

1 Basic concept

1.1 Bias-variance decomposition

Define $e(x) = f(x) - \hat{f}(x)$ and $p(e)$ as its pdf.

$$\text{MSE} = E(e^2) = \int e^2 p de = (\int e p de)^2 + (\int e^2 p de - (\int e p de)^2) = E(e)^2 + \text{Var}(e) \quad (1)$$

1.2 A simple example

(A) Y follows binomial distribution $B(n, \pi_h)$. $E(Y) = n\pi_h$, $\text{Var}(Y) = n\pi_h(1 - \pi_h)$. One can estimate $f(0)$ by

$$f(0)_{\text{estimate}} = \frac{Y}{nh} \quad (2)$$

(B)

$$\begin{aligned} \pi_h &\approx hf(0) + \frac{f''(0)}{2} \int x^2 dx \\ &= hf(0) + \frac{f''(0)h^3}{24} \end{aligned} \quad (3)$$

Choosing h so that both $h \ll 1$ and $\pi_h \ll 1$ are true, we have

$$\begin{aligned} \text{MSE}(0) &= (E(\hat{f}(0)) - f(0))^2 + \text{Var}(\hat{f}(0)) \\ &\approx (\frac{\pi_h}{h} - f(0))^2 + \frac{\pi_h}{nh^2} \\ &= (\frac{f''(0)}{24})^2 h^4 + \frac{1}{nh} (f(0) + \frac{f''(0)h^2}{24}) \\ &\approx (\frac{f''(0)}{24})^2 h^4 + \frac{1}{nh} f(0) \end{aligned} \quad (4)$$

(C)

$$\frac{\partial \text{MSE}(0)}{\partial h} = 4Ah^3 - \frac{f(0)}{nh^2} \quad (5)$$

In order to minimize the MSE, $h = (\frac{f(0)}{4An})^{\frac{1}{5}}$

2 Curve Fitting by linear smoothing

(A)

$$y_{\text{estimate}} = \beta_{\text{estimate}} x$$

$$\begin{aligned}
&= (\vec{x}^T \vec{x})^{-1} \vec{x}^T \vec{y} x \\
&= \frac{\sum_i x_i y_i}{\sum_i x_i^2} x \\
&= \sum_i \frac{x_i x}{\vec{x}^2} y_i
\end{aligned} \tag{6}$$

So we have $\omega_i = \frac{x_i x}{\vec{x}^2}$

(B) See "linearsmoothing.R" and "weight.R"

3 Cross validation

(A) See "predictionerror.R"

(B) See "testmodel.R"

(C)

4 Local polynomial regression

(A) Define matrix R where $R_{ij}(\vec{x}) = (x_i - x)^{j-1}$. Then $g(x_i, a) = (R\vec{a})_i$. The cost function can be written as

$$\sum_{i=1}^n \omega_i (y_i - (R\vec{a})_i)^2 \tag{7}$$

Optimizing this with \vec{a} is equivalent to linear regression on $f(\vec{x}) = \vec{x}^T \vec{a}$ given observed data $X = R$ and \vec{y} and weighted cost function. The answer is

$$\vec{a} = (R^T \Omega R)^{-1} R^T \Omega \vec{y} \tag{8}$$

$$\Omega = \text{diag}(\omega_1, \omega_2, \dots) \tag{9}$$

The estimate of $f(x)$ is just $g(x, a) = a_0$.

