X) $= \sum_{x=1}^{n} \dfrac{n(n-1)!}{(x-1)!\,(n-x)!} (p) p^{x-1} q^{n-x}$

$= np \sum_{x=1}^{n} \dfrac{(n-1)!}{(x-1)!\,(n-x)!} p^{x-1} q^{n-x}$

$= np \sum_{x=1}^{n} \dfrac{(n-1)!}{(x-1)!\,[(n-1)-(x-1)]!} p^{x-1} q^{n-x}$

$= np \sum_{x=1}^{n} \binom{n-1}{x-1} p^{x-1} q^{n-x}$

$= np[\,{}^{n-1}\mathbf{C}_0 q^{n-1} + {}^{n-1}\mathrm{C}_1 p q^{n-2} + {}^{n-1}\mathrm{C}_2 p^2 q^{n-3} + \cdots . + {}^{n-1}\mathrm{C}_{n-1} p^{n-1}]$

$= np$ [This is a blnomial expansion of $(p+q)^{n-1}$]

$= np[p+q]^{n-1}$

But we know that $p + q = 1$ So E(X) $= np(1)^{n-1} = np$

Thus the mean of binomial distribution is np.

Now variance of binomial distribution is: Var $(\mathbf{X}) = \mathbf{E}(\mathbf{X}^2) - [\mathbf{E}(\mathbf{X})]^2$

$\mathrm{E}(\mathrm{X}^2) = \sum_{x=0}^{n} x^2 \binom{n}{x} p^x q^{n-x}$

$= \sum_{x=0}^{n} [x(x-1) + x] \binom{n}{x} p^x q^{n-x}$

$= \sum_{x=0}^{n} [x(x-1)] \binom{n}{x} p^x q^{n-x} + \sum_{x=0}^{n} x \binom{n}{x} p^x q^{n-x}$

$= \sum_{x=2}^{n} \dfrac{x(x-1)n!}{(n-x)!\,x(x-1)(x-2)!} p^x q^{n-x} + np$

$= n(n-1)p^2 \sum_{x=2}^{n} \dfrac{(n-2)!}{(x-2)!\,(n-x)!} p^{x-2} q^{n-x} + np$

$= n(n-1)p^2 \sum_{x=2}^{n} \dfrac{(n-2)!}{(x-2)!\,[(n-2)-(x-2)]!} p^{x-2} q^{n-x} + np$

$= n(n-1)p^2 \sum_{x=2}^{n} \binom{n-2}{x-2} p^{x-2} q^{n-x} + np$

$= n(n-1)p^2[\,{}^{n-2}\mathrm{C}_0 q^{n-2} + {}^{n-2}\mathrm{C}_1 p q^{n-3} + \cdots + {}^{n-2}\mathrm{C}_{n-2} p^{n-2}] + np$

$= n(n-1)p^2[(p+q)^{n-2}] + np$

Since $p + q = 1$, we have $\mathrm{E}(\mathrm{X}^2) = n(n-1)p^2 + np$

Using this, Var (X) $= n(n-1)p^2 + np - (np)^2$

$= n^2p^2 - np^2 + np - n^2p^2$

$\boldsymbol{= np(1-p) = npq}$

Hence the variance of binomial distribution is npq.

Ans 7 b)

Multivariate Gaussian:

$$\boldsymbol{N(x|\mu,\Sigma) = \left(\frac{1}{(2\pi)^{\frac{D}{2}}\,|\Sigma|^{\frac{1}{2}}}\right) * e^{-\left(\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right)}}$$

$$E[x] = \int N(x|\mu,\Sigma).x.dx$$

$$E[x] = \left(\frac{1}{(2\pi)^{\frac{D}{2}}\,|\Sigma|^{\frac{1}{2}}}\right)\int e^{-\left(\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right)}.x.dx$$

$$Z = x - \mu$$

$$x = Z + \mu$$

$$dx = dZ$$

$$E[x] = \left(\frac{1}{(2\pi)^{\frac{D}{2}}\,|\Sigma|^{\frac{1}{2}}}\right)\int e^{-\left(\frac{1}{2}Z^T \Sigma^{-1}Z\right)}(Z+\mu)dZ$$

$$E[x] = \left(\frac{1}{(2\pi)^{\frac{D}{2}}\,|\Sigma|^{\frac{1}{2}}}\right)\int_{-\infty}^{+\infty} e^{-\left(\frac{1}{2}Z^T \Sigma^{-1}Z\right)}(Z)dZ + \int_{-\infty}^{+\infty} e^{-\left(\frac{1}{2}Z^T \Sigma^{-1}Z\right)}(\mu)dZ$$

