1.

let $y = x - \mu$

$$\int_{R^k} \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)} dx$$

$$= \int_{R^k} \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} e^{-\frac{1}{2} y^T \Sigma^{-1} y} dy$$

$\because$ Σ is a positive define matrix

$\therefore$ exist a orthogonal matrix P s.t. $P^T P = P P^T = I$

and a diagonal matrix $D = \begin{pmatrix} \lambda_1 & & 0 \\ \vdots & \lambda_2 & \\ 0 & & \ddots \ \lambda_k \end{pmatrix}$ where $\lambda_1, \lambda_2 \dots \lambda_k$ are eigenvalues of Σ

therefore $\Sigma = PDP^T$, $\Sigma^{-1} = (PDP^T)^{-1} = PD^{-1}P^T$, where $D = \begin{pmatrix} \frac{1}{\lambda_1} & 0 \cdots 0 \\ 0 & \frac{1}{\lambda_2} & \\ \vdots & & \ddots \\ 0 & & \frac{1}{\lambda_k} \end{pmatrix}$

$$= \int_{R^k} \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} e^{-\frac{1}{2} y^T (PD^{-1}P^T) y} dy$$

let $z = P^T y$ $\quad \because |P| = 1$

$$= \int_{R^k} \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} e^{-\frac{1}{2}(z^T P^T)(PD^{-1}P^T)(Pz)} dz$$

$$= \int_{R^k} \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} e^{-\frac{1}{2} z^T D^{-1} z} dz$$

$$= \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} \int_{R^k} e^{-\frac{1}{2} z^T D^{-1} z} dz$$

$$= \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} \int_{R^k} e^{-\frac{1}{2} \sum_{i=1}^{k} \frac{z_i^2}{\lambda_i}} dz$$

$$= \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} \prod_{i=1}^{k} \left(\int_{-\infty}^{\infty} e^{-\frac{z_i^2}{2\lambda_i}} dz_i \right)$$

$$= \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} \prod_{i=1}^{k} \left(\sqrt{\frac{\pi}{\frac{1}{2\lambda_i}}} \right)$$

$$= \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} \sqrt{(2\pi)^k} \sqrt{\prod_{i=1}^{k} \lambda_i}$$

$$= \frac{1}{\sqrt{(2\pi)^k \prod_{i=1}^{k} \lambda_i}} \sqrt{(2\pi)^k} \sqrt{\prod_{i=1}^{k} \lambda_i} = 1 \ \#$$

2.

(a)

$$(AB)_{ij} = \sum_{k=1}^{n} A_{ik} B_{kj}$$

$$\text{trace}(AB) = \sum_{i=1}^{n} (AB)_{ii} = \sum_{i=1}^{n} \sum_{k=1}^{n} A_{ik} B_{ki}$$

$$\left(\frac{\partial\, \text{trace}(AB)}{\partial A} \right)_{ab} = \frac{\partial}{\partial A_{ab}} \left(\sum_{i=1}^{n} \sum_{k=1}^{n} A_{ik} B_{ki} \right)$$

$$\frac{\partial A_{ik}}{\partial A_{ab}} = \begin{cases} 1 & \text{if } a=i \text{ and } k=b \\ 0 & \text{, o.w} \end{cases}$$

hence $\frac{\partial}{\partial A_{ab}} \left(\sum_{i=1}^{n} \sum_{k=1}^{n} A_{ik} B_{ki} \right) = 0 \cdot \sum_{i \neq a}^{n} \sum_{k \neq b}^{n} B_{ki} + 1 \cdot B_{ba} = (B_{ab})^T$

$$\therefore \frac{\partial}{\partial A} \text{trace}(AB) = B^T$$

(b) $x^T A x$ is a 1x1 matrix also a scalar

$$\therefore x^T A x = \text{trace}(x^T A x)$$

let A be a nxm metrix B be a mxn matrix

$$\text{trace}(AB) = \sum_{i=1}^{n} \sum_{j=1}^{m} A_{ij} B_{ji}$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{m} B_{ji} A_{ij}$$

$$= \text{trace}(BA)$$

let A.B.C be any matrix

$$\operatorname{trace}(ABC)$$
$$= \operatorname{trace}((BC)A)$$
$$= \operatorname{trace}(BCA)$$
$$= \operatorname{trace}((CA)B)$$
$$= \operatorname{trace}(CAB)$$

$$\therefore \quad x^T A x = \operatorname{trace}(x^T A x) = \operatorname{trace}(x x^T A)$$

