Chapter 7

Model Assessment and Selection

Problem 7.4

$$\mathbb{E}\left[\overline{\text{err}}(\boldsymbol{y})\right]_{\boldsymbol{y}} = \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left[(y_i - \hat{y}_i(\boldsymbol{y}))^2\right]_{\boldsymbol{y}}$$

$$\mathbb{E}\left[\text{Err}_{\text{in}}(\boldsymbol{y})\right]_{\boldsymbol{y}} = \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left[\mathbb{E}\left[(z_i - \hat{y}_i(\boldsymbol{y}))^2\right]_{\boldsymbol{z}}\right]_{\boldsymbol{y}} \text{ with } \boldsymbol{z} \text{ having the same distribution as } \boldsymbol{y}$$

$$\begin{aligned}
&\mathbb{E}\left[(y_i - \hat{y}_i(\boldsymbol{y}))^2\right]_{\boldsymbol{y}} \\
=&\mathbb{E}\left[\Big(y_i \underbrace{-\mathbb{E}[y_i] + \mathbb{E}[y_i]}_{=0} \underbrace{-\mathbb{E}[\hat{y}_i(\boldsymbol{y})] + \mathbb{E}[\hat{y}_i(\boldsymbol{y})]}_{=0} - \hat{y}_i(\boldsymbol{y})\Big)^2\right]_{\boldsymbol{y}} \\
=&\mathbb{E}\left[\Big(\underbrace{y_i - \mathbb{E}[y_i]}_{\text{expectation}=0} + \underbrace{\mathbb{E}[y_i] - \mathbb{E}[\hat{y}_i(\boldsymbol{y})]}_{\text{constant}} + \underbrace{\mathbb{E}[\hat{y}_i(\boldsymbol{y})] - \hat{y}_i(\boldsymbol{y})}_{\text{expectation}=0}\Big)^2\right]_{\boldsymbol{y}} \\
=&\left(\mathbb{E}[y_i] - \mathbb{E}[\hat{y}_i(\boldsymbol{y})]\right)^2 + \mathbb{E}\left[(y_i - \mathbb{E}[y_i] + \mathbb{E}[\hat{y}_i(\boldsymbol{y})] - \hat{y}_i(\boldsymbol{y}))^2\right]_{\boldsymbol{y}} \\
=&\left(\mathbb{E}[y_i] - \mathbb{E}[\hat{y}_i(\boldsymbol{y})]\right)^2 + \mathbb{E}\left[(y_i - \mathbb{E}[y_i])^2\right] + \mathbb{E}\left[(\hat{y}_i(\boldsymbol{y}) - \mathbb{E}[\hat{y}_i(\boldsymbol{y})])^2\right] \\
&- 2\mathbb{E}\left[(y_i - \mathbb{E}[y_i])(\hat{y}_i(\boldsymbol{y}) - \mathbb{E}[\hat{y}_i(\boldsymbol{y})])\right] \\
=&\text{ModelBias}_i + \text{Var}(y_i) + \text{Var}(\hat{y}_i(\boldsymbol{y})) - \text{Cov}(y_i, \hat{y}_i(\boldsymbol{y})).
\end{aligned}$$

Similarly, we can show

$$\begin{aligned}
&\mathbb{E}\left[\mathbb{E}\left[(z_i - \hat{y}_i(\boldsymbol{y}))^2\right]_{\boldsymbol{z}}\right]_{\boldsymbol{y}} \\
=&\text{ModelBias}_i + \text{Var}(y_i) + \text{Var}(\hat{y}_i(\boldsymbol{y})).
\end{aligned}$$

Combining all the above equations yields

$$\mathbb{E}\left[\overline{\text{err}}(\boldsymbol{y})\right]_{\boldsymbol{y}} - \mathbb{E}\left[\text{Err}_{\text{in}}(\boldsymbol{y})\right]_{\boldsymbol{y}} = \frac{2}{N}\sum_{i=1}^{N}\text{Cov}(y_i, \hat{y}_i(\boldsymbol{y}))$$