概率知识补充 — 高斯分布

Data: $X = (x_1, x_2 \cdots x_N)^T = \begin{pmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_N^T \end{pmatrix}_{N \times p}$

$x_i \in \mathbb{R}^p$

$x_i \overset{iid}{\sim} N(\mu, \Sigma) \quad \theta = (\mu, \Sigma)$

iid → 独立同分布

MLE: $\theta_{MLE} = \arg\max_{\theta} P(X|\theta)$

极大似然估计

简化 ⇒ 令 $p=1$（1维） $x \sim (\mu, \sigma^2) \quad \theta = (\mu, \sigma^2)$

一维: $P(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$

p维: $P(x) = \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right)$

一维为例:

① μ_{MLE}

$$\log P(X|\theta) = \log \prod_{i=1}^{N} P(x_i|\theta) = \sum_{i=1}^{N} \log P(x_i|\theta)$$

$$= \sum_{i=1}^{N} \log \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)$$

exp 与 log 抵消

$$= \sum_{i=1}^{N} \left[\underbrace{\log \frac{1}{\sqrt{2\pi}} + \log \frac{1}{\sigma}}_{与\mu无关} - \frac{(x_i-\mu)^2}{2\sigma^2} \right]$$

①

$$\mu_{MLE} = \arg\max_{\mu} \log P(X|\theta)$$

$$= \arg\max \sum^{N} \left(-\frac{(x_i-\mu)^2}{\cdots}\right)$$

$$= \underset{\mu}{\arg\min} \sum_{i=1}^{N} (x_i - \mu)^2$$

求导：

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

$$\Rightarrow \sum_{i=1}^{N} (x_i - \mu) = 0$$

$$\Rightarrow \boxed{\mu_{MLE} = \frac{1}{N} \sum_{i=1}^{N} x_i}$$

即μ的极大似然估计为样本均值.

② σ_{MLE}

$$\sigma^2_{MLE} = \underset{\sigma^2}{\arg\max}\ P(X|\theta)$$

①式中 log项与σ无关，消去.

$$= \underset{\sigma^2}{\arg\max} \sum_{i=1}^{N} \underbrace{\Big(\underset{= -\log\sigma}{\log \frac{1}{\sigma}} - \frac{(x_i - \mu)^2}{2\sigma^2} \Big)}_{L(\sigma)}$$

$$\frac{\partial L}{\partial \sigma} = \sum_{i=1}^{N} \left[-\frac{1}{\sigma} - \frac{1}{2} (x_i - \mu)^2 \cdot (-2) \sigma^{-3} \right] = 0$$

$$\Rightarrow \sum_{i=1}^{N} \left[-\frac{1}{\sigma} + (x_i - \mu)^2 \sigma^{-3} \right] = 0$$

$$\Rightarrow \sum_{i=1}^{N} \left(-\sigma^2 + (x_i - \mu)^2 \right) = 0$$

即：$-\sum_{i=1}^{N} \sigma^2 + \sum_{i=1}^{N} (x_i - \mu)^2 = 0$

$$\Rightarrow \boxed{\sigma^2_{MLE} = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu_{MLE})^2}$$

有偏/无偏估计：

①
$$E(\mu_{MLE})$$
$$= \frac{1}{N}\sum_{i=1}^{N} E(x_i) \rightarrow \text{独立同分布}$$
$$= \frac{1}{N}\sum_{i=1}^{N} \mu = \frac{1}{N}\cdot N \cdot \mu = \mu \quad \text{无偏估计}$$

②
$$\sigma^2_{MLE} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu_{MLE})^2$$
$$= \frac{1}{N}\sum_{i=1}^{N}(x_i^2 - 2x_i\mu_{MLE} + \mu_{MLE}^2)$$
$$= \frac{1}{N}\sum_{i=1}^{N}x_i^2 - \underbrace{\frac{1}{N}\sum_{i=1}^{N}2x_i\mu_{MLE}}_{2\mu_{MLE}\cdot\mu_{MLE}} + \underbrace{\frac{1}{N}\sum_{i=1}^{N}\mu_{MLE}^2}_{\mu_{MLE}^2}$$
$$= \frac{1}{N}\sum_{i=1}^{N}x_i^2 - \mu_{MLE}^2$$

$$E(\sigma^2_{MLE}) = E\left(\frac{1}{N}\sum_{i=1}^{N}x_i^2 - \mu_{MLE}^2\right) = E\left[\frac{1}{N}\sum_{i=1}^{N}x_i^2 - \mu^2 - (\mu_{MLE}^2 - \mu^2)\right]$$
$$= E\left(\frac{1}{N}\sum_{i=1}^{N}x_i^2 - \mu^2\right) - E(\mu_{MLE}^2 - \mu^2)$$

$$E\left(\frac{1}{N}\sum_{i=1}^{N}x_i^2 - \mu^2\right) \Rightarrow E\left[\frac{1}{N}\sum_{i=1}^{N}(x_i^2 - \mu^2)\right] \rightarrow \frac{1}{N}\sum_{i=1}^{N}E(x_i^2 - \mu^2) \rightarrow \frac{1}{N}\sum_{i=1}^{N}\left(E(x_i^2) - E(\mu^2)\right) \rightarrow \frac{1}{N}\sum_{i=1}^{N}\left(E(x_i^2) - \mu^2\right)$$

$$\mu^2 = E(x_i)^2$$

$$= (\text{方差}) = Var(x_i)$$

$$E(\mu_{MLE}^2 - \mu^2) \Rightarrow E(\mu_{MLE}^2) - E(\mu^2) \rightarrow E(\mu_{MLE}^2) - \mu^2$$

$$\mu^2 \rightarrow E(\mu_{MLE})^2$$

$$\rightarrow Var(\mu_{MLE}) \Rightarrow Var\left[\frac{1}{N}\sum_{i=1}^{N}x_i\right]$$
$$= \frac{1}{N^2}\sum_{i=1}^{N}Var(x_i)$$
$$= \frac{1}{N^2}\sum_{i=1}^{N}\sigma^2$$

$$\Downarrow$$

$$\sigma^2$$

$$\underbrace{\frac{1}{N}\sum_{i=1}^{N}\sigma^2}=\sigma^2 \qquad = \frac{\sigma^2}{N}$$

$$\downarrow$$

$$\sigma^2 - \frac{\sigma^2}{N} = \frac{N-1}{N}\sigma^2$$

$$\Rightarrow \sigma^2_{MLE} = \frac{1}{N}\sum_{i=1}^{N}(x_i-\mu_{MLE})^2 \rightarrow \text{有偏估计}$$

无偏：$\hat{\sigma}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i-\mu_{MLE})^2$