Machine Learning

313704071 陳安定 Assignment 5

1. Given

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

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

Proof:

Σ can be transformed into $\Sigma=\mathrm{Q}\Lambda\mathrm{Q}^{\mathrm{T}}$, where Q is k-by-k orthogonal matrix, Λ is k-by-k diagonal matrix. Let $y=x-\mu$, then $(x-\mu)^T\Sigma^{-1}(x-\mu)=y^T\mathrm{Q}\Lambda^{-1}\mathrm{Q}^{\mathrm{T}}y$. And let $u=\mathrm{Q}^{\mathrm{T}}y$, so $(x-\mu)^T\Sigma^{-1}(x-\mu)=u^T\Lambda^{-1}u$.

Continuously, let $z=\Lambda^{-\frac{1}{2}}u$, $u=\Lambda^{\frac{1}{2}}z$, then $u^T\Lambda^{-1}u=\|z\|^2$.

Since $z=\Lambda^{-\frac{1}{2}}\mathrm{Q}^{\mathrm{T}}(x-\mu)$, $\frac{dz}{dx}=\left|\Lambda^{-\frac{1}{2}}\mathrm{Q}^{\mathrm{T}}\right|$, $dx=\left|\Lambda^{\frac{1}{2}}\mathrm{Q}^{\mathrm{T}}\right|dz=\sqrt{|\Sigma|}\,dz$

Then,

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

$$\int f(x)dx=\int\frac{1}{\sqrt{(2\pi)^k}}e^{-\frac{1}{2}\|z\|^2}dz=1$$

Notice that $\frac{1}{\sqrt{(2\pi)^k}}e^{-\frac{1}{2}\|z\|^2}$ is the density of k independent $N(0,1)$variables, hence its integral is equal to 1.

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

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

Assume A, B is 2-by-2 matrices, $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, $B = \begin{bmatrix} p & q \\ r & s \end{bmatrix}$

$AB = \begin{bmatrix} ap + br & \# \\ \# & cq + ds \end{bmatrix}$, $trace(AB) = \boldsymbol{a}p + \boldsymbol{b}r + \boldsymbol{c}q + \boldsymbol{d}s$.

and derived by every element of A (a, b, c, d) respectively, then it will be

$\frac{\partial}{\partial A} trace(AB) = \begin{bmatrix} p & r \\ q & s \end{bmatrix}$, which is B^T.

$trace(AB)_{n\times n} = \sum_i (AB)_{ii} = \sum_i \sum_j A_{ij}B_{ji}$, and derive it by A_{kl},

The result will be $B_{lk} = B^{-1}$.

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

$x^T Ax$ is **scalar**, so $x^T Ax = trace(x^T Ax)$,

by $trace(AB) = trace(BA)$,

$trace(x^T Ax) = trace(Axx^T) = trace(Axx^T)$.

(c) Derive the maximum likelihood estimators for a multivariate Gaussian.

$$L(\mu, \Sigma) = \prod_i^n f(x) = \left(\frac{1}{(2\pi)^{k/2}|\Sigma|^{1/2}}\right)^N e^{-\frac{1}{2}\sum(x-\mu)^T\Sigma^{-1}(x-\mu)},$$

$$\ln L(\mu, \Sigma) = -\frac{\text{nk}}{2}\ln(2\pi) - \frac{n}{2}\ln|\Sigma| - \frac{1}{2}\sum(x_i - \mu)^T\Sigma^{-1}(x_i - \mu),$$

$$\frac{\partial \ln L(\mu,\Sigma)}{\partial \mu} = -\frac{1}{2}\Sigma^{-1}\left(-2\left(\sum x_i - n\mu\right)\right) = 0,\ \boxed{\hat{\boldsymbol{\mu}} = \frac{1}{n}\sum x_i.}$$

$\frac{\partial \ln L(\mu,\Sigma)}{\partial \Sigma}$, first compute $\frac{d}{d\Sigma}\ln|\Sigma|$, $\frac{d}{d\Sigma}\ln|\Sigma| = (\Sigma^{-1})^T = \Sigma^{-1}$,

$$\frac{\partial \ln L(\mu,\Sigma)}{\partial \Sigma} = -\frac{n}{2}\Sigma^{-1} + \frac{1}{2}\Sigma^{-2}\sum(x_i - \mu)^T(x_i - \mu) = 0,$$

$$-\frac{n}{2} + \frac{1}{2}\Sigma^{-1}\sum(x_i - \mu)^T(x_i - \mu) = 0,\ \boxed{\hat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum(x_i - \mu)^T(x_i - \mu)}.$$

3. Unanswered Questions

上一個作業：台灣氣候資料分類問題，很難用一條線去區分有無資料，因為台灣是有許多離島且外圍是一個番薯形狀，那我的問題是，假設圖畫出來，有資料的地方僅僅是一個圓，包圍的部分就是無資料，這樣又如何去分類呢? 問題會變得比較簡單嗎?

還是最大的問題是在離島分布不規則的狀況?