Machine Learning HW1

1)

$$E[x_n x_m] = \mu^2 + I_{nm}\delta^2$$

$x_n, x_m \Rightarrow$ points sampled from Gaussian

$\mu \Rightarrow$ mean

$\delta^2 \Rightarrow$ variance

$$\begin{cases} I_{nm} = 1 & \text{if } n=m \\ I_{nm} = 0 & \text{otherwise } (n \neq m) \end{cases}$$

from equation 1 $\Rightarrow$

$$E[x] = \int_{-\infty}^{\infty} N(x|\mu,\delta^2)\, x\, dx = \mu$$

$$\left.\begin{aligned} x_n &\Rightarrow E[x_n] = \int_{-\infty}^{\infty} N(x_n|\mu,\delta^2)\, x_n\, dx_n = \mu \\ x_m &\Rightarrow E[x_m] = \int_{-\infty}^{\infty} N(x_m|\mu,\delta^2)\, x_m\, dx_m = \mu \end{aligned}\right\} \text{Let there be 2 points } x_n, x_m$$

if $x_n \neq x_m$, $E[x_n * x_m] \Rightarrow \mu^2$

elif $x_n = x_m$, $E[x_n * x_m] \Rightarrow E[x^2] \Rightarrow \int_{-\infty}^{\infty} N(x|\mu,\delta^2)\, x^2\, dx = \mu^2 + \delta^2$

(using equation 2)

Therefore for formula

$$E[x_n, x_m] = \mu^2 + I_{nm}\delta^2$$ where the property is

$$\begin{cases} I_{nm} = 1 & \text{if } n=m\ (x_n = x_m \\ I_{nm} = 0 & \text{if } n\neq m\ (x_n \neq x_m \end{cases}$$

1.2

So, if $n=m (x_n = x_m) \Rightarrow$ it is $E[x_n * x_m] \Rightarrow E[x^2] \Rightarrow \mu^2 + \delta^2$ (eq. 2) w

elif $n \neq m (x_n \neq x_m) \Rightarrow$ it is $E[x_n * x_m] \Rightarrow E[x_n] * E[x_m] \Rightarrow \mu \cdot \mu \Rightarrow \mu^2$ w

(eq. 1)

Prove Eq. 4 and Eq. 5 =>

4 $E[\mu_{ML}] = \mu$

from the book (pg. 45-46), we know that

$$p(x|\mu,\sigma^2) = \prod_{n=1}^{N} N(x_n|\mu,\sigma^2) \quad \text{and}$$

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

Then, Maximizing with respect to μ, we obtain

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

So, therefore

$$E[\mu_{ML}] \Rightarrow \frac{1}{N}\sum_{n=1}^{N} E[x_n]$$

$$\sum_{n=1}^{N} E[x_n] \Rightarrow E[x_1] + E[x_2] + /// + E[x_n]$$ until N, so there are N number of μ

$$N\cdot\mu$$

$$E[\mu_{ML}] = \frac{1}{\cancel{N}}\cdot\cancel{N}\cdot\mu \Rightarrow \mu$$ W

5 $E[\sigma^2_{ML}] = \left(\frac{N-1}{N}\right)\sigma^2$

from the book (pg. 46), we know that

$$\sigma^2_{ML} = \frac{1}{N}\sum_{n=1}^{N}(x_n-\mu_{ML})^2$$ where σ^2_{ML} is sample variance to sample mean μ_{ML} by maximizing with respect to σ^2_{ML}.

$$E[\sigma^2_{ML}] = \frac{1}{N}\sum_{n=1}^{N} E[(x_n-\mu_{ML})^2] \Rightarrow$$ by using $(a-b)^2 \Rightarrow a^2-2ab+b^2$ formula =>

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

So, there will be N number of the formula

$$\frac{1}{\cancel{N}} \cdot \cancel{N}\left(E[x_n^2] - 2E[x_n \cdot \mu_{ML}] + E[(\mu_{ML})^2]\right) \Rightarrow$$

$$\underbrace{E[x_n^2]}_{eq.2} - 2\underbrace{E[x_n \cdot \mu_{ML}]} + \underbrace{E[(\mu_{ML})^2]}_{eq.4} \Rightarrow$$

$$\mu^2 + \delta^2 - \frac{2}{N} E\left[\sum_{n=1}^{N} x_n * \frac{1}{N}\sum_{n=1}^{N} x_n\right] + E\left[\frac{1}{N}\sum_{n=1}^{N} x_n * \frac{1}{N}\sum_{n=1}^{N} x_n\right] \Rightarrow$$

