1.

$\because \Sigma$ is positive definite $\Rightarrow \Sigma = R^T R$ for some $R \in \mathbb{R}^{k\times k}$ and R is also positive (R^{-1} is also poistive)

$$\Rightarrow R^{-1} = S^{-1}\Lambda S, \text{ where } \Lambda = \begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_k \end{pmatrix}, \lambda_1, \cdots, \lambda_k \text{ are eigenvalues of } R^{-1}$$

$$\Rightarrow R^{-1} = \sum_{n=1}^{k} S^{-1}\Lambda_n S, \text{ where } \Lambda_n = \begin{cases} a_{nn} = \lambda_n \\ 0, \text{ otherwise} \end{cases}$$

$$\therefore (x-\mu)^T \Sigma^{-1}(x-\mu) = \left(R^{-1}(x-\mu)\right)^T R^{-1}(x-\mu) = \|R^{-1}(x-\mu)\|^2$$
$$= \left\|\sum_{n=1}^{k} S^{-1}\Lambda_n S(x-\mu)\right\|^2$$
$$= \sum_{n=1}^{k} \|S^{-1}\Lambda_n S(x-\mu)\|^2$$

$$\Rightarrow \int_{\mathbb{R}^k} e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}dx = \int_{\mathbb{R}^k} e^{-\frac{1}{2}\sum_{n=1}^{k}\|S^{-1}\Lambda_n S(x-\mu)\|^2}dx$$

$$= \int_{\mathbb{R}^k} e^{-\frac{1}{2}\|S^{-1}\Lambda_1 S(x-\mu)\|^2}\cdot e^{-\frac{1}{2}\|S^{-1}\Lambda_2 S(x-\mu)\|^2}\cdot \ldots \cdot e^{-\frac{1}{2}\|S^{-1}\Lambda_k S(x-\mu)\|^2}dx$$

$$= \left(\int_{-\infty}^{\infty} e^{-\frac{1}{2}\lambda_1^2(x-\mu_1)^2}dx_1\right)\cdot\left(\int_{-\infty}^{\infty} e^{-\frac{1}{2}\lambda_2^2(x-\mu_2)^2}dx_2\right)\cdot \ldots \cdot\left(\int_{-\infty}^{\infty} e^{-\frac{1}{2}\lambda_k^2(x-\mu_k)^2}dx_k\right)$$

$$= \left(\int_{-\infty}^{\infty} e^{-\left(\frac{\lambda_1(x-\mu_1)}{\sqrt{2}}\right)^2}dx_1\right)\left(\int_{-\infty}^{\infty} e^{-\left(\frac{\lambda_2(x-\mu_2)}{\sqrt{2}}\right)^2}dx_2\right) \quad \left(\int_{-\infty}^{\infty} e^{-\left(\frac{\lambda_k(x-\mu_k)}{\sqrt{2}}\right)^2}dx_k\right)$$

$$= \frac{\sqrt{(2\pi)^k}}{\lambda_1\cdot\lambda_2\cdot\ldots\cdot\lambda_k} \quad \text{—— (1)}$$

$$\because \Sigma = R^T R \Rightarrow \frac{1}{\lambda_1^2}, \frac{1}{\lambda_2^2} \cdots, \frac{1}{\lambda_k^2} \text{ are the eigenvalues of } \Sigma$$

$$\Rightarrow |\Sigma| = \frac{1}{(\lambda_1\lambda_2\cdots\lambda_k)^2} \quad \text{—— (2)}$$

$$\therefore \int_{\mathbb{R}^k} f(x)\,dx = \sqrt{\frac{(\lambda_1\lambda_2\cdots\lambda_k)^2}{(2\pi)^k}}\cdot\frac{\sqrt{(2\pi)^k}}{\lambda_1\lambda_2\cdots\lambda_k} = 1 \quad \text{by (1) and (2)}$$

2.