(c)

Suppose $X_i \xrightarrow{iid} N_n(\mu, \Sigma) \quad i = 1 \ldots m$

$$f(x_i, \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x_i - \mu)^T \Sigma^{-1} (x_i - \mu)\right)$$

$$L(\mu, \Sigma) = \prod_{i=1}^{m} f(x_i, \mu, \Sigma)$$

$$l(\mu, \Sigma) = \sum_{i=1}^{m} \left[-\frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma| - \frac{1}{2}(x_i - \mu)^T \Sigma^{-1} (x_i - \mu)\right]$$

$$= -\frac{mn}{2}\ln(2\pi) - \frac{m}{2}\ln|\Sigma| - \frac{1}{2}\sum_{i=1}^{m}(x_i - \mu)^T \Sigma^{-1} (x_i - \mu)$$

$$\frac{\partial l(\mu, \Sigma)}{\partial \mu} = -\frac{1}{2}\sum_{i=1}^{m}\frac{\partial}{\partial \mu}\left((x_i - \mu)^T \Sigma^{-1} (x_i - \mu)\right)$$

$$= -\frac{1}{2}\sum_{i=1}^{m}\left(-2\Sigma^{-1}(x_i - \mu)\right) = \Sigma^{-1}\sum_{i=1}^{m}(x_i - \mu) = 0$$

$$\Rightarrow \sum_{i=1}^{m} x_i - m\mu = 0$$

$$\Rightarrow \underline{\hat{\mu} = \frac{\sum_{i=1}^{m} x_i}{m}} \ \#$$

$$l(\mu,\Sigma) = -\frac{mn}{2}\ln(2\pi) - \frac{m}{2}\ln|\Sigma| - \frac{1}{2}\sum_{i=1}^{m}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu)$$

$$= -\frac{mn}{2}\ln(2\pi) - \frac{m}{2}\ln|\Sigma| - \frac{1}{2}\operatorname{trace}\left(\left(\sum_{i=1}^{m}(x_i-\mu)\Sigma^{-1}(x_i-\mu)\right)\right)$$

$$= -\frac{mn}{2}\ln(2\pi) - \frac{m}{2}\ln|\Sigma| - \frac{1}{2}\operatorname{trace}\left(\left(\sum_{i=1}^{m}(x_i-\mu)(x_i-\mu)^T\right)\Sigma^{-1}\right)$$

let $S = \sum_{i=1}^{m}(x_i-\mu)(x_i-\mu)^T$

$$= -\frac{mn}{2}\ln(2\pi) - \frac{m}{2}\ln|\Sigma| - \frac{1}{2}\operatorname{trace}(S\Sigma^{-1})$$

$$\because \frac{d}{dA}\ln|A^{-1}| = \frac{d}{dA}(-\ln|A|) = -(A^{-1})^T$$

$$\therefore \frac{dl}{d\Sigma^{-1}} = \frac{m}{2}(\Sigma^{-1})^T - \frac{1}{2}S^T$$

$$= \frac{m}{2}\Sigma^T - \frac{1}{2}S^T = 0$$

$$\Rightarrow m\Sigma^T = S^T$$

$$\Rightarrow m\Sigma = S$$

$$\Rightarrow \hat{\Sigma} = \frac{S}{m} = \frac{\sum_{i=1}^{m}(x_i-\mu)(x_i-\mu)^T}{m} = \underline{\frac{\sum_{i=1}^{m}(x_i-\hat{\mu})(x_i-\hat{\mu})^T}{m}} \ *$$

3.

1. GDA假設成立時，其後驗 $P(y|x)$ 是 logistic 函數，是不是可以說 GDA模型比 logistic model 更強大或更嚴格？那如果是這樣訓練效果一定會更好？如果不是的話 使用這兩者 model 是否有規則可以來判斷