ESL Ex 7.4 p258

Squared error loss.

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

In sample prediction error $\quad Err_{in} = \frac{1}{N}\sum_{i=1}^{N} E_{Y^0}\left[(Y_i^0 - \hat{f}(x_i))^2\right]$

observations in training set, x's kept fixed. E only over new Y^0's

TASK: establish that the average optimism is

$$E_y(op) = \omega = \frac{2}{N}\sum_{i=1}^{N} Cov(\hat{y}_i, y_i) \qquad (7.21)$$

FIRST : short solution from Efron & Hastie (2016) 'CASI' page 220

with slightly differing notation from ESL

Notation: $\mathrm{Cov}(\hat{\mu}_i, y_i) = E\left((\hat{\mu}_i - \mu_i)(y_i - \mu_i)\right)$ [alt. μ_i^* if $E(\hat{\mu}_i) = \mu_i^* \neq \mu$]

And the proof works on one element i (and no N).

What is shown is: $\boxed{E(Err_i) = E(err_i) + 2\,\mathrm{Cov}(\hat{\mu}_i, y_i)}$

where $Err_i = E_0\left((y_{i0} - \hat{\mu}_i)^2\right)$ and $err_i = (y_i - \hat{\mu}_i)^2$

Err_{in} $\quad$ err

Let

$\varepsilon_i = y_i - \mu_i \;,\quad \delta_i = \hat{\mu}_i - \mu_i$

$(\varepsilon_i - \delta_i)^2 = \varepsilon_i^2 - 2\varepsilon_i\delta_i + \delta_i^2$, so that

(1) $(y_i - \hat{\mu}_i)^2 = (y_i - \mu_i)^2 - 2(\hat{\mu}_i - \mu_i)(y_i - \mu_i) + (\hat{\mu}_i - \mu_i)^2$

If we now instead look at a new y_{0i} – that is independent of $\hat{\mu}_i$ (not part of estimating $\hat{\mu}_i$) then similarly

(2) $(y_{i0} - \hat{\mu}_i)^2 = (y_{i0} - \mu_i)^2 - 2(\hat{\mu}_i - \mu_i)(y_{i0} - \mu_i) + (\hat{\mu}_i - \mu_i)^2$

E of (1) gives

$$\underbrace{E\left((y_i - \hat{\mu}_i)^2\right)}_{E(err_i)} = \underbrace{E\left((y_i - \mu_i)^2\right)}_{\substack{\sigma^2 \\ \text{def of variance}}} - 2\cdot\mathrm{Cov}(\hat{\mu}_i, y_i) + E\left((\hat{\mu}_i - \mu_i)^2\right)$$

E of (2)

$\hat{\mu}_i$ independent of y_{i0}

$$\underbrace{E\left((y_{i0} - \hat{\mu}_i)^2\right)}_{E(Err_i)} = \sigma^2 - 2\underbrace{\mathrm{Cov}(\hat{\mu}_i, y_{i0})}_{0} + E\left((\hat{\mu}_i - \mu_i)^2\right)$$

Diff of E of (2) and (1) gives $E(Err_i) = E(err_i) + 2\,\mathrm{Cov}(\hat{\mu}_i, y_i)$

SECOND VERSION – in ESL-notation

$$\underset{\substack{\uparrow \\ \text{expected value over} \\ \text{the responses in the} \\ \text{training set}}}{E_y}(op) = E_y(Err_{in} - \overline{err}) \qquad \text{"definition"}$$

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

$$= E_y\left[\frac{1}{N}\sum_{i=1}^{N}\left(E_{Y_0}\left[(Y_i^0)^2 - 2Y_i^0\cdot\hat{f}(x_i) + \hat{f}(x_i)^2\right] - (y_i^2 - 2y_i\hat{f}(x_i) + \hat{f}(x_i)^2)\right)\right]$$

$$= E_y\left[\frac{1}{N}\sum_{i=1}^{N}\left(E_{Y_0}[(Y_i^0)^2] - 2\underset{\substack{\uparrow \\ \text{not a} \\ \text{function of} \\ Y_i^0}}{\hat{f}(x_i)}\, E_{Y_0}(Y_i^0) + \cancel{\hat{f}(x_i)^2} - y_i^2 + 2y_i\hat{f}(x_i) - \cancel{\hat{f}(x_i)^2}\right)\right]$$

$E_{Y_0}(\hat{f}(x_i)) = \hat{f}(x_i)$ and $E_{Y_0}(\hat{f}(x_i)^2) = \hat{f}(x_i)^2$

$$= \frac{1}{N}\sum_{i=1}^{N}\left(\underbrace{E_y \underbrace{E_{Y_0}(Y_i^{0\,2})}_{E_y(y_i^2)}}_{\cancel{E_y(y_i^2)}} - 2E_y\left(\hat{f}(x_i)\cdot E_{Y_0}(Y_i^0)\right) - \cancel{E_y(y_i^2)} + 2E_y(y_i\cdot\hat{f}(x_i))\right)$$

same distribution

Also $E_{Y_0}(Y_i^0) = E_y(y_i)$ and

$E_y(E_{Y_0}(Y_i^0)) = E_y(y_i)$

$$= \frac{1}{N}\sum_{i=1}^{N}\left(-2\,E_y(\hat{f}(x_i))\cdot E_y(y_i) + 2E_y(y_i\cdot\hat{f}(x_i))\right)$$

$$Cov(y_i, \hat{y}_i) = Cov(y_i, \hat{f}(x_i)) = E_y(y_i\cdot\hat{f}(x_i)) - E_y(y_i)\cdot E_y(\hat{f}(x_i))$$

$$\text{Cov}(V,W) = E(V \cdot W) - E(V) \cdot E(W)$$

def. of covariance

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

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