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

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

$$\mu^2 + \delta^2{}^{\backslash N} - \frac{1}{N^{\cancel{2}}}\left[\cancel{N}(N\mu^2 + \delta^2)\right] \Rightarrow \frac{\cancel{N\mu^2} + N\delta^2 - \cancel{N\mu^2} + \delta^2}{N} \Rightarrow \frac{N\delta^2 - \delta^2}{N} \Rightarrow$$

$$\delta^2 \frac{(N-1)}{N}$$

$\boxed{2}$

Let the sum be

$$E[a+b] = E[a] + E[b] : \text{(prove)}$$

$$= \iint (a+b)\, p(a,b)\, da\, db \Rightarrow \quad p(a,b) \Rightarrow p(a) * p(b)$$

$$\iint (a+b)\, p(a)\, p(b)\, da\, db \Rightarrow$$

$$\iint p(a) * a * p(b) + p(a) * b * p(b)\, da\, db \Rightarrow$$

$$\iint a\, p(a) * p(b)\, da\, db + \iint b\, p(a)\, p(b)\, da\, db \Rightarrow$$

$$\int a\, p(a)\, da * \int (p(b)\, db) + \int b\, p(b)\, db + \int p(a)\, da \Rightarrow$$

$$\int p(a)\, da = \int p(b)\, db \Rightarrow 1$$

$$\int a\, p(a)\, da + \int b\, p(b)\, db \Rightarrow$$

$$E[a] + E[b]$$

Covariance matrix => $Cov(y) = Cov(a) + Cov(b)$ (prove) ; $Cov(y) = Cov(a+b)$

$$Cov(a+b) \Rightarrow (E[(a+b - E[a+b]) * (a+b - E[a+b])^T] \Rightarrow$$

$$E[a+b] \Rightarrow E[a] + E[b]$$

$$E[(a+b - E[a] - E[b]) * (a+b - E[a] - E[b])^T] \Rightarrow$$

$$E[(a - E[a] + b - E[b]) * (a - E[a] + b - E[b])^T] \Rightarrow$$

$$E[(a - E[a])(a - E[a]^T + (a - E[a]) * (b - E[b]^T) +$$

$$(b - E[b])(a - E[a])^T + (b - E[b]) * (b - E[b]^T)] \Rightarrow$$

$$\underbrace{E[(a - E[a])(a - E[a])^T]}_{Cov(a)} + \underbrace{E[(b - E[b])(b - E[b]^T)]}_{Cov(b)}$$

$$Cov(a+b) \Rightarrow Cov(a) + Cov(b)$$

$\boxed{3}$

$$\sigma_N^2(x) = \frac{1}{\beta} + \phi(x)^T * S_N \phi(x).$$

prove =>

$$\sigma_{N+1}^2(x) \le \sigma_N^2(x)$$

$$\frac{1}{\beta} + \phi(x)^T * S_{N+1}\phi(x) \le \frac{1}{\beta} + \phi(x)^T * S_N \phi(x) \Rightarrow$$

$$\phi(x)^T * S_{N+1}\phi(x) \le \phi(x)^T * S_N \phi(x) \Rightarrow$$

$$S_{N+1} \Rightarrow S_N^{-1} + \beta\phi(x_{N+1})\phi(x_{N+1})^T)^{-1}$$

$v = \phi(x_{N+1})$

$v^T = \phi(x_{N+1})^T$

$M = S_N$

$$(M + vv^T)^{-1} \Rightarrow M^{-1} - \frac{(M^{-1} * v)(v^T * M^{-1})}{1 + v^T * M^{-1} v}$$

$$S_N - \frac{[S_N * \phi(x_{N+1})][\phi(x_{N+1})^T * S_N]}{1 + \phi(x_{N+1})^T * S_N * \phi(x_{N+1})}$$

$$1 + \phi(x_{N+1})^T * S_N * \phi(x_{N+1}) \ne 0$$

Therefore,

$$\underbrace{\phi(x)^T * \left[S_N - \frac{[S_N * \phi(x)][\phi(x_{N+1})^T * S_N]}{1 + \phi(x_{N+1})^T * S_N * \phi(x_{N+1})} \right]}_{S_{N+1}} * \phi(x) \geq 0$$

along as S_N is positive where $\phi(x)^T * S_N * \phi(x)$ & the equation $\sigma^2_{N+1}(x) \leq \sigma^2_N(x)$ is satisfied. ▲

