2. (a)

Let $A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \ddots & & \\ a_{n1} & & & a_{nn} \end{pmatrix}$ $B = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ \vdots & \ddots & & \\ b_{n1} & & & b_{nn} \end{pmatrix}$

$$\Rightarrow tr(AB) = tr \begin{pmatrix} \sum_{j=1}^{n} a_{1j} b_{j1} & & \cdots & \\ & \sum_{j=1}^{n} a_{2j} b_{j2} & & \\ \vdots & & \ddots & \\ & & & \sum_{j=1}^{n} a_{nj} b_{jn} \end{pmatrix}$$

$$= \sum_{k=1}^{n} \left(\sum_{j=1}^{n} a_{kj} b_{jk} \right)$$

$$\therefore \frac{\partial}{\partial A} tr(AB) = \frac{\partial}{\partial A} \left(\sum_{k=1}^{n} \sum_{j=1}^{n} a_{kj} b_{jk} \right)$$

$$= \begin{pmatrix} \frac{\partial s}{\partial a_{11}} & \frac{\partial s}{\partial a_{12}} & \cdots & \frac{\partial s}{\partial a_{1n}} \\ \vdots & \ddots & & \\ \frac{\partial s}{\partial a_{n1}} & & & \frac{\partial s}{\partial a_{nn}} \end{pmatrix} \quad \text{where } s := \sum_{k=1}^{n} \sum_{j=1}^{n} a_{kj} b_{jk}$$

$$= \begin{pmatrix} b_{11} & b_{21} & b_{31} & \cdots & b_{n1} \\ \vdots & \ddots & & & \\ b_{1n} & & & & b_{nn} \end{pmatrix}$$

$$= B^T$$

2. (b)

Let $x = (x_1, x_2, \cdots x_n)^t$ A is similar to (a)

$$\Rightarrow x^T A x = x^T (Ax)$$

$$= x^T \begin{pmatrix} \sum_{j=1}^{n} a_{1j} x_j \\ \sum_{j=1}^{n} a_{2j} x_j \\ \vdots \\ \sum_{j=1}^{n} a_{nj} x_j \end{pmatrix}$$

$$= (x_1, x_2, \cdots x_n) \begin{pmatrix} \sum_{j=1}^{n} a_{1j} x_j \\ \sum_{j=1}^{n} a_{2j} x_j \\ \vdots \\ \sum_{j=1}^{n} a_{nj} x_j \end{pmatrix}$$

$$= x_1 \cdot \sum_{j=1}^{n} a_{1j} x_j + x_2 \cdot \sum_{j=1}^{n} a_{2j} x_j + \cdots + x_n \sum_{j=1}^{n} a_{nj} x_j$$

$$= \sum_{k=1}^{n} \sum_{j=1}^{n} x_k a_{kj} x_j$$

and $$tr(xx^t A) = tr\left(\begin{pmatrix} x_1 x_1 & x_1 x_2 & \cdots & x_1 x_n \\ \vdots & \ddots & & \\ x_n x_1 & & & x_n x_n \end{pmatrix} A \right)$$

$$= tr \begin{pmatrix} x_1 \cdot \sum_{k=1}^{n} x_k a_{k1} & & \cdots \\ & \ddots & \\ \vdots & & x_n \sum_{k=1}^{n} x_k a_{kn} \end{pmatrix}$$

$$= \sum_{j=1}^{n} x_j \sum_{k=1}^{n} a_{kj} x_k = \sum_{j=1}^{n} \sum_{k=1}^{n} x_j a_{kj} x_k$$

$$= \sum_{j=1}^{n} \sum_{k=1}^{n} x_k a_{kj} x_j = \sum_{k=1}^{n} \sum_{j=1}^{n} x_k a_{kj} x_j$$

$$\therefore x^t A x = tr(xx^t A)$$

2(c)

Let $x_1, x_2, \cdots, x_m \in \mathbb{R}^k$ be i.i.d. sample from $x_i \sim N(\mu, \Sigma)$

where $\mu \in \mathbb{R}^k$ is mean $\Sigma \in \mathbb{R}^{k \times k}$ is covariance matrix

and the p.d.f is $f(x) = \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} \exp\{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\}$

$\therefore$ The likelihood function is

$$L(\mu, \Sigma) = \prod_{i=1}^{m} \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} \exp\{-\frac{1}{2}(x_i-\mu)^T \Sigma^{-1}(x-\mu)\}$$

$$= \left[(2\pi)^k |\Sigma|\right]^{-\frac{m}{2}} \cdot \exp\left\{\sum_{i=1}^{m}\left[-\frac{1}{2}(x_i-\mu)^T \Sigma^{-1}(x-\mu)\right]\right\}$$

$$\Rightarrow \ln(L) = -\frac{m}{2}\left(\log(2\pi)^k + \log|\Sigma|\right) + \sum_{i=1}^{m}\left[-\frac{1}{2}(x_i-\mu)^T \Sigma^{-1}(x-\mu)\right]$$

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

Find $\hat{\mu}$

Let $\frac{\partial}{\partial\mu}(\ln(L)) = 0$

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

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

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

#

Find $\hat{\Sigma}$

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

$$\overset{\text{by (b)}}{=} C - \frac{m}{2}\log|\Sigma| - \frac{1}{2}\operatorname{tr}\left(\sum_{i=1}^{m}(x-\mu)(x-\mu)^T \Sigma^{-1}\right), \quad C \in \mathbb{R}$$

$$\therefore \frac{\partial}{\partial \Sigma}(\ln(L)) = \frac{\partial}{\partial \Sigma}\left(-\frac{m}{2}\log|\Sigma|\right) - \frac{1}{2}\frac{\partial}{\partial \Sigma}\left(tr\left(\sum_{i=1}^{m}(x-\mu)(x-\mu)^T\Sigma^{-1}\right)\right)$$

$$\overset{\text{by (a)}}{=} -\frac{m}{2}\left(\Sigma^{-1}\right)^T + \frac{1}{2}\left(\Sigma^{-1}\sum_{i=1}^{m}(x-\mu)(x-\mu)^T\Sigma^{-1}\right)^T$$

$$= -\frac{m}{2}\Sigma^{-1} + \frac{1}{2}\Sigma^{-1}\sum_{i=1}^{m}(x-\mu)(x-\mu)^T\Sigma^{-1}$$

Let $\frac{\partial}{\partial \Sigma}(\ln(L)) = 0$

$$\Rightarrow -\frac{m}{2}\Sigma^{-1} + \frac{1}{2}\Sigma^{-1}\sum_{i=1}^{m}(x-\mu)(x-\mu)^T\Sigma^{-1} = 0$$

$$\Rightarrow \sum_{i=1}^{m}(x-\mu)(x-\mu)^T = m\cdot\Sigma$$

$$\therefore \hat{\Sigma} = \frac{1}{m}\sum_{i=1}^{m}(x-\mu)(x-\mu)^T$$

#

3.

GDA假設條件為高斯分布，但若資料不屬於高斯分布的話，模型會如何變化？