2.11

$$Q_{n+1}(\hat\beta_{(n+1)}) = \|Y_{n+1} - \hat Y_{n+1}\|^2$$
$$= (Y_{n+1} - \hat Y_{n+1})^T (Y_{n+1} - \hat Y_{n+1})$$
$$= (Y_{n+1} - X_{n+1}\hat\beta_{(n+1)})^T (Y_{(n+1)} - X_{n+1}\hat\beta_{(n+1)})$$

$$= (Y_{n+1} - X_{n+1}[\hat\beta_{(n)} + k_n e_{n+1}])^T (Y_{n+1} - X_{n+1}[\hat\beta_{(n)} + k_n e_{n+1}])$$
$$= \left(\begin{bmatrix} Y_n \\ y_{n+1}\end{bmatrix} - \begin{bmatrix} X_n \\ x_{n+1}^T\end{bmatrix}[\hat\beta_{(n)} + k_n e_{n+1}]\right)^T \left(\begin{bmatrix} Y_n \\ y_{n+1}\end{bmatrix} - \begin{bmatrix} X_n \\ x_{n+1}^T\end{bmatrix}[\hat\beta_{(n)} + k_n e_{n+1}]\right)$$

$$= (Y_n - X_n[\hat\beta_{(n)} + k_n e_{n+1}])^T (Y_n - X_n[\hat\beta_{(n)} + k_n e_{n+1}])$$
$$+ (y_{n+1} - x_{n+1}^T[\hat\beta_{(n)} + k_n e_{n+1}])^2$$

$$= ([Y_n - X_n\hat\beta_{(n)}] - X_n k_n e_{n+1})^T ([Y_n - X_n\hat\beta_{(n)}] + K_n k_n e_{n+1})$$
$$+ ([y_{n+1} - x_{n+1}^T\hat\beta_{(n)}] + x_{n+1}^T k_n e_{n+1})^2$$

$$= (Y_n - X_n\hat\beta_{(n)})^T (Y_n - X_n\hat\beta_{(n)}) + \overbrace{(Y_n - X_n\hat\beta_{(n)})^T X_n}^{=0,\ \text{orthogonal}} k_n e_{n+1}$$
$$- e_{n+1} k_n^T \underbrace{X_n^T (Y_n - X_n\hat\beta_{(n)})}_{0} + e_{n+1} k_n^T X_n^T X_n k_n e_{n+1}$$
$$+ (\underbrace{y_{n+1} - x_{n+1}^T\hat\beta_{(n)}}_{e_{n+1}})^2 - 2(y_{n+1} - x_{n+1}^T\hat\beta_{(n)})(x_{n+1}^T k_n e_{n+1}) + (x_{n+1}^T k_n)^2 e_{n+1}^2$$

$$= Q_n(\hat\beta_{(n)}) + k_n^T X_n^T X_n k_n e_{n+1}^2 + e_{n+1}^2 - 2 e_{n+1} x_{n+1}^T k_n e_{n+1} + (x_{n+1}^T k_n)^2 e_{n+1}^2$$

$$= Q_n(\hat\beta_{(n)}) + [k_n^T X_n^T X_n k_n + 1 - 2 x_{n+1}^T k_n + (x_{n+1}^T k_n)^2] e_{n+1}^2$$

$$= Q_n(\hat\beta_{(n)}) + [(h_n V_n x_{n+1})^T X_n^T X_n h_n V_n x_{n+1} + 1 - 2 x_{n+1}^T h_n V_n x_{n+1} + (x_{n+1}^T h V_n x_{n+1})^2] e_{n+1}^2$$

$$= Q_n(\hat\beta_{(n)}) + [h_n^2 x_{n+1}^T V_n^T X_n^T X_n V_n x_{n+1} + 1 - 2h_n x_{n+1}^T V_n x_{n+1} + h^2 (x_{n+1}^T V_n x_{n+1})^2] e_{n+1}^2$$

$$= Q_n(\hat\beta_{(n)}) + [h_n^2 x_{n+1}^T V_n x_{n+1} + 1 - 2h_n x_{n+1}^T V_n x_{n+1} + h^2 (x_{n+1}^T V_n x_{n+1})^2] e_{n+1}^2$$

$$= Q_n(\hat\beta_{(n)}) + \left[h_n^2 \frac{1-h_n}{h_n} + 1 - 2h_n \frac{1-h_n}{h_n} + h_n^2 \frac{(1-h_n)^2}{h_n^2}\right] e_{n+1}^2$$

$$= Q_n(\hat\beta_{(n)}) + [h_n - h_n^2 + 1 - 2 + 2h_n + 1 - 2h_n + h_n^2] e_{n+1}^2$$

$$= Q_n(\hat\beta_{(n)}) + h_n e_{n+1}^2$$

$$h_n = \frac{1}{1 + x_{n+1}^T V_n x_{n+1}}$$
$$h_n + h_n x_{n+1}^T V_n x_{n+1} = 1$$
$$x_{n+1}^T V_n x_{n+1} = \frac{1-h}{h}$$