(a)

Let $A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & & \vdots \\ \vdots & & \ddots & \vdots \\ a_{n1} & \cdots & \cdots & a_{nn} \end{pmatrix}$, $B = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & & \vdots \\ \vdots & & \ddots & \vdots \\ b_{n1} & \cdots & \cdots & b_{nn} \end{pmatrix}$

$$trace(AB) = \sum_{k=1}^{n}\left(\sum_{s=1}^{n} a_{ks} b_{sk}\right)$$

$$\Rightarrow \frac{\partial}{\partial a_{ij}} trace(AB) = b_{ji}, \text{ for } 1 \le i,j \le n$$

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

(b)

$$x^T A x = (x_1\ x_2\ \cdots\ x_n) \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & & \vdots \\ \vdots & & \ddots & \vdots \\ a_{n1} & \cdots & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \sum_{1 \le i,j \le n} a_{ij} x_i x_j$$

$$x x^T A = \begin{pmatrix} x_1 x_1 & x_1 x_2 & \cdots & x_1 x_n \\ x_2 x_1 & x_2 x_2 & & \vdots \\ \vdots & & \ddots & \vdots \\ x_n x_1 & \cdots & & x_n x_n \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & & \vdots \\ \vdots & & \ddots & \vdots \\ a_{n1} & \cdots & \cdots & a_{nn} \end{pmatrix}$$

$$\therefore trace(x x^T A) = \sum_{k=1}^{n} \sum_{s=1}^{n} x_k x_s a_{sk} = \sum_{1 \le i,j \le n} a_{ij} x_i x_j$$

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

(c)

Let Likelihood function $L(\theta) = \prod_{i=1}^{N} \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} e^{-\frac{1}{2}(x_i-\mu)^T \Sigma^{-1} (x_i-\mu)}$

$$= \left(\frac{1}{(2\pi)^k |\Sigma|}\right)^{\frac{N}{2}} \prod_{i=1}^{N} e^{-\frac{1}{2}(x_i-\mu)^T \Sigma^{-1} (x_i-\mu)}$$

Let $l(\theta) = \ln L(\theta) = \frac{N}{2} \ln \frac{1}{(2\pi)^k} + \frac{N}{2} \ln \frac{1}{|\Sigma|} - \frac{1}{2} \sum_{i=1}^{N} (x_i-\mu)^T \Sigma^{-1} (x_i-\mu)$

①

$$\Rightarrow \frac{\partial l}{\partial \mu_s} = \sum_{i=1}^{N} \left(-\frac{1}{2} \frac{\partial}{\partial \mu_s} \left(\sum_{1 \le m,n \le k} \Sigma^{-1}_{mn} (x_i-\mu)_m (x_i-\mu)_n\right)\right)$$

$$= \sum_{i=1}^{N} \left(\Sigma^{-1}_{mn} (x_i-\mu)_s + \sum_{m=s,\, n \ne s} \Sigma^{-1}_{mn} (x_i-\mu)_n + \sum_{m \ne s,\, n=s} \Sigma^{-1}_{mn} (x_i-\mu)_m\right)$$

$\because \Sigma$ is symmetric (by positive)

$$\Rightarrow \frac{\partial l}{\partial \mu} = \sum_{i=1}^{N} (x_i-\mu)^T \Sigma^{-1} = 0$$

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

#

②

$$l(\theta) = \frac{N}{2} \ln \frac{1}{(2\pi)^k} + \frac{N}{2} \ln |\Sigma^{-1}| - \frac{1}{2} \sum_{i=1}^{N} trace\left(\Sigma^{-1} (x_i - \mu)(x_i - \mu)^T\right)$$

$$\Rightarrow \frac{\partial l}{\partial \Sigma^{-1}} = \frac{N}{2} \Sigma - \frac{1}{2} \sum_{i=1}^{N} (x_i - \mu)(x_i - \mu)^T = 0$$

$$\Rightarrow \underline{\hat{\Sigma} = \frac{1}{N} \sum_{i=1}^{N} (x_i - \hat{\mu})(x_i - \hat{\mu})^T} \quad \#$$

3. How does GDA perform if the true class-conditional distributions are significantly non-Gaussian.