7.4

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:

$$\text{Err}_{\text{in}} = \frac{1}{N}\sum_{i=1}^{N} \text{E}_{Y^0}(Y_i^0 - \hat{f}(x_i))^2$$

$$\overline{\text{err}} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{f}(x_i))^2.$$

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).

$$\overline{err} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{f}(x_i))^2$$

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

and

$$Err_{in} = \frac{1}{N}\sum_{i=1}^{N} E_{Y^0}(Y_i^0 - \hat{f}(x_i))^2$$

$$= \frac{1}{N}\sum_{i=1}^{N} E_{Y^0}\Big((Y_i^0 - f(x_i)) + (f(x_i - E\hat{f}(x_i)) + (E(\hat{f}(x_i) - \hat{f}(x_i)\Big)^2$$

So then $w = E_y(op) = E_y(Err_{in} - \overline{err})$

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

then this simplifies after working through the algebra to

$$E\Big[2\sum_i E_{y^0}\big[(y_i^0 - f(x_i))(E\hat{f}(x_i) - \hat{f}(x_i))$$

$$- 2\sum_i E\big[y - f(x_i)\, E\hat{f}(x_i) - \hat{f}(x_i)\big]\Big]$$

Where $E_{y^0}(y_i^0) = f(x_i)$ so then

$$\omega = \frac{2}{N}\sum_{i=1}^{N} \mathrm{cov}(\hat{f}(x_i), y_i)$$

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

7.6

Ex. 7.6 Show that for an additive-error model, the effective degrees-of-freedom for the k-nearest-neighbors regression fit is N/k.

A linear fitting method is defined as

$$\hat{Y} = Sy$$

where S depends on x_i but not y_i

We can write the KNN regression as

$$\hat{Y}(x) = \frac{1}{K} \underbrace{\sum_{i : x_i \in N_K(x)}}_{\text{sum over the K nearest neighbors}} y_i \quad = \frac{1}{K} \sum_{i=1}^{N} I_{(x_i \in N_K(x))} y_i$$

$$= \frac{1}{K} S y$$

Because the point itself is always included in the averge, the diagonal of S is $\begin{bmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & \ddots & \\ & & & & 1 \end{bmatrix}$

so $tr\left(\frac{1}{K} S\right) = \boxed{\frac{N}{K}}$