$\int_{-\infty}^{+\infty} e^{-\left(\frac{1}{2}Z^T \Sigma^{-1}Z\right)}(Z)dZ = 0$, since its an odd function with limits -inf to +inf

$$\left(\frac{1}{(2\pi)^{\frac{D}{2}}\,|\Sigma|^{\frac{1}{2}}}\right)\int_{-\infty}^{+\infty} e^{-\left(\frac{1}{2}Z^T \Sigma^{-1}Z\right)}dZ = 1$$

Therefore,

$$E[x] = 1 * \mu$$

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

We considered the second order moment given by $\mathbb{E}[x^2]$. For the multivariate Gaussian, there are D^2 second order moments given by $\mathbb{E}[x_i x_j]$, which we can group together to form the matrix $\mathbb{E}[\mathrm{xx}^{\mathrm{T}}]$. This matrix can be written as:

$$\mathbb{E}[\mathrm{xx}^{\mathrm{T}}] = \frac{1}{(2\pi)^{\frac{D}{2}}} \frac{1}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \int \exp\left\{-\frac{1}{2}(\mathrm{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathrm{x}-\boldsymbol{\mu})\right\} \mathrm{xx}^{\mathrm{T}}\mathrm{dx}$$

$$Z = x - \mu$$

$$x = Z + \mu$$

$$dx = dZ$$

$$\mathbb{E}[\mathrm{xx}^{\mathrm{T}}] = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} \int \exp\left\{-\frac{1}{2}\mathbf{Z}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{z}\right\}(\mathbf{z}+\boldsymbol{\mu})(\mathbf{z}+\boldsymbol{\mu})^{\mathrm{T}}\mathrm{d}\mathbf{z}$$

$$\mathbb{E}[\mathrm{xx}^{\mathrm{T}}] = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} \int \exp\left\{-\frac{1}{2}\mathbf{Z}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{z}\right\}(\mathbf{z}\mathbf{z}^{\boldsymbol{T}} + \boldsymbol{\mu}\mathbf{z}^{\mathbf{T}} + \boldsymbol{\mu}^{\boldsymbol{T}}\mathbf{z} + \boldsymbol{\mu}\boldsymbol{\mu}^{\boldsymbol{T}})\mathrm{d}\mathbf{z}$$

$\boldsymbol{\mu}\mathbf{z}^{\mathbf{T}} + \boldsymbol{\mu}^{\boldsymbol{T}}\mathbf{z}$, these terms cancel out due to symmetry

$\boldsymbol{\mu}\boldsymbol{\mu}^{\boldsymbol{T}}$, is a constant which can be taken out of the integral

Consider the term involving zz^{T}. Again, we can make use of the eigenvector expansion of the covariance matrix, together with the completeness of the set of eigenvectors, to write:

$$\mathbf{z} = \sum_{j=1}^{D} y_j \mathbf{u}_j$$

where $y_j = \mathbf{u}_j^{\mathrm{T}}\mathbf{z}$, which gives

$$\begin{aligned}
&\frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} \int \exp\left\{-\frac{1}{2}\mathbf{z}^{\mathrm{T}}\Sigma^{-1}\mathbf{z}\right\}\mathrm{zz}^{\mathrm{T}}\mathrm{dz} \\
&= \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} \sum_{i=1}^{D}\sum_{j=1}^{D} \mathbf{u}_i\mathbf{u}_j^{\mathrm{T}} \int \exp\left\{-\sum_{k=1}^{D}\frac{y_k^2}{2\lambda_k}\right\} y_i y_j \mathrm{dy} \\
&= \sum_{i=1}^{D} \mathrm{u}_i \mathbf{u}_i^{\mathrm{T}} \lambda_i = \Sigma
\end{aligned}$$

where I have made use of the eigenvector equation in PRML book equation (2.45), together with the fact that the integral on the right-hand side of the middle line vanishes by symmetry unless $i = j$, and in the final line I have made use of the results (1.50) and (2.55), together with (2.48) from PRML book. Thus, we have:

$$\mathbb{E}[\mathbf{xx}^{\mathbf{T}}] = \boldsymbol{\mu}\boldsymbol{\mu}^{\mathbf{T}} + \boldsymbol{\Sigma}$$