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).

$$w = E_y[\text{Err}_{in}] - E_y[\overline{err}] = E_y\left[\frac{1}{N}\sum_{i=1}^{N} E_{Y^0}[(Y_i^0 - \hat{f}(x_i))^2]\right] - E_y\left[\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{f}(x_i))^2\right] = E_y\left[\frac{1}{N}\sum_{i=1}^{N}\left\{E_{Y^0}[(Y_i^0)^2] - 2\hat{f}(x_i)E_{Y^0}[Y_i^0] + \cancel{\hat{f}(x_i)^2}\right\}\right] - E_Y\left[\frac{1}{N}\sum_{i=1}^{N}\left\{y_i^2 - 2y_i\hat{f}(x_i) + \cancel{\hat{f}(x_i)^2}\right\}\right]$$

$$= E_y\left[\frac{1}{N}\sum_{i=1}^{N}\left\{E_{Y^0}[(Y_i^0)^2] - y_i^2 + 2\hat{f}(x_i)(y_i - E_{Y^0}[Y_i^0])\right\}\right]$$

Since $E_{Y^0}[(Y_i^0)^2] = E_Y[Y_i^2]$, $w = E_y\left[\frac{1}{N}\sum_{i=1}^{N}\left\{E_Y[Y_i^2] - y_i^2 + 2\hat{f}(x_i)(y_i - E_{Y^0}[Y_i^0])\right\}\right]$

Also, since $E_{Y^0}[Y_i^0] = E_Y[Y_i]$, $w = E_y\left[\frac{1}{N}\sum_{i=1}^{N}\left\{E_Y[Y_i^2] - y_i^2 + 2\hat{f}(x_i)(y_i - E_y[Y_i])\right\}\right] = E_Y[Y_i^2] - E_Y\left[\frac{1}{N}\sum_{i=1}^{N} y_i^2\right] + E_Y\left[\frac{1}{N}\sum_{i=1}^{N} 2\hat{f}(x_i)(y_i - E_Y[Y_i])\right]$

$$= \cancel{E_Y[Y_i^2]} - \cancel{\frac{1}{N}\sum_{i=1}^{N}E_Y[Y_i^2]} + \frac{2}{N}\sum_{i=1}^{N}E_Y[\hat{f}(x_i)(y_i - E_Y(Y_i))]$$

$$= \frac{2}{N}\sum_{i=1}^{N}\left\{E_Y[y_i\hat{f}(x_i)] - E_Y(Y_i)\hat{f}(x_i)\right\} = \frac{2}{N}\sum_{i=1}^{N}\left\{E_Y[y_i\hat{f}(x_i)] - E_Y[y_i]E[\hat{f}(x_i)]\right\}$$

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

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