111652015 曹晉嘉

1. Given

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

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

Since Σ is positive definite matrix, there exists unique lower-triangular L with positive diagonal s.t. $\Sigma = LL^T$

Let $y := L^{-1}(x-\mu) \Rightarrow x = \mu + Ly$

Then $(x-\mu)^T\Sigma^{-1}(x-\mu) = (Ly)^T(LL^T)^{-1}(Ly) = y^Ty = \|y\|^2$

The Jacobian of the transformation is $dx = |\det L|\,dy = \sqrt{\Sigma}\,dy$

$$\int_{\mathbb{R}^k} f(x)\,dx = \frac{1}{\sqrt{(2\pi)^k|\Sigma|}} \int_{\mathbb{R}^k} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$

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

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

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

Since $\|y\|^2 = y_1^2 + y_2^2 \cdots + y_k^2$, $e^{-\frac{1}{2}\|y\|^2} = \prod_{i=1}^{k} e^{-\frac{1}{2}y_i^2}$

$$\int_{\mathbb{R}^k} f(x)\,dx = \frac{1}{(2\pi)^{\frac{k}{2}}} \int_{\mathbb{R}^k} e^{-\frac{1}{2}\|y\|^2}\,dy = (2\pi)^{-\frac{k}{2}} \prod_{i=1}^{k} \int_{-\infty}^{\infty} e^{-\frac{1}{2}y_i^2}\,dy$$

$$= \prod_{i=1}^{k} \underbrace{\int_{-\infty}^{\infty} \frac{e^{-\frac{1}{2}y_i^2}}{\sqrt{2\pi}}\,dy}_{=1} = 1 \quad \#$$

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 $x^T A x = \mathrm{trace}(x x^T A)$.
   (b) Derive the maximum likelihood estimators for a multivariate Gaussian.

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

$\frac{\partial}{\partial A_{pq}} \mathrm{trace}(AB) = \frac{\partial}{\partial A_{pq}} \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij} B_{ji} = B_{qp}$

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

(b) $\mathrm{trace}(x x^T A) = \sum_{j=1}^{n}\sum_{i=1}^{n} (x x^T)_{ji} A_{ij} = \sum_{j=1}^{n}\sum_{i=1}^{n} x_j x_i A_{ij} = \sum_{j=1}^{n}\sum_{i=1}^{n} x_i A_{ij} x_j$

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

$\Rightarrow \mathrm{trace}(x x^T A) = x^T A x$

(C) Suppose we have iid. sample $y_1\ y_2 \ldots y_n \sim N_k(\mu, \Sigma)$

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

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

The likelihood function of the sample $L(\mu, \Sigma) = \prod_{i=1}^{n} f(y_i|\mu, \Sigma)$

Let $l(\mu, \Sigma) := \ln(L(\mu, \Sigma)) = -\frac{nk}{2}\ln(2\pi) - \frac{n}{2}\ln|\Sigma| - \frac{1}{2}\sum_{i=1}^{n}(y_i-\mu)^T \Sigma^{-1}(y_i-\mu)$

Let $\frac{\partial l}{\partial \mu} = 0 \Rightarrow \frac{\partial l}{\partial \mu} = \Sigma^{-1}\sum_{i=1}^{n}(y_i - \mu) = 0 \Rightarrow \hat{\mu} = \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$

$l(\hat{\mu}, \Sigma) = -\frac{n}{2}\ln|\Sigma| - \frac{1}{2}\underline{\sum_{i=1}^{n}(y_i - \bar{y})^T \Sigma^{-1}(y_i - \bar{y})}$ (Omitting the constant term does not affect the solution)

$$\sum_{i=1}^{n}(y_i-\bar{y})^T\Sigma^{-1}(y_i-\bar{y}) = \mathrm{trace}\left(\Sigma^{-1}\sum_{i=1}^{n}(y_i-\bar{y})(y_i-\bar{y})^T\right)$$

Let $\bar{\Sigma} := \frac{1}{n}\sum_{i=1}^{n}(y_i-y)(y_i-y)^T$

Rewrite $\ell(\hat{u},\Sigma)$: $\ell(\hat{u},\Sigma) = -\frac{n}{2}\ln|\Sigma| - \frac{n}{2}\,tr(\Sigma^{-1}\bar{\Sigma})$

Using two matrix calculus Identities

$$d\ln|\Sigma| = tr(\Sigma^{-1}d\Sigma) \quad , \quad d\,tr(\Sigma^{-1}\bar{\Sigma}) = -tr(\Sigma^{-1}\bar{\Sigma}\Sigma^{-1}d\Sigma)$$

We have $d\ell = -\frac{n}{2}\mathrm{trace}(\Sigma^{-1}d\Sigma) + \frac{n}{2}\mathrm{trace}(\Sigma^{-1}\bar{\Sigma}\Sigma^{-1}d\Sigma)$

$= \langle G, d\Sigma\rangle$ where $G = -\frac{n}{2}\Sigma^{-1} + \frac{n}{2}\Sigma^{-1}\bar{\Sigma}\Sigma^{-1}$

The first-order condition $G=0 \Rightarrow -\frac{n}{2}\Sigma^{-1} + \frac{n}{2}\Sigma^{-1}\bar{\Sigma}\Sigma^{-1} = 0$

$$\Rightarrow \Sigma^{-1}\bar{\Sigma}\Sigma^{-1} = \Sigma^{-1}$$

$$\Rightarrow \bar{\Sigma} = \Sigma$$

$$\Rightarrow \hat{\Sigma} = \bar{\Sigma}$$

Final Result : $\hat{u} = \frac{1}{n}\sum_{i=1}^{n}y_i$, $\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{u})(y_i-\hat{u})^T$

3 Unanswered Questions :

上課有說到啟動函數用 $\sigma(x) = \frac{1}{1+e^x}$ 的原因. 那若用其他函數做為啟動函數會有什麼影響？