Homework of Chapter 7

Chen Cheng, 1130339005

January 12, 2015

Ex. 7.1

Solution. Because $\hat{y}$ is obtained by a linear fit, we have $\hat{\mathbf{y}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$. We note $\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ to be $\mathbf{H}$.

$$\begin{aligned}
\sum_{i=1}^{N} Cov(\hat{y}_i, y_i) &= trace(Cov(\hat{\mathbf{y}}, \mathbf{y})) \\
&= trace(Cov(\mathbf{H}\mathbf{y}, \mathbf{y})) \\
&= trace(\mathbf{H}\, Cov(\mathbf{y}, \mathbf{y})) \\
&= trace(\mathbf{H}\, Var(\mathbf{y})) \\
&= trace(\mathbf{H}(\sigma_\varepsilon^2 \mathbf{I})) \\
&= \sigma_\varepsilon^2 trace(\mathbf{H}) \\
&= \sigma_\varepsilon^2 trace(\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T) \\
&= \sigma_\varepsilon^2 trace(\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}) \\
&= \sigma_\varepsilon^2 trace(\mathbf{I}_d) \\
&= d\sigma_\varepsilon^2
\end{aligned}$$

combine this result with (7.22):

$$E_{\mathbf{y}}(Err_{in}) = E_{\mathbf{y}}(\overline{err}) + \frac{2}{N}\sum_{i=1}^{N} Cov(\hat{y}_i, y_i)$$

we have

$$E_{\mathbf{y}}(Err_{in}) = E_{\mathbf{y}}(\overline{err}) + 2 * \frac{d}{N}\sigma_\varepsilon^2$$

□

Ex. 7.3

Solution. (a) We can use the solution of Ex 5.13 to prove this problem.
Smoothing matrix:

$$\mathbf{S}_\lambda = \mathbf{X}(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{\Omega})^{-1}\mathbf{X}^T$$
$$\mathbf{S}_{ii} = \mathbf{x}_i^T(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{\Omega})^{-1}\mathbf{x}_i$$

The linear fitting function:

$$\hat{f}(\mathbf{x}_i) = \mathbf{x}_i^T(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{\Omega})^{-1}\mathbf{X}^T\mathbf{y}$$

Denote$\mathbf{X}_{(-i)}$and$\mathbf{y}_{(-i)}$be the matrix and the corresponding output after the removal of $\mathbf{x}_i$. So the linear fitting function turns to:

$$\hat{f}^{(-i)}(\mathbf{x}_i) = \mathbf{x}_i^T(\mathbf{X}_{(-i)}^T\mathbf{X}_{(-i)} + \lambda\mathbf{\Omega})^{-1}\mathbf{X}_{(-i)}^T\mathbf{y}_{(-i)}$$

Note that

$$\mathbf{X}_{(-i)}^T\mathbf{X}_{(-i)} = \mathbf{X}^T\mathbf{X} - \mathbf{x}_i\mathbf{x}_i^T$$
$$\mathbf{X}_{(-i)}^T\mathbf{y}_{(-i)} = \mathbf{X}^T\mathbf{y} - \mathbf{x}_i\mathbf{y}_i$$

Lemma:

$$(\mathbf{A} + \mathbf{u}\mathbf{v}^T)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{u}\mathbf{v}^T\mathbf{A}^{-1}}{1 + \mathbf{u}^T\mathbf{A}^{-1}\mathbf{v}}$$

Denote:

$$\mathbf{A} = \mathbf{X}^T\mathbf{X} + \lambda\mathbf{\Omega}$$

So,

$$\begin{aligned}
\mathbf{x}_i^T(\mathbf{X}_{(-i)}^T\mathbf{X}_{(-i)} + \lambda\mathbf{\Omega})^{-1} &= \mathbf{x}_i^T(\mathbf{A} - \mathbf{x}_i\mathbf{x}_i^T)^{-1} \\
&= \mathbf{x}_i^T\mathbf{A}^{-1} + \frac{\mathbf{x}_i^T\mathbf{A}^{-1}\mathbf{x}_i\mathbf{x}_i^T\mathbf{A}^{-1}}{1 - \mathbf{x}_i^T\mathbf{A}^{-1}\mathbf{x}_i} \\
&= \mathbf{x}_i^T\mathbf{A}^{-1} + \frac{\mathbf{S}_{ii}\mathbf{x}_i^T\mathbf{A}^{-1}}{1 - \mathbf{S}_{ii}}
\end{aligned}$$

