

6(d) MLE for nonparametric
density estimator $p(x|\theta)$
with Gaussian kernel

$$p(x|\theta) = \frac{1}{N} \sum_{i=1}^N \frac{1}{\sqrt{2\pi}\theta} \exp \left\{ -\frac{(x-x_i)^2}{2\theta^2} \right\}$$

Intuition

θ is MLE of σ^2 (variance) of x (data)!

観測値 $\{x_1, \dots, x_n\}$ について θ を推定する

$$\text{尤度 } L(\theta^2) = \log p(x|\theta)$$

最適な θ を求める

$$L(\theta^2) = \frac{1}{N} \log \left(\sum \dots \right)$$

$$= \frac{1}{N} \left\{ \log \frac{1}{\sqrt{2\pi}\theta} \exp \left(-\frac{(x-x_i)^2}{2\theta^2} \right) + \log \left(\sum \dots \right) \right\}$$

$$= \frac{1}{N} \log \left\{ \left(\frac{1}{\sqrt{2\pi}\theta} \right)^N \exp \left(-\frac{(x-x_i)^2}{2\theta^2} \right) \exp(\dots) \right\}$$

$$L(\theta^2) = \frac{1}{N} \log \left\{ \left(\frac{1}{\sqrt{2\pi}\theta} \right)^n \times \exp\left(-\frac{(x-x_1)^2}{2\theta^2}\right) \times \exp(\dots) \right\}$$

$$= \frac{1}{N} \log \left\{ \left(\frac{1}{\sqrt{2\pi}\theta} \right)^n \times \exp \left\{ \frac{(x-x_1)^2 + \dots + (x-x_n)^2}{2\theta^2} \right\} \right\}$$

$$= \frac{1}{N} \left\{ n \log \frac{1}{\sqrt{2\pi}} + n \log \frac{1}{\theta} + \log e^{\dots} \right\}$$

$$= \frac{1}{N} \left\{ \underbrace{-n \log \sqrt{2\pi}}_{\text{constant}} + \frac{n}{2} \log \frac{1}{\theta^2} + \left\{ \frac{(x-x_1)^2 + \dots + (x-x_n)^2}{2\theta^2} \right\} \right\}$$

$$\hat{\theta}^2 = \underset{\theta^2}{\operatorname{argmax}} L(\theta^2) = \underset{\theta^2}{\operatorname{argmax}} \left\{ \frac{n}{2} \log \frac{1}{\theta^2} + \left\{ \dots \right\} \right\}$$

$$\hat{L}(\theta^2) = \hat{L}(\varepsilon) \quad \varepsilon = \theta^2$$

$$\frac{\hat{L}(\varepsilon)}{d\varepsilon} = -\frac{n}{2} (\log \varepsilon)' + \left\{ (x-x_1)^2 + \dots \right\} (2\varepsilon)^{-1} \cdot (-1)$$

$$= -\frac{n}{2} \frac{1}{\varepsilon} + \left\{ \dots \right\} (2\varepsilon)^{-2} \times (2\varepsilon)'$$

$$= \frac{-n}{2\varepsilon} + \left\{ \dots \right\} \frac{2}{4\varepsilon^2} = \frac{-n}{2\theta^2} + \left\{ \dots \right\} \frac{1}{2\theta^4}$$

x に ついて

$$\frac{\partial L(x)}{\partial x} = \frac{((x-x_1)^2)' + \dots + ((x-x_n)^2)'}{2\theta^2}$$

$$= \frac{1}{\theta^2} \{ (x_1 + x_2 + \dots + x_n) - nx \}$$

$$\frac{\partial L(x)}{\partial x} = 0 \text{ となる } x = \frac{1}{n} \sum x_i \text{ となる (平均値)}$$

$$\text{よって } \frac{\partial \hat{L}(\theta^2)}{\partial \theta^2} = 0 \text{ となる } \theta^2 \text{ と}$$

