26 Aug

26 August 2025 10:03

$X \sim N_p(\mu, \Sigma)$

Random sample $\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n$ are all iid random variable ($\underset{\sim}{X}_1 \to p \times 1$)

$\underset{\sim}{X}_i \sim N_p(\mu, \Sigma) \quad 1 \le i \le n$

$$f_{X_i}(x_i ; \mu, \Sigma) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2} (x_i - \mu)^T \Sigma^{-1} (x_i - \mu) \right)$$

Joint pdf of $\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n$ = product of marginals

Likelihood function $L(x_1, \ldots, x_n ; \mu, \Sigma) = \prod_{i=1}^{n} f_{X_i}(x_i ; \mu, \Sigma$

[x_i is realization of $\underset{\sim}{X}_i$] ($x_i \to p \times 1$, $\underset{\sim}{X}_i \to p \times 1$)

$$= \frac{1}{(2\pi)^{np/2} |\Sigma|^{n/2}} \exp\left(-\frac{1}{2} \sum_{i=1}^{n} (x_i - \mu)^T \Sigma^{-1} (x\right.$$

What is maximum likelyhood estimator (MLE) of μ and Σ.

$\hookrightarrow$ $\hat{\mu}$ and $\hat{\Sigma}$ such that $L(\quad)$ is maximized.

$x_1, \ldots, x_n$ are given

$\mu_{p \times 1}$ is unknown $\longrightarrow p$ parameters

$\Sigma_{p \times p}$ (symetric) is unknown $\longrightarrow \frac{p(p+1)}{2}$ parameters

} this grows with p

classical calculus: evaluate derivative wrt each parameter

Trace

$Tr(A+B) = Tr(A) + Tr(B)$

$Tr(KA) = K\,Tr(A)$ (K: real constant)

$Tr(AB) = Tr(BA)$

$\hookrightarrow Tr(X^T A X) = Tr(A X X^T)$

if A is symetric then A is diagonalizable: $A = P \Lambda P^T$

$$(PP^T = I = P^T P)$$

$$\hookrightarrow \operatorname{Tr}(A) = \operatorname{Tr}(P \Lambda P^T)$$

$$= \operatorname{Tr}(\Lambda) = \boxed{\text{sum of eigen values of } A}$$

MLE:

$$\max_{\mu, \Sigma} L(x_1, \dots, x_n; \mu, \Sigma) \equiv L(\mu, \Sigma)$$

$$= \frac{1}{(2\pi)^{np/2} |\Sigma|^{n/2}} \exp\left(-\frac{1}{2} \sum_{i=1}^{n} (x_i - \mu)^T \Sigma^{-1} (x_i - \mu)\right)$$

$$\sum_{i=1}^{n} (x_i - \mu)^T \Sigma^{-1} (x_i - \mu) = \operatorname{Trace}\left(\sum_{i=1}^{n} (x_i - \mu)^T \Sigma^{-1} (x_i - \mu)\right)$$

[this is 1x1 matrix ≡ trace of matrix]

$$= \sum_{i=1}^{n} \operatorname{Tr}\left((x_i - \mu)^T \Sigma^{-1} (x_i - \mu)\right)$$

$$= \sum_{i=1}^{n} \operatorname{Tr}\left(\underset{p \times p}{\Sigma^{-1}} \underset{p \times 1}{(x_i - \mu)} \underset{1 \times p}{(x_i - \mu)^T}\right)$$

$$= \operatorname{Tr}\left(\sum_{i=1}^{n} \Sigma^{-1} (x_i - \mu)(x_i - \mu)^T\right)$$

$$= \operatorname{Tr}\left(\Sigma^{-1} \sum_{i=1}^{n} (x_i - \mu)(x_i - \mu)^T\right)$$

from notes:

$$(n-1) S = \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})^T$$

$$= \sum_{i=1}^{n} (x_i - \mu + \mu - \bar{x})(x_i - \mu + \mu - \bar{x})^T$$

$$= \sum_{i=1}^{n} (x_i - \mu)(x_i - \mu)^T - n(\bar{x} - \mu)(\bar{x} - \mu)^T$$

