1   Basic concept

1.1   Bias-variance decomposition

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

$$\text{MSE} = E(e^2) = \int e^2 pde = (\int epde)^2 + (\int e^2 pde - (\int epde)^2) = E(e)^2 + Var(e) \tag{1}$$

1.2   A simple example

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

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

(B)

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

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

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

(C)

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

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

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

$$
\Omega = \mathrm{diag}(\omega_1, \omega_2, \cdots)
\tag{9}
$$

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

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

$$
\begin{aligned}
(R^T \Omega R)^{-1} &= \begin{pmatrix} \sum_i \omega_i & \sum_i R_{i2}\omega_i \\ \sum_i R_{i2}\omega_i & \sum_i R_{i2}^2\omega_i \end{pmatrix}^{-1} \\
&= \frac{1}{(\sum_i \omega_i)(\sum_i R_{i2}^2\omega_i) - (\sum_i R_{i2}\omega_i)^2} \begin{pmatrix} \sum_i R_{i2}^2\omega_i & -\sum_i R_{i2}\omega_i \\ -\sum_i R_{i2}\omega_i & \sum_i \omega_i \end{pmatrix}
\end{aligned}
\tag{10}
$$

$$
R^T \Omega \vec{y} = \begin{pmatrix} \sum_i \omega_i y_i \\ \sum_i R_{i2}\omega_i y_i \end{pmatrix}
\tag{11}
$$

$$f(x) = a_0 \quad = \quad \frac{(\sum_i R_{i2}^2 \omega_i)(\sum_i \omega_i y_i) - (\sum_i R_{i2}\omega_i)(\sum_i R_{i2}\omega_i y_i)}{(\sum_i \omega_i)(\sum_i R_{i2}^2 \omega_i) - (\sum_i R_{i2}\omega_i)^2}$$

$$= \quad \sum_i \frac{((\sum_j R_{j2}^2 \omega_j) - (\sum_j R_{j2}\omega_j)R_{i2})\omega_i}{\sum_k ((\sum_j R_{j2}^2 \omega_j) - (\sum_j R_{j2}\omega_j)R_{k2})\omega_k} y_i \qquad (12)$$

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:

$$\gamma(x_i, x) \quad = \quad ((\sum_j R_{j2}^2 \omega_j) - (\sum_j R_{j2}\omega_j)R_{i2})\omega_i$$

$$= \quad \frac{1}{h^2}((\sum_j (x_j - x)^2 K_j) - (\sum_j (x_j - x)K_j)(x_i - x))K_i$$

$$= \quad \frac{1}{h^2}(s_2 - s_1(x_i - x))K_i \qquad (13)$$

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 $y = f(x) + \epsilon$

$$\mathrm{Var}(\vec{y}) \quad = \quad \sigma^2 \qquad (14)$$
$$\mathrm{Mean}(\vec{y}) \quad = \quad f(\vec{x}) \qquad (15)$$

$$\mathrm{Mean}(a_0) \quad = \quad \sum_i \gamma_i f(x_i)$$

$$\mathrm{Var}(a_0) \quad = \quad \mathrm{TrVar}(H\vec{y})$$

$$= \quad \sum_i |\gamma(x_i, x)|^2 \sigma^2 \qquad (16)$$

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

$$E(\sigma^2) \quad \propto \quad E((\vec{y} - H\vec{y})^T(\vec{y} - H\vec{y}))$$
$$= \quad E(\vec{y}^T \vec{y}) - 2E(\vec{y}^T H\vec{y}) + E(\vec{y}^T H^2 \vec{y})$$
$$= \quad E(\vec{y}^T \vec{y}) - 2 * (\vec{\mu}^T H\vec{\mu} + \sigma^2 \mathrm{Tr}(H)) + \vec{\mu}^T (H^2)\vec{\mu} + \sigma^2 \mathrm{Tr}(H^2) \quad (17)$$

Apply trace trick to the first terms:

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

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

3

5   Gaussian Process

(A)

(B) Define C_{ij} to be element of the inverse of covariance matrix computed by covariance function, from the joint distribution

$$f(x_1, x_2, \cdots, x_n, x^*) \quad \propto \quad e^{-\frac{1}{2}\sum_{i,j} C_{ij} x_i x_j} \tag{20}$$

we have

$$f(x^*|x_1, x_2, \cdots, x_n) \quad \propto \quad e^{-\frac{1}{2}(C_{**}x^{*2} + 2\sum_i C_{*i} x^* x_i)} \tag{21}$$

So we have $(x^*|x_1, \cdots, x_n) \sim N(-\frac{1}{C_{**}}\sum_i C_{*i} x_i, \frac{1}{C_{**}})$

(C) For any linear combination of $(\vec{y}, \vec{\theta})$: $(\vec{a}^T \vec{y} + \vec{b}^T \vec{\theta})$. Compute the moment generating function

$$E(e^{t\vec{a}^T \vec{y} + t\vec{b}^T \vec{\theta}}) \quad = \quad \int e^{t\vec{a}^T \vec{y} + t\vec{b}^T \vec{\theta}} p(\vec{y}|\vec{\theta}) p(\vec{\theta}) d\vec{y} d\vec{\theta} \tag{22}$$

Change the variable being integrated

$$\vec{y} \quad \to \quad \vec{y}' = \vec{y} - R\vec{\theta} \tag{23}$$

$$\vec{\theta} \quad \to \quad \vec{\theta}' = \vec{\theta} - \vec{m} \tag{24}$$

The Jacobian of this transformation is 1. The logrithmic pdf of $e^{t\vec{a}^T \vec{y} + t\vec{b}^T \vec{\theta}}$ is:

$$t\vec{a}^T \vec{y}' + t(\vec{b}^T + \vec{a}^T R)\vec{\theta}' + t(\vec{b}^T + \vec{a}^T R)\vec{m} - \frac{\vec{y}'^T \Sigma^{-1} \vec{y}'}{2} - \frac{\vec{\theta}'^T V^{-1} \vec{\theta}'}{2} + \cdots \tag{25}$$

Complete the square for $\vec{y}'$,

$$-\frac{\vec{y}'^T \Sigma^{-1} \vec{y}'}{2} + t\vec{a}^T \vec{y}' = -\frac{(\vec{y}' - t\Sigma\vec{a})^T \Sigma^{-1}(\vec{y}' - t\Sigma\vec{a})}{2} + \frac{\vec{a}^T \Sigma \vec{a}}{2} t^2 \tag{26}$$

and similarly for $\vec{\theta}'$. The t-dependent part after the the integrating is

$$\exp\left(\frac{\vec{a}^T \Sigma \vec{a} + (\vec{b} + R^T \vec{a})^T V(\vec{b} + R^T \vec{a})}{2} t^2 + (\vec{b}^T + R^T \vec{a})t \right) \tag{27}$$

The coefficient of t^2 is positive definite. And the constant multiplied to this exponential must be 1 because $M(0) = E(1) = 1$.

6   In nonparametric regression and spatial smoothing

(A)

$$p(f(\vec{x}), \vec{y}|\vec{x}) \quad = \quad p(\vec{y}|f(\vec{x})) p(f(\vec{x})|\vec{x}) \tag{28}$$

$(\vec{y}|f(\vec{x})) \sim N(f(\vec{x}), \sigma^2 I)$