## Written assignment

1. Given

$$f(x) = \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)},$$

where $x, \mu \in \mathbb{R}^k$, $\Sigma$ is a $k$-by-$k$ positive definite matrix and $|\Sigma|$ is its determinant. Show that $\int_{\mathbb{R}^k} f(x)\, dx = 1$.

1. Since $\Sigma$ is positive definite matrix.

$\exists L \in$ 下三角 metrix s.t. $\Sigma = LL^T$, and $|\Sigma| = |L|^2$, then $|L| = \sqrt{|\Sigma|}$

Let $z = L^{-1}(x-\mu)$, then $x = \mu + Lz$

By the differential transformation, we have $dx = |L|\, dz$.

Thus, $(x-\mu)^T \Sigma^{-1}(x-\mu) = (x-\mu)^T (LL^T)^{-1}(x-\mu) = (x-\mu)^T L^{-T} L^{-1}(x-\mu)$

$$= (L^{-1}(x-\mu))^T (L^{-1}(x-\mu)) = z^T z.$$

Then we have

$$\int_{\mathbb{R}^k} f(x)\, dx = \int_{\mathbb{R}^k} \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} \exp\{-\tfrac{1}{2} z^T z\} \cdot |L|\, dz$$

By $|L| = \sqrt{|\Sigma|}$, we have

$$\int_{\mathbb{R}^k} \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} \sqrt{|\Sigma|} \exp\{-\tfrac{1}{2} z^T z\}\, dz = \int_{\mathbb{R}^k} \frac{1}{\sqrt{(2\pi)^k}} \exp\{-\tfrac{1}{2} z^T z\}\, dz$$

Since $z^T z = \sum_{i=1}^{k} z_i^2$, we have

$$\exp\{-\tfrac{1}{2} z^T z\} = \prod_{i=1}^{k} \exp\{-\tfrac{1}{2} z_i^2\}$$

Thus,

$$\int_{\mathbb{R}^k} \frac{1}{\sqrt{(2\pi)^k}} \exp\{-\tfrac{1}{2} z^T z\}\, dz = \prod_{i=1}^{k} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\{\tfrac{-1}{2} z_i^2\}\, dz_i$$

$$= \frac{1}{\sqrt{(2\pi)^k}} \prod_{i=1}^{k} \sqrt{2\pi} = 1. \quad \square$$

2. Let $A, B$ be $n$-by-$n$ matrices and $x$ be a $n$-by-1 vector.

(a) Show that $\frac{\partial}{\partial A}\mathrm{trace}(AB) = B^T$.

(b) Show that $x^T A x = \mathrm{trace}(x x^T A)$.

(C) ~~3.~~ Derive the maximum likelihood estimators for a multivariate Gaussian.

(a) we have $\mathrm{trace}(AB) = \sum_{i=1}^{n} [AB]_{ii} = \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}B_{ji}$

Thus,

$$\frac{\partial}{\partial A_{ij}}\mathrm{trace}(AB) = B_{ji} = B^T. \square$$

(b) left side:

$$x^T A x = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i A_{ij} x_j$$

right side:

$$\mathrm{trace}(xx^TA) = \sum_{i=1}^{n}[xx^TA]_{ii}$$

where $[xx^TA]_{ii} = \sum_{k=1}^{n}[xx^T]_{ik}A_{ki} = \sum_{k=1}^{n} x_i x_k A_{ki}$,

Then we have

$$\mathrm{trace}(xx^TA) = \sum_{i=1}^{n}\sum_{k=1}^{n} x_i x_k A_{ki}$$

let $k = j$ and since $A_{ji}$ is scalar, we have

$$\sum_{i=1}^{n}\sum_{k=1}^{n} x_i x_k A_{ki} = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i A_{ij} x_j = x^T A x. \square$$

(c) Suppose $x_1, x_2, \dots, x_n$ with i.i.d $\sim N(\mu, \Sigma)$

the likelihood function is:

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

and the log-likelihood function is

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

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

derive to $\mu$:

$$\frac{\partial \ell}{\partial \mu}(\mu,\Sigma) = \frac{\partial}{\partial \mu}\left(\underline{-\frac{nk}{2}\ln(2\pi)} - \underline{\frac{n}{2}\ln|\Sigma|} - \frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu)\right)$$

$$= -\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu) = -\frac{1}{2}\sum_{i=1}^{n}(-2\Sigma^{-1}(x_i-\mu)) = \sum_{i=1}^{n}\Sigma^{-1}(x_i-\mu) := 0 \text{ (to find } \hat{\mu})$$

$\Rightarrow \sum_{i=1}^{n}\Sigma^{-1}(x_i-\mu) = 0 \Rightarrow \sum_{i=1}^{n}(x_i-\mu) = 0 \Rightarrow n\mu = \sum_{i=1}^{n}x_i$. Thus, $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}x_i$. $\square$

derive to $\Sigma$ :

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

$$\sum_{i=1}^{n}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu) = \mathrm{trace}(\Sigma^{-1}S)$$

we can simplify log-likelihood function to

$$l(\Sigma) = -\frac{n}{2}\ln|\Sigma| - \frac{1}{2}\mathrm{trace}(\Sigma^{-1}S) + C, \quad C \in \mathbb{R}$$

we have $\frac{\partial \ln|\Sigma|}{\partial \Sigma} = \Sigma^{-1}$ , $\frac{\partial \mathrm{trace}(\Sigma^{-1}S)}{\partial \Sigma} = -\Sigma^{-1}S\Sigma^{-1}$

Thus,

$$\frac{\partial l}{\partial \Sigma} = -\frac{n}{2}\Sigma^{-1} + \frac{1}{2}\Sigma^{-1}S\Sigma^{-1} := 0 \quad (\text{to find } \hat{\Sigma})$$

Then

$$-\frac{n}{2}\Sigma^{-1} + \frac{1}{2}\Sigma^{-1}S\Sigma^{-1} = 0 \Rightarrow \Sigma^{-1}S\Sigma^{-1} = n\Sigma^{-1}$$

$$\Rightarrow \Sigma\Sigma^{-1}S\Sigma^{-1}\Sigma = \Sigma n\Sigma^{-1}\Sigma$$

$$\Rightarrow S = n\Sigma$$

Thus,

$$S = n\Sigma \Rightarrow \hat{\Sigma} = \frac{1}{n}S = \frac{1}{n}\sum_{i=1}^{n}(x_i-\hat{\mu})(x_i-\hat{\mu})^T \text{ ..}$$

3. 如果 $\Sigma$ 是 positive semidefinite matrix 而不是 positive definite matrix, multrvariate Gaussian 會變什麼樣子, 對模型又會有什麼影響?