

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:

$$\text{Err}_{\text{in}} = \frac{1}{N} \sum_{i=1}^{N} \mathcal{E}_{Y^{0}} (Y_{i}^{0} - \hat{f}(x_{i}))^{2}
 \overline{\text{err}} = \frac{1}{N} \sum_{i=1}^{N} (y_{i} - \hat{f}(x_{i}))^{2}.$$

Add and subtract $f(x_i)$ and $E\hat{f}(x_i)$ 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).

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

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

$$+ (E[\hat{f}(x_i)] - \hat{f}(x_i))]^2$$

and

$$Err_{in} = \frac{1}{N} \sum_{i=1}^{N} E_{i} \left(Y_{i}^{0} - \hat{f}(x_{i})^{2} \right)$$

$$= \frac{1}{N} \sum_{i=1}^{N} E_{i} \left((Y_{i}^{0} - \hat{f}(x_{i})) + (\hat{f}(x_{i}^{-} + \hat{f}(x_{i})) + (\hat{f}(x_{i}^{-} + \hat{f}(x_{i}))^{2} \right)$$

$$+ \left(E(\hat{f}(x_{i}) - \hat{f}(x_{i})^{2} \right)$$

So then
$$W = E_y(op) = E_y(Err_{in} - err)$$

$$\begin{bmatrix}
\frac{1}{N} \sum_{i=1}^{N} \mathbb{E}_{Y^{\circ}} \left((Y_{i}^{\circ} - \mathcal{E}(x_{i})) + (\mathcal{E}(x_{i}^{\circ} - \mathcal{E}(x_{i})) - \mathcal{E}(x_{i})) \\
+ (\mathbb{E}(\hat{\mathcal{E}}(x_{i}) - \hat{\mathcal{E}}(x_{i}))^{2}
\end{bmatrix} - \begin{bmatrix}
\frac{1}{N} \sum_{i=1}^{N} \left[(y_{i}^{\circ} - \mathcal{E}(x_{i})) + (\mathcal{E}(x_{i}^{\circ}) - \mathcal{E}(x_{i})) \\
+ (\mathbb{E}(\hat{\mathcal{E}}(x_{i})) - \hat{\mathcal{E}}(x_{i}))^{2}
\end{bmatrix}^{2}$$

then this Simplifies after working through the algebra to

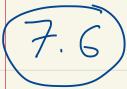
$$E\left[2 = F_{0}\left(Y_{i}^{o} - f(x_{i})\right)\left(E \hat{f}(x_{i}) - \hat{f}(x_{i})\right)\right]$$

$$= 2 = F\left[y - f(x_{i})\right] = \hat{f}(x_{i}) - \hat{f}(x_{i})$$
Where $F_{yo}\left(Y_{i}^{o}\right) = f(x_{i})$ So then

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

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

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



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

A linear fitting method is defined as

Where 5 depends on X; but not y;

We can write the KNN regression as

$$\hat{Y}(x) = \frac{1}{K} \sum_{i:X_i \in N_K(x)} y_i$$

$$= \frac{1}{K} \sum_{i:X_i \in N_K(x)} y_i$$

Becare the Point itself is always included in

the averse, the diasonal of S is

$$50 + r\left(\frac{1}{K}s\right) = \boxed{N}$$