$K_n = h_n V_n x_{n+1}$

$V_n = (X_n^T X_n)^{-1}$

$(X^T X)^T = X^T X$

$W_n^T = W_n$

$V_n^T = W^{-1T} = (W^T)^{-1} = W^{-1} = V_n$

3.1

$$E[(\hat{Y}-f(x'))^2] = (E[\hat{Y}-f(x')])^2 + Var(\hat{Y}-f(x'))$$
$$= (E[\hat{Y}] - E[f(x')])^2 + Var(\hat{Y}) + \underbrace{Var(f(x'))}_{=0}$$
$$= (E[\hat{Y}] - f(x'))^2 + Var(\hat{Y}).$$

Recall:
$Var(X) = E[X^2] - E[X]^2$
$\Updownarrow$
$E[X^2] = E[X]^2 + Var(X)$

3.3

Training set: $(x_1, y_1) \dots (x_n, y_n)$
Test set: $(x_{n+1}, y_{n+1}) \dots (x_{n+m}, y_{n+m})$.
$\Big\}$ randomly iid selected data.

Let $\hat{\beta} = \arg\min_\beta R_{train}(\beta) = \arg\min_\beta \frac{1}{n}\sum_{i=1}^{n} (y_i - \beta^T x_i)^2$ and
$\tilde{\beta} = \arg\min_\beta R_{test}(\beta) = \arg\min_\beta \frac{1}{m}\sum_{i=1}^{m} (y_{n+i} - \beta^T x_{n+i})^2$.

Then, $R_{test}(\tilde{\beta}) \leq R_{test}(\hat{\beta})$.

Furthermore, from class and iid assumption, we have that

$$E[R_{train}(\hat{\beta})] = E[R_{test}(\tilde{\beta})].$$

Combine these results, and we get that

$$E[R_{train}(\hat{\beta})] = E[R_{test}(\tilde{\beta})] \leq E[R_{test}(\hat{\beta})].$$

Which we were asked to show.

3.4

Define: $\hat{y}_{-i} = x_i^T \hat{\beta}_{-i}$, here $\hat{\beta}_{-i}$ is fitted based on all data except the i'th; $\hat{\beta}_{-i} = (X_{-i}^T X_{-i})^{-1} X_{-1}^T y_{-i}$.

$P = X(X^TX)^{-1}X^T$, Projection / hat matrix.

Furthermore, we have that

$$X_{-1}^T X_{-1} = X^TX - x_i x_i^T \quad \text{and} \quad X_{-i}^T y_{-i} = X^T y - x_i y_i .$$

From (A.2),

$$(X_{-i}^T X_{-i})^{-1} = (X^TX - x_i x_i^T)^{-1}$$

$$= (X^TX)^{-1} + \frac{(X^TX)^{-1} x_i x_i^T (X^TX)^{-1}}{1 - \underbrace{x_i^T (X^TX) x_i}_{P_{ii}}} .$$

Then

$$\hat{\beta}_{-i} = (X_{-i}^T X_{-i})^{-1} X_{-i}^T y_{-i}$$

$$= (X_{-i} X_{-i})^{-1} (X^T y - x_i y_i)$$

$$= \underbrace{(X^TX)^{-1} X^T y}_{\hat{\beta}} - (X^TX)^{-1} x_i y_i + \frac{(X^TX)^{-1} x_i}{1 - P_{ii}} \Big(x_i^T \underbrace{(X^TX)^{-1} X^T y}_{\hat{\beta}} - \overbrace{x_i^T (X^TX) x_i}^{P_{ii}} y_i \Big)$$

$$= \hat{\beta} - (X^TX)^{-1} x_i y_i + \frac{(X^TX)^{-1} x_i}{1 - P_{ii}} \big(\underbrace{x_i^T \hat{\beta}}_{\hat{y}_i} - P_{ii} y_i \big)$$

$$= \hat{\beta} - \frac{(X^TX)^{-1} x_i}{1 - P_{ii}} \left[(1 - P_{ii}) y_i - \hat{y}_i - P_{ii} y_i \right]$$

$$= \hat{\beta} - \frac{(X^TX)^{-1} x_i}{1 - P_{ii}} \left[y_i - \hat{y}_i \right] .$$

This means that

$$y_i - \hat{y}_{-i} = y_i - x_i^T \left[\hat{\beta} - \frac{(X^TX)^{-1} x_i}{1 - P_{ii}} (y_i - \hat{y}_i) \right]$$

$$= y_i - x_i^T \hat{\beta} + \frac{x_i^T (X^TX)^{-1} x_i}{1 - P_{ii}} (y_i - \hat{y}_i)$$

$$= y_i - \hat{y}_i + \frac{P_{ii}}{1 - P_{ii}} (y_i - \hat{y}_i)$$

$$= (y_i - \hat{y}_i) \left[1 + \frac{P_{ii}}{1 - P_{ii}} \right] = \frac{y_i - \hat{y}_i}{1 - P_{ii}} .$$