$\boxed{4}$

$E[\mu] \Rightarrow \frac{a}{a+b}$ (prove) $\Rightarrow$

$$E[\mu] = \int_0^1 \mu \, Beta(\mu | a, b) \, d\mu \Rightarrow$$

$$\int_0^1 \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \mu^{a} (1-\mu)^{b-1} \, d\mu \Rightarrow$$

$$\frac{\Gamma(a+b)\Gamma(a+1)}{\Gamma(a+1+b)\Gamma(a)} \int_0^1 Beta(\mu | a+1, b) \, d\mu$$

$$\frac{\Gamma(a+b)}{\Gamma(a+1+b)} * \frac{\Gamma(a+1)}{\Gamma(a)} \Rightarrow$$

$\Gamma(a+1) \Rightarrow a\Gamma(a)$

$\Gamma(a+1+b) \Rightarrow (a+b)\Gamma(a+b)$

$$\frac{\cancel{\Gamma(a+b)}}{(a+b)\cancel{\Gamma(a+b)}} * \frac{a\cancel{\Gamma(a)}}{\cancel{\Gamma(a)}} \Rightarrow \frac{a}{a+b}$$

$\mathrm{var}[\mu] = \dfrac{ab}{(a+b)^2\ (a+b+1)}$ (prove) $\Rightarrow \underbrace{E[\mu^2]}_{\text{find}} - E[\mu]^2$

$E[\mu^2] \Rightarrow \int_0^1 \mu^2\ \mathrm{Beta}(\mu|a,b)\,d\mu \Rightarrow$

$\int_0^1 \dfrac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\ \mu^{a+1}\ (1-\mu)^{b-1}\ d\mu \Rightarrow$

$\dfrac{\Gamma(a+b)\ \Gamma(a+2)}{\Gamma(a+2+b)\ \Gamma(a)} \int_0^1 \dfrac{\Gamma(a+2+b)\ \Gamma(a)}{\Gamma(a+2)\ \Gamma(b)}\ \mu^{a+1}(1-\mu)^{b-1}\ d\mu \Rightarrow$

$\dfrac{\Gamma(a+b)\ \Gamma(a+2)}{\Gamma(a+2+b)\ \Gamma(a)} \int_0^1 \mathrm{Beta}(\mu|a+2,b)\ d\mu \Rightarrow$

$\dfrac{\Gamma(a+b)}{\Gamma(a+2+b)\ \Gamma(a)} * \dfrac{\Gamma(a+2)}{\Gamma(a)}$

$\Gamma(a+2) \Rightarrow (a+1)\ \Gamma(a+1) \Rightarrow (a+1)\ a \cdot \Gamma(a)$

$\Gamma(a+2+b) \Rightarrow (a+1+b)\ \Gamma(a+1+b) \Rightarrow (a+1+b)\ (a+b)\ \Gamma(a+b)$

$\dfrac{\cancel{\Gamma(a+b)}}{(a+1+b)(a+b)\ \cancel{\Gamma(a+b)}} * \dfrac{(a+1)\ a \cdot \cancel{\Gamma(a)}}{\cancel{\Gamma(a)}} \Rightarrow \dfrac{a(a+1)}{(a+1+b)(a+b)}$

$\mathrm{Var}[\mu] = E[\mu^2] - E[\mu]^2 \Rightarrow$

$\dfrac{a(a+1)}{(a+1+b)\ (a+b)}_{\llcorner a+b} - \dfrac{a^2}{(a+b)^2}_{\llcorner a+1+b} \Rightarrow$

$\dfrac{a(a+1)(a+b) - a^2(a+1+b)}{(a+b)^2\ (a+1+b)} \Rightarrow$

$\underbrace{(a^2+a)(a+b)}_{a^3+a^2b+a^2+ab} - (a^3+a^2+a^2b) \Rightarrow \cancel{a^3} + \cancel{a^2b} + \cancel{a^2} + ab \cancel{-a^3} \cancel{-a^2} \cancel{-a^2b} \Rightarrow ab$

$\mathrm{var}[\mu] \Rightarrow \dfrac{ab}{(a+b)^2\ (a+1+b)}$