(B) Call the old weight function ω and the new ones γ

$$\begin{aligned}
(R^T \Omega R)^{-1} &= \left(\begin{array}{cc} \sum_i \omega_i & \sum_i R_{i2} \omega_i \\ \sum_i R_{i2} \omega_i & \sum_i R_{i2}^2 \omega_i \end{array} \right)^{-1} \\
&= \frac{1}{(\sum_i \omega_i)(\sum_i R_{i2}^2 \omega_i) - (\sum_i R_{i2} \omega_i)^2} \begin{pmatrix} \sum_i R_{i2}^2 \omega_i & -\sum_i R_{i2} \omega_i \\ -\sum_i R_{i2} \omega_i & \sum_i \omega_i \end{pmatrix} \\
R^T \Omega \vec{y} &= \begin{pmatrix} \sum_i \omega_i y_i \\ \sum_i R_{i2} \omega_i y_i \end{pmatrix}
\end{aligned} \tag{11}$$

$$\begin{aligned}
f(x) = a_0 &= \frac{(\sum_i R_{i2}^2 \omega_i)(\sum_i \omega_i y_i) - (\sum_i R_{i2} \omega_i)(\sum_i R_{i2} \omega_i y_i)}{(\sum_i \omega_i)(\sum_i R_{i2}^2 \omega_i) - (\sum_i R_{i2} \omega_i)^2} \\
&= \sum_i \frac{((\sum_j R_{j2}^2 \omega_j) - (\sum_j R_{j2} \omega_j) R_{i2}) \omega_i}{\sum_k ((\sum_j R_{j2}^2 \omega_j) - (\sum_j R_{j2} \omega_j) R_{k2}) \omega_k} y_i \quad (12)
\end{aligned}$$

Define weight function $\gamma(x_i, x) = ((\sum_j R_{j2}^2 \omega_j) - (\sum_j R_{j2} \omega_j) R_{i2}) \omega_i$, the above can be written as a weighted sum of y_i . Further evaluate γ :

$$\begin{aligned}
\gamma(x_i, x) &= ((\sum_j R_{j2}^2 \omega_j) - (\sum_j R_{j2} \omega_j) R_{i2}) \omega_i \\
&= \frac{1}{h^2} ((\sum_j (x_j - x)^2 K_j) - (\sum_j (x_j - x) K_j)(x_i - x)) K_i \\
&= \frac{1}{h^2} (s_2 - s_1(x_i - x)) K_i \quad (13)
\end{aligned}$$

Here we used K_i to denote $K(\frac{x-x_i}{h})$. Since the function will be normalized so we can ignore $\frac{1}{h^2}$.

(C) From $y = f(x) + \epsilon$

$$\text{Var}(\vec{y}) = \sigma^2 \quad (14)$$

$$\text{Mean}(\vec{y}) = f(\vec{x}) \quad (15)$$

$$\begin{aligned}
\text{Mean}(a_0) &= \sum_i \gamma_i f(x_i) \\
\text{Var}(a_0) &= \text{TrVar}(H \vec{y}) \\
&= \sum_i |\gamma(x_i, x)|^2 \sigma^2 \quad (16)
\end{aligned}$$

(D) Define $\vec{\mu} = \text{Mean}(\vec{y}) = f(\vec{x})$

$$\begin{aligned}
E(\sigma^2) &\propto E((\vec{y} - H \vec{y})^T (\vec{y} - H \vec{y})) \\
&= E(\vec{y}^T \vec{y}) - 2E(\vec{y}^T H \vec{y}) + E(\vec{y}^T H^2 \vec{y}) \\
&= E(\vec{y}^T \vec{y}) - 2 * (\vec{\mu}^T H \vec{\mu} + \sigma^2 \text{Tr}(H)) + \vec{\mu}^T (H^2) \vec{\mu} + \sigma^2 \text{Tr}(H^2) \quad (17)
\end{aligned}$$

Apply trace trick to the first terms:

$$E(\vec{y}^T \vec{y}) = \text{Tr}(E(\vec{y} \vec{y}^T)) = \text{Tr}(\sigma^2 + \vec{\mu} \vec{\mu}^T) = n\sigma^2 + \vec{\mu}^T \vec{\mu} \quad (18)$$

$$E(\sigma^2) = \sigma^2 + \frac{\vec{\mu}^T (1 - H)^2 \vec{\mu}}{n - 2\text{Tr}(H) + \text{Tr}(H^2)} \quad (19)$$