then,

$$\begin{aligned}
\hat{f}^{(-i)}(\mathbf{x}_i) &= (\mathbf{x}_i^T\mathbf{A}^{-1} + \frac{\mathbf{S}_{ii}\mathbf{x}_i^T\mathbf{A}^{-1}}{1-\mathbf{S}_{ii}})(\mathbf{X}^T\mathbf{y} - \mathbf{x}_i\mathbf{y}_i) \\
&= \mathbf{x}_i^T\mathbf{A}^{-1}\mathbf{X}^T\mathbf{y} + \frac{\mathbf{S}_{ii}\mathbf{x}_i^T\mathbf{A}^{-1}}{1-\mathbf{S}_{ii}}\mathbf{X}^T\mathbf{y} - \frac{\mathbf{S}_{ii}\mathbf{x}_i^T\mathbf{A}^{-1}}{1-\mathbf{S}_{ii}}\mathbf{x}_i\mathbf{y}_i - \mathbf{x}_i^T\mathbf{A}^{-1}\mathbf{x}_i\mathbf{y}_i \\
&= \hat{f}(\mathbf{x}_i) + \frac{\mathbf{S}_{ii}\hat{f}(\mathbf{x}_i)}{1-\mathbf{S}_{ii}} - \frac{\mathbf{S}_{ii}\mathbf{S}_{ii}}{1-\mathbf{S}_{ii}}\mathbf{y}_i - \mathbf{S}_{ii}\mathbf{y}_i \\
&= \frac{\hat{f}(\mathbf{x}_i)}{1-\mathbf{S}_{ii}} - \frac{\mathbf{y}_i\mathbf{S}_{ii}}{1-\mathbf{S}_{ii}}
\end{aligned}$$

Then we have:

$$\mathbf{y}_i - \hat{f}^{(-i)}(\mathbf{x}_i) = \frac{\mathbf{y}_i - \hat{f}(\mathbf{x}_i)}{1-\mathbf{S}_{ii}}$$

(b) It's obvious that $0 \leq S_{ii} \leq 1$, so we get the conclusion.
(c) If the recipe for producing $\hat{f}$ from y does not depend on y itself and S depends only on the x_i and λ, result (7.64) hold.

□

Ex. 7.4

Solution. Because $\mathbf{y}$ and Y^0 are i.i.d., we have $E_{Y^0}(Y_i^0) = E_{\mathbf{y}}(y_i)$ and $E_{Y^0}((Y_i^0)^2) = E_{\mathbf{y}}(y_i^2)$, let $\hat{y}_i = \hat{f}(x_i)$, Therefore,

$$\begin{aligned}
\omega =& E_{\mathbf{y}}(op) \\
=& E_{\mathbf{y}}(Err_{in} - \overline{err}) \\
=& \frac{1}{N}\sum_{i=1}^{N} E_{\mathbf{y}}[E_{Y^0}(Y_i^0 - \hat{f}(x_i))^2 - (y_i - \hat{f}(x_i)^2)] \\
=& \frac{1}{N}\sum_{i=1}^{N} E_{\mathbf{y}}[E_{Y^0}((Y_i^0)^2) - 2\hat{y}_i E_{Y^0}(Y_i^0) + \hat{y}_i{}^2 - (y_i^2 - 2y_i\hat{y}_i + \hat{y}_i^2)] \\
=& \frac{1}{N}\sum_{i=1}^{N}[E_{Y^0}((Y_i^0)^2) - 2E_{\mathbf{y}}(\hat{y}_i)E_{Y^0}(Y_i^0) - E_{\mathbf{y}}(y_i^2) + 2E_{\mathbf{y}}(y_i\hat{y}_i)] \\
=& \frac{2}{N}\sum_{i=1}^{N}[-E_{\mathbf{y}}(\hat{y}_i)E_{\mathbf{y}}(y_i) + E_{\mathbf{y}}(y_i\hat{y}_i)] \\
=& \frac{2}{N}\sum_{i=1}^{N} Cov(\hat{y}_i, y_i)
\end{aligned}$$

□

Ex. 7.5

Solution.

$$\begin{aligned}
\sum_{i=1}^{N} Cov(\hat{\mathbf{y}}_i, \mathbf{y}_i) =& trace(Cov(\hat{\mathbf{y}}, \mathbf{y})) \\
=& trace(Cov(\mathbf{S}\mathbf{y}, \mathbf{y})) \\
=& trace(\mathbf{S}\, Cov(\mathbf{y}, \mathbf{y})) \\
=& trace(\mathbf{S}\, Var(\mathbf{y})) \\
=& trace(\mathbf{S}(\sigma_\varepsilon^2 \mathbf{I})) \\
=& trace(\mathbf{S})\sigma_\varepsilon^2
\end{aligned}$$

□