$$\therefore \sum_{i=1}^{n} (x_i - \mu)(x_i - \mu)^T = (n-1) S + n(\bar{x} - \mu)(\bar{x} - \mu)^T$$

$$= n S_n + n(\bar{x} - \mu)(\bar{x} - \mu)^T$$

where $S_n = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})^T$

$$= \frac{n-1}{n} \times \frac{1}{n-1} \times \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})^T$$

$$= \frac{n-1}{n} S$$

so now

$$L(\mu,\Sigma) = \frac{1}{(2\pi)^{np/2} |\Sigma|^{n/2}} \exp\left(-\frac{1}{2} Tr\left(\Sigma^{-1}\left(n S_n + n(\bar{x}-\mu)(\bar{x}-\mu)^T\right)\right)\right)$$

$$= \frac{1}{(2\pi)^{np/2} |\Sigma|^{n/2}} \exp\left\{-\frac{n}{2} Tr(\Sigma^{-1} S_n) - \frac{n}{2} tr\left(\Sigma^{-1}(\bar{x}-\mu)(\bar{x}-\mu)^T\right)\right\}$$

$$= \frac{1}{(2\pi)^{np/2} |\Sigma|^{n/2}} \exp\left\{-\frac{n}{2} Tr(\Sigma^{-1} S_n) - \frac{n}{2} Tr\left((\bar{x}-\mu)^T \Sigma^{-1} (\bar{x}-\mu)\right)\right\}$$

Aim is to maximizing this

wrt μ : only term : $(\bar{x}-\mu)^T \Sigma^{-1} (\bar{x}-\mu)$ ← positive definate

if we minimize this, L maximize

@ $\mu = \bar{x}$ this is zero (minimum)

$\therefore \boxed{\hat{\mu}_{MLE} = \bar{x}}$

now $$L = \frac{1}{(2\pi)^{np/2} |\Sigma|^{n/2}} \exp\left\{-\frac{n}{2} Tr(\Sigma^{-1} S_n)\right\}$$

conider $Tr(\Sigma^{-1} S_n)$

$= Tr(\underbrace{\Sigma^{-1} S_n^{1/2}}\, \underbrace{S_n^{1/2}})$

$= Tr(\underbrace{S_n^{1/2} \Sigma^{-1} S_n^{1/2}})$

(product of sym pd
$\therefore$ this is also pd)

= sum of eigen values of $S_n^{1/2} \Sigma^{-1} S_n^{1/2}$

$$= \sum_{i=1}^{p} \eta_i$$

← i^{th} eigen value

both Σ and S_n are positive definate and symetr

$S_n = P \Lambda P^T$ $\quad (P^T P = P P^T = I)$

define $\Lambda^{1/2} = \begin{pmatrix} \sqrt{\lambda_1} & & 0 \\ & \ddots & \\ 0 & & \sqrt{\lambda_p} \end{pmatrix}$

$S_n^{1/2} = P \Lambda^{1/2} P^T$ $\quad \lambda_i > 0$

$S_n^{1/2} S_n^{1/2} = S_n$

$\vert S_n^{1/2} \Sigma^{-1} S_n^{1/2} \vert \quad \prod^{n} \eta_i$

now consider $\left| S_n^{1/2} \Sigma^{-1} S_n^{1/2} \right| = \prod_{i=1}^{p} \eta_i$

$$= |\Sigma^{-1}| \times |S_n|$$

$$\Rightarrow |\Sigma| = \frac{|S_n|}{\prod_{i=1}^{p} \eta_i}$$

substitute in L:

$$L = \frac{1}{(2\pi)^{np/2}} \times \left(\frac{\prod \eta_i}{|S_n|} \right)^{n/2} \times \exp\left(-\frac{n}{2} \sum_{i=1}^{p} \eta_i \right)$$

$$= \frac{1}{(2\pi)^{np/2} \times |S_n|^{n/2}} \times \prod_{i=1}^{p} \left\{ \eta_i^{n/2} \exp\left(-\frac{n \eta_i}{2} \right) \right\}$$