Marisa Blackman

BIOS 7362 HW4 : HTF 7.4 + 7.6

Ex 7.4: Consider Err_{in} and $\overline{err}$ for squared error loss. Add & subtract $f(x_i)$ and $E\hat{f}(x_i)$ in each expression & expand to establish that avg optimism for training error is. $\frac{2}{N}\sum_{i=1}^{N} Cov(\hat{y}_i, y_i)$.

$$Err_{in} = \frac{1}{N}\sum E_{Y^0}(Y_i^0 - \hat{f}(x_i))^2 \qquad \overline{err} = \frac{1}{N}\sum (y_i - \hat{f}(x_i))^2$$

We know we can define average optimism as $E[\text{optimism}] = E_y[Err_{in} - \overline{err}]$. First let's evaluate $E[Err_{in}]$.

$$E_y[Err_{in}] = E_y\left[\frac{1}{n}\sum E_{Y_0}[(Y_i^0 - \hat{f}(X_i))^2]\right]$$
$$= \frac{1}{n}\sum\left[E_y E_{Y^0}[Y_i^{0\,2}] - 2E_y E_{Y^0}[Y_i^0\hat{f}(X_i)] + E_y[\hat{f}(X_i)^2]\right].$$

As stated on p.228, Y^0 is simply another set of observations of y, so Y_0 and y must have the same distribution and are independent of eachother ($Y_0 \sim y$, $Y_0 \perp y$). Also, $\hat{f}(x_i)$ can be expressed as $\hat{y}_i$:

$$E_y[Err_{in}] = \frac{1}{n}\sum\left[E_y[y_i^2] - 2(E_y E_{Y^0}[Y_i^0])(E_y E_{Y_0}[\hat{y}_i]) + E_y[\hat{y}_i^2]\right]$$
$$= \frac{1}{n}\sum\left[E_y[y_i^2] - 2E_y[y_i]E_y[\hat{y}_i] + E_y[\hat{y}_i^2]\right].$$

Now lets evaluate $E[\overline{err}]$.

$$E[\overline{err}] = E_y\left[\frac{1}{n}\sum (y_i - \hat{f}(x_i))^2\right]$$
$$= \frac{1}{n}\sum\left[E_y[y_i^2] - 2E_y[y_i\hat{y}_i] + E_y[\hat{y}_i^2]\right]$$

Putting this together, we have

$$E[\text{optimism}] = \frac{1}{n}\sum_{i=1}^{n}\left[E_y[y_i^2] - 2E_y[y_i]E_y[\hat{y}_i] + E_y[\hat{y}_i^2] - E_y[y_i^2] + 2E_y[y_i\hat{y}_i] - E_y[\hat{y}_i^2]\right]$$
$$= \frac{1}{n}\sum\left[2E_y[y_i\hat{y}_i] - 2E_y[y_i]E_y[\hat{y}_i]\right]$$
$$= \frac{2}{N}\sum E[y_i\hat{y}_i] - E[y_i]E[\hat{y}_i] = \frac{2}{N}\sum_{i=1}^{N} Cov(y_i, \hat{y}_i).$$

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

page 233

From 7.33, $df(\hat{y}) = \frac{1}{\sigma_\varepsilon^2} \sum_{i=1}^{N} \mathrm{Cov}(\hat{Y}_i, y_i)$, where $\hat{Y}_i = \frac{1}{k}[y_i + \sum_{j \neq i} y_j]$.

$$= \frac{1}{\sigma_\varepsilon^2} \sum_{i=1}^{N} \mathrm{Cov}\left(\frac{1}{k} y_i + \frac{1}{k} \sum_{j \neq i} y_j \, , \, y_i\right)$$

$$= \frac{1}{\sigma_\varepsilon^2} \sum \left[\underbrace{\mathrm{Cov}\left(\frac{1}{k} y_i, y_i\right)}_{= \frac{1}{k}\sigma_\varepsilon^2} + \underbrace{\mathrm{Cov}\left(\frac{1}{k} \sum y_j, y_i\right)}_{=0} \right]$$

$=0$ because observations of Y are independent.

$$= \frac{1}{\sigma_\varepsilon^2} \sum_{i=1}^{N} \frac{1}{k} \sigma_\varepsilon^2 = \left(\sigma_\varepsilon^2 / \sigma_\varepsilon^2\right)\left(N/k\right)$$

$$= N/k$$

$\therefore$ The effective df is N/k.