Tutorial 2

CS5487 Lecture Notes (2022A)
Prof. Antoni B. Chan
Dept of Computer Science
City University of Hong Kong

P2.6 - MLE for m.v. Gaussian.

$D = \{x_1, \dots, x_N\}$

data log-likelihood: $\log p(D) = \sum_{i=1}^{N} \log p(x_i)$

$$= \sum_{i=1}^{N} \log \left[\frac{1}{(2\pi)^{d/2} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}\|x_i-\mu\|^2_\Sigma} \right]$$

$$= \sum_{i=1}^{N} \left[-\frac{d}{2}\log 2\pi - \frac{1}{2}\log|\Sigma| - \frac{1}{2}\|x_i-\mu\|^2_\Sigma \right]$$

$$\boxed{\log p(D) = -\frac{Nd}{2}\log 2\pi - \frac{N}{2}\log|\Sigma| - \frac{1}{2}\sum_{i=1}^{N}\|x_i-\mu\|^2_\Sigma}$$

a) MLE for μ

$$\frac{\partial}{\partial\mu}\log p(D) = \frac{\partial}{\partial\mu}\left(-\frac{1}{2}\sum_{i=1}^{N}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu)\right)$$

$$= \frac{\partial}{\partial\mu}\left(-\frac{1}{2}\sum_{i=1}^{N}\left[\underbrace{x_i^T\Sigma^{-1}x_i}_{\times} + \mu^T\Sigma^{-1}\mu - 2x_i^T\Sigma^{-1}\mu\right]\right)$$

$$= -\frac{1}{2}\frac{\partial}{\partial\mu}\left(N\mu^T\Sigma^{-1}\mu - 2\left(\sum_{i=1}^{N}x_i^T\right)\Sigma^{-1}\mu\right)$$

① $\frac{\partial}{\partial x} a^T x = a$ $\qquad$ $\frac{\partial}{\partial x} ax = a$

② $\frac{\partial}{\partial x} x^T A x = Ax + A^T x = 2Ax$ if A symmetric.

$\frac{\partial}{\partial x} ax^2 = 2ax$

$$\frac{\partial}{\partial\mu}\log p(D) = \cancel{-\frac{1}{2}}\left(N\cdot\cancel{2}\,\Sigma^{-1}\mu - \cancel{2}\,\Sigma^{-1}\left(\sum_{i=1}^{N}x_i\right)\right) = 0$$

$$N\Sigma^{-1}\mu = \Sigma^{-1}\sum_{i=1}^{N}x_i$$

pre-multiply by Σ

$$N\mu = \sum_{i=1}^{N}x_i$$

$$\Rightarrow \boxed{\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N}x_i}$$

b) MLE for Σ

$$\frac{\partial}{\partial\Sigma}\log p(D) = \frac{\partial}{\partial\Sigma}\left[-\frac{N}{2}\log|\Sigma| - \frac{1}{2}\sum_{i=1}^{N}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu)\right]$$

$a^T b = \text{trace}(a^T b) = \text{trace}(b a^T)$

$$= \frac{\partial}{\partial\Sigma}\left[-\frac{N}{2}\log|\Sigma| - \frac{1}{2}\sum_{i=1}^{N}\text{tr}\left(\Sigma^{-1}(x_i-\mu)(x_i-\mu)^T\right)\right]$$

$$= \frac{\partial}{\partial\Sigma}\left[-\frac{N}{2}\log|\Sigma| - \frac{1}{2}\text{tr}\left(\Sigma^{-1}\underbrace{\sum_{i=1}^{N}(x_i-\mu)(x_i-\mu)^T}_{A}\right)\right]$$

$\frac{\partial}{\partial X}\log|X| = X^{-T}$ $\quad\longleftrightarrow\quad$ $\frac{\partial}{\partial x}\log x = \frac{1}{x} = x^{-1}$

