1. (a). $x^T A x$

$$= x^T \cdot \begin{bmatrix} A_{1\cdot} x \\ \vdots \\ A_{d\cdot} x \end{bmatrix}$$

$A_{i\cdot}$ is ith row of A.

$$= x_1 A_{1\cdot} x + \cdots + x_d A_{d\cdot} x$$

$$= x_1 \cdot (A_{11} x_1 + A_{12} x_2 + \cdots + A_{1d} x_d) + \cdots + x_d (A_{d1} x_1 + \cdots + A_{dd} x_d)$$

$$= A_{11} x_1^2 + A_{12} x_1 x_2 + \cdots + A_{1d} x_1 x_d$$
$$+ \cdots$$
$$+ A_{d1} x_d x_1 + A_{d2} x_d x_2 + \cdots + A_{dd} x_d^2$$

let

$$x x^T = \begin{bmatrix} x_1^2 & \cdots & x_1 x_d \\ \vdots & \ddots & \vdots \\ x_d x_1 & \cdots & x_d^2 \end{bmatrix}.$$

$= B$. (symmetric)

$$= \sum_{j=1}^{d} A_{1j} x_1 \cdot x_j + \sum_{j=1}^{d} A_{2j} x_2 \cdot x_j + \cdots + \sum_{j=1}^{d} A_{dj} \cdot x_d x_j$$

$$= \sum_{j=1}^{d} \sum_{i=1}^{d} A_{ij} x_i x_j$$

$$= A_{1\cdot} B_{\cdot 1} + A_{2\cdot} B_{\cdot 2} + \cdots + A_{d\cdot} \cdot B_{d\cdot}$$

$$= \sum_{i=1}^{d} A_{i\cdot} B_{\cdot i}$$

$$= tr[AB]$$

$$= tr[A x x^T]$$

1 b).

$$p(D|\Sigma) = \prod_{n=1}^{N} \left\{ \frac{1}{(2\pi)^{d/2} |\Sigma|^{\frac{1}{2}}} \cdot \exp\left\{ -\frac{1}{2}(x_n-\mu)^T \Sigma^{-1} (x_n-\mu) \right\} \right.$$

$$= \left(\frac{1}{(2\pi)^{d/2} |\Sigma|^{\frac{1}{2}}} \right)^N \cdot \exp\left\{ \sum_{n=1}^{N} -\frac{1}{2}(x_n-\mu)^T \Sigma^{-1} (x_n-\mu) \right\}.$$

$$= \frac{1}{(2\pi)^{Nd/2}} |\Sigma^{-1}|^{N/2} \cdot \exp\left\{ -\frac{1}{2}\sum_{n=1}^{N} tr\left[\Sigma^{-1} (x_n-\mu)^T (x_n-\mu) \right] \right\}$$

$$= \frac{1}{(2\pi)^{Nd/2}} |\Sigma^{-1}|^{N/2} \exp\left\{ -\frac{1}{2} tr\left[\sum_{n=1}^{N} \Sigma^{-1} (x_n-\mu)^T (x_n-\mu) \right] \right\}$$

1.c) $\tilde{\Sigma} = \frac{1}{N}\sum^{N}(x_n-\mu)(x_n-\mu)^T$ $\quad A = \Sigma^{-1}\tilde{\Sigma} \quad eig(A) = \{\lambda_1, \dots, \lambda_d\}$

$$p(D|\Sigma) = \frac{1}{(2\pi)^{Nd/2}} \cdot |\Sigma^{-1}|^{N/2} \cdot \exp\left\{ -\frac{N}{2} tr\left[\Sigma^{-1} \tilde{\Sigma} \right] \right\} \quad , \text{ let } \quad k = \frac{1}{(2\pi)^{Nd/2}}$$

$$= k \cdot |A\tilde{\Sigma}^{-1}|^{N/2} \exp\left\{ -\frac{N}{2} tr[A] \right\}$$

$$= k \cdot \frac{1}{|\tilde{\Sigma}|^{N/2}} |A|^{N/2} \exp\left\{ -\frac{N}{2} \sum_{i=1}^{d} \lambda_i \right\}$$

$$= \frac{1}{(2\pi)^{Nd/2} |\tilde{\Sigma}|^{N/2}} \left(\prod_{i=1}^{d} \lambda_i \right)^{N/2} \exp\left\{ -\frac{N}{2} \sum_{i=1}^{d} \lambda_i \right\}$$

1 d). $\hat{\Sigma} = \underset{\Sigma}{\text{argmax}} \; P(D|\Sigma)$

$$= \underset{\lambda_i}{\text{argmin}} - \ln\left[\frac{1}{(2\pi)^{Nd/2} |\tilde{\Sigma}|^{N/2}} \left(\prod_{i=1}^{d} \lambda_i \right)^{\frac{N}{2}} \cdot \exp\left\{ -\frac{N}{2} \sum_{i=1}^{d} \lambda_i \right\} \right]$$

$$= \underset{\lambda_i}{\text{argmin}} \; \frac{Nd}{2} \ln(2\pi) + \frac{N}{2} \ln(|\tilde{\Sigma}|) - \frac{N}{2} \ln\left(\prod_{i=1}^{d} \lambda_i \right) + \frac{N}{2} \sum_{i=1}^{d} \lambda_i$$

$$\nabla_\lambda - \ln(P(D|\Sigma)) = \nabla_\lambda \left[-\frac{N}{2} \sum_{i=1}^{d} \ln(\lambda_i) + \frac{N}{2} \sum_{i=1}^{d} \lambda_i \right]$$

$$0 = \nabla_\lambda \left[\sum_{i=1}^{d} \lambda_i - \ln(\lambda_i) \right].$$

$$0 = \left[1 - \frac{1}{\lambda_1} \quad 1 - \frac{1}{\lambda_2} \quad \cdots \quad 1 - \frac{1}{\lambda_i} \right] \Rightarrow \lambda = 1. \Rightarrow A = I.$$

(e) Prove $\hat{\Sigma}_{ML} = \frac{1}{N} \sum_{n=1}^{N} (x_n - \mu)(x_n - \mu)^T = \tilde{\Sigma}$

$$A = \tilde{\Sigma} \hat{\Sigma}^{-1} = I.$$

$$\hat{\Sigma} = \tilde{\Sigma} = \frac{1}{N} \sum_{n=1}^{N} (x_n - \mu)(x_n - \mu)^T.$$

(f) Take derivative w.r.t to Σ, do some tedious matrix calculus, then get $\hat{\Sigma}$.

set $\frac{d(P)}{d\Sigma} = 0$,

(g) $\hat{\Sigma}_{ML}$ is a biased estimator of Σ

$$E(\hat{\Sigma}_{ML}) = \frac{N-1}{N} \Sigma \neq \Sigma$$

by defining $\hat{\mu} = \frac{1}{N} \sum_{n=1}^{N} x_n$ & $\hat{\Sigma}_{ML} = \underline{\frac{1}{N-1} \sum_{n=1}^{N} (x_n - \hat{\mu})(x_n - \hat{\mu})^T}$

we can obtain an unbiased estimator of Σ.

2. See Jupyter Notebook

3. See Jupyter Notebook