Ex. 7.4 Consider the in-sample prediction error (7.18) and the training error $\overline{\text{err}}$ in the case of squared-error loss:

$$\begin{aligned}\text{Err}_{\text{in}} &= \frac{1}{N}\sum_{i=1}^{N} \text{E}_{Y^0}(Y_i^0 - \hat{f}(x_i))^2 \\ \overline{\text{err}} &= \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{f}(x_i))^2.\end{aligned}$$

Add and subtract $f(x_i)$ and $\text{E}\hat{f}(x_i)$ in each expression and expand. Hence establish that the average optimism in the training error is

$$\frac{2}{N}\sum_{i=1}^{N}\text{Cov}(\hat{y}_i, y_i),$$

as given in (7.21).

$op = Err_{in} - \overline{err}$

We have to show that

$W = E_y(op) = E_y(Err_{in} - \overline{err}) = \frac{2}{N}\cdot\sum_{i=1}^{N} Cov(\hat{y}_i, y_i)$

Proof) $E_y(Err_{in} - \overline{err})$

let $\hat{f}(x_i) = \hat{y}_i$

$= \frac{1}{N}\cdot\sum_{i=1}^{n} E_{y^0} E_y (Y_i^0 - \hat{f}(x_i))^2 - E_y \cdot \frac{1}{N}\sum_{i=1}^{n}(y_i - \hat{f}(x_i))^2$

$= \frac{1}{N}\cdot\sum_{i=1}^{n}\left[E_{y^0}E_y(Y_i^0 - \hat{y}_i)^2 - E_y(y_i - \hat{y}_i)^2\right]$

$= \frac{1}{N}\cdot\sum_{i=1}^{n}\left[E_{y^0}E_y(Y_i^{0^2} - \hat{y}_i)^2 - E_y(y_i - \hat{y}_i)^2\right]$

$= \frac{1}{N}\cdot\sum_{i=1}^{n}\left[E_{y^0}E_y(Y_i^{0^2} - 2Y_i^0\hat{y}_i + \hat{y}_i^2) - E_y(y_i^2 - 2y_i\hat{y}_i + \hat{y}_i^2)\right]$

$= \frac{1}{N}\cdot\sum_{i=1}^{n}\left[E_{y^0}(Y_i^{0^2}) - 2E_{y^0}E_y(Y_i^0\hat{y}_i) + E_y(\hat{y}_i^2)\right.$

$\left. + E_y(\hat{y}_i^2) - E_y(y_i^2) + 2E_y(y_i\hat{y}_i) - E_y(\hat{y}_i^2)\right]$

$= \frac{1}{N}\cdot\sum_{i=1}^{n}\left[E_{y^0}(Y_i^{0^2}) - 2E_{y^0}E_y(Y_i^0\hat{y}_i) - E_y(y_i^2) + 2E_y(y_i\hat{y}_i)\right]$

using that $E_{y^0}(y_i^0) = E_y(y_i^2)$

$$= \frac{1}{N} \cdot \sum_{i=1}^{n} \left[2E_y(y_i \hat{y}_i) - 2E_{y^0} E_y(y_i^0 \hat{y}_i) \right]$$

$$= \frac{1}{N} \cdot \sum_{i=1}^{n} \left[2E_y(y_i \hat{y}_i) - 2E_{y^0}(y_i^0) \cdot E_y(\hat{y}_i) \right]$$

We know that $E_{y^0}(y_i^0) = E_y(y_i)$

$$= \frac{1}{N} \cdot \sum_{i=1}^{n} \left[2E_y(y_i \hat{y}_i) - 2E_{y^0}(y_i) \cdot E_y(\hat{y}_i) \right]$$

$$= \frac{2}{N} \cdot \sum_{i=1}^{n} Cov(y_i, \hat{y}_i)$$