$$\theta^2 = \mu = \frac{1}{n} \sum (x_i - \mu)^2 \text{ となる}$$

分散の分散

すなわち 正規分布の分散を推定するパラメータは

$$\mu = \mu$$

$$\theta^2 = \sigma^2 \text{ のときである。}$$

よって $\frac{1}{n}$ として割り出す

分散は $\frac{1}{n-1}$ として割り出す。

$$\mu_{ML} = \bar{x} = \frac{1}{N} \sum_{n=1}^N x_n$$

$$\sigma_{ML}^2 = \theta^2 = \frac{1}{N} \sum_{n=1}^N (x_n - \mu)^2$$

最大尤度で分散を求めると導出された分布の分散は
過小に評価されてしまう。

この現象を バイアス とする

多項式でモデルを fitting する際に 過学習を防ぐには
関連

ガウス分布の期待値

$$\mathbb{E}[\mu_{ML}] = \mathbb{E}\left[\frac{1}{N} \sum x_i\right]$$

$$= \frac{1}{N} \sum \mathbb{E}[x_i]$$

$$= \frac{1}{N} \cdot N \cdot \mu = \mu$$

Sample variance

Unbiased variance

サンプル分散は 標本から計算した分散で、母集団に比べて
標本数が少ない場合は
サンプル分散が母分散よりも小さくなる



サンプル分散が母分散に等しくなるよう補正したものを
不偏分散という。

Sample Variance $\{x_1, \dots, x_n\}$ 平均は $\bar{x}$ とする

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

Unbiased Variance

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

$$\begin{aligned}
s^2 &\equiv \frac{1}{n} \sum (x_i - \bar{x})^2 \quad \leftarrow \text{Sample Variance} \\
&= \frac{1}{n} \sum \left\{ (x_i - \mu) - (\bar{x} - \mu) \right\}^2 \\
&= \frac{1}{n} \sum \left\{ (x_i - \mu)^2 + (\bar{x} - \mu)^2 - 2(x_i - \mu)(\bar{x} - \mu) \right\} \\
&= \frac{1}{n} \sum (x_i - \mu)^2 - \frac{2}{n} (\bar{x} - \mu) \sum (x_i - \mu) + \frac{(\bar{x} - \mu)^2}{n} \sum 1 \\
&= \frac{1}{n} \sum (x_i - \mu)^2 - \frac{2}{n} (\bar{x} - \mu) \{ x_1 + x_2 + \dots + x_n - n\mu \} + \frac{(\bar{x} - \mu)^2}{n} \times n \\
&= \frac{1}{n} \sum (x_i - \mu)^2 - \frac{2}{n} (\bar{x} - \mu) n (\bar{x} - \mu) + (\bar{x} - \mu)^2 \\
&= \frac{1}{n} \sum (x_i - \mu)^2 - (\bar{x} - \mu)^2
\end{aligned}$$

1) の期待値は以下の通り、母分散に比べて $E[(\bar{x} - \mu)^2]$ だけ小さい

$$E[s^2] = \frac{1}{n} \sum \underbrace{E[(x_i - \mu)^2]}_{\sigma^2} - E[(\bar{x} - \mu)^2]$$

$$\begin{aligned}
&= \frac{1}{n} \sum \sigma^2 - E[(\bar{x} - \mu)^2] \\
&= \sigma^2 - E[(\bar{x} - \mu)^2]
\end{aligned}$$

標本分散の期待値 = 母分散 - $\overline{E}[(\bar{x} - \mu)^2]$

これを評価する

$$\overline{E}[(\bar{x} - \mu)^2] = \overline{E}\left[\left\{\frac{1}{n}(x_1 + \dots + x_n - n\mu)\right\}^2\right]$$

$$= \overline{E}\left[\frac{1}{n^2}\{(x_1 - \mu) + (x_2 - \mu) + \dots\}^2\right]$$

$$= \frac{1}{n^2} \sum \overline{E}[(x_i - \mu)^2]$$

$$= \frac{1}{n^2} \sum \sigma^2$$

$$= \frac{1}{n} \sigma^2$$

$$\overline{E}[\sigma^2] = \sigma^2 - \frac{1}{n} \sigma^2 = \frac{n-1}{n} \sigma^2$$

$$\sigma^2 = \frac{n}{n-1} \overline{[s^2]}$$

$$= \frac{n}{n-1} \left\{ \frac{1}{n} \sum (x_i - \bar{x})^2 \right\}$$

$$= \frac{1}{n-1} \sum (x_i - \bar{x})^2$$

↑
標本への計算した分散は n で割り込む。

よって $n-1$ で割り込む分散と同じになる。

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$$

→ Unbiased Variance

と呼ぶ。