Ex. 7.7.

$$GCV(\hat{f}) = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{y_i - \hat{f}(x_i)}{1-\text{trace}(S)/N}\right]^2$$

$$= \frac{1}{N}\sum_{i=1}^{N}\left[y_i - \hat{f}(x_i)\right]^2\left(1+2\,\frac{\text{trace}(S)}{N}\right)$$

$$= \overline{err} + \frac{2}{N}\,\overline{err}\cdot\text{trace}(S).$$

While

$$AIC = \overline{err} + 2\cdot\frac{d}{N}\,\hat{\sigma}_\varepsilon^2.$$

Now let's justify that

$$\overline{err}\cdot\text{trace}(S) \simeq d\cdot\hat{\sigma}_\varepsilon^2.$$

(1) In the case of linear regression, $S = X(X^TX)^{-1}X^T$, and $\text{trace}(S) = d$.

(2) We can obtain $\hat{\sigma}_\varepsilon^2$ from the mean squared error of a low-biased model, which can be $\overline{err}$.

□