$\frac{\partial}{\partial X}\text{tr}(X^{-1}A) = -(X^{-T}A^TX^{-T})$ $\quad\longleftrightarrow\quad$ $\frac{\partial}{\partial x}\frac{a}{x} = \frac{-a}{x^2}$

$$= -\frac{N}{2}\Sigma^{-T} - \frac{1}{2}\left(-\Sigma^{-T}\left[\sum_{i=1}^{N}(x_i-\mu)(x_i-\mu)^T\right]\Sigma^{-T}\right) = 0$$

pre-post multiply by Σ

$$= -\frac{N}{2}\Sigma + \frac{1}{2}\sum_{i=1}^{N}(x_i-\mu)(x_i-\mu)^T = 0$$

$$\Rightarrow \boxed{\hat{\Sigma} = \frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)(x_i-\mu)^T}$$

PS 2-8 LS regression & MLE

function
$$f(x,\theta) = \underbrace{\phi(x)^T\theta} \qquad \theta \in \mathbb{R}^P, \ \phi(x) \in \mathbb{R}^P$$

observation model
$$y = \underbrace{f(x,\theta)} + \underbrace{\epsilon}, \qquad \epsilon \sim N(0, \sigma^2)$$

$$p(y \mid x, \theta) = N(y \mid \phi(x)^T\theta, \sigma^2)$$

$$x \to \begin{bmatrix} 1 \\ x \\ x^2 \\ \vdots \\ x^P \end{bmatrix} = \phi(x) \Rightarrow f(x)\ \phi(x)^T\theta = \begin{bmatrix} 1 \\ x \\ x^2 \\ \vdots \\ x^P \end{bmatrix}^T \begin{bmatrix} \theta_0 \\ \vdots \\ \theta_P \end{bmatrix}$$

$$= \theta_0 + \theta_1 x + \theta_2 x^2 + \dots \theta_P x^P$$

Given dataset: $D = \{(x_i, y_i)\}_{i=1}^N$

$$\log p(D) = \sum_{i=1}^N \log p(y_i \mid x_i, \theta)$$

$$= \sum_{i=1}^N \underbrace{-\frac{1}{2}\log 2\pi - \frac{1}{2}\log \sigma^2}_{\times} - \frac{1}{2\sigma^2}(y_i - \phi(x_i)^T\theta)^2$$

$$\propto \underbrace{-\frac{1}{2\sigma^2}} \sum_{i=1}^N (y_i - \phi(x_i)^T\theta)^2$$

negating the objective → max into min.

$$\arg\max_\theta \log p(D) = \arg\min_\theta \sum_{i=1}^N (\underbrace{y_i} - \underbrace{\phi(x_i)^T\theta}_{f(x_i)})^2$$

⇐ Least squares formulation

$$= \arg\min_\theta \begin{bmatrix} y_1 - \phi(x_1)^T\theta \\ \vdots \\ y_N - \phi(x_N)^T\theta \end{bmatrix}^T \underbrace{\begin{bmatrix} y_1 - \phi(x_1)^T\theta \\ \vdots \\ y_N - \phi(x_N)^T\theta \end{bmatrix}} = \arg\min_\theta \|y - \Phi^T\theta\|^2$$

$$\underbrace{\begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}}_{y} - \underbrace{\begin{bmatrix} \phi(x_1) \dots \phi(x_N) \end{bmatrix}}_{\Phi}^T \theta$$

Solve for θ

$$\frac{\partial}{\partial\theta}\|y - \Phi^T\theta\|^2 = \frac{\partial}{\partial\theta}(y - \Phi^T\theta)^T(y - \Phi^T\theta)$$

$$= \frac{\partial}{\partial\theta}\Big[\underbrace{y^Ty}_{\times} - \underbrace{2y^T\Phi^T\theta}_{a^Tx} + \underbrace{\theta^T\Phi\Phi^T\theta}_{x^TAx}\Big]$$

$$= -2\Phi y + 2\Phi\Phi^T\theta = 0$$

$$\Rightarrow \Phi\Phi^T\theta = \Phi y$$

premultiply by $(\Phi\Phi^T)^{-1}$

$$\Rightarrow \boxed{\hat{\theta} = (\Phi\Phi^T)^{-1}\Phi y}$$

$$\Phi = \underbrace{\begin{bmatrix} \phi(x_1) & \dots & \phi(x_N) \end{bmatrix}}_{N} \Big\} P$$

We assume $\Phi\Phi^T$ is invertible

$\Phi\Phi^T \in \mathbb{R}^{P \times P}$, ⇒ we want the columns of Φ to span $\mathbb{R}^P$