Ex 7.4

Wish to show:

$$\mathbb{E}_y[\mathrm{Err}_{in} - \overline{\mathrm{err}}] = 2/N \sum_{i=1}^{N} \mathrm{Cov}(\hat{y}_i, y_i)$$

as suggested in the question we will add and subtract $f(x_i)$ and $\mathbb{E}(\hat{f}(x_i))$ in each expression and expand e.g.

$$\omega = \mathbb{E}_y\Big[1/N \sum_{i=1}^{N} \mathbb{E}_{y_0}\big[y_i^0 - f(x_i) + f(x_i) - \mathbb{E}(\hat{f}(x_i)) + \mathbb{E}(\hat{f}(x_i)) - \hat{f}(x_i)\big]^2 - 1/N \sum_{i=1}^{N} \big(y_i - f(x_i) + f(x_i) - \mathbb{E}(\hat{f}(x_i)) + \mathbb{E}(\hat{f}(x_i)) - \hat{f}(x_i)\big)^2 \Big]$$

Now we will need to expand the left part and the right part.

left expression,

$$\mathbb{E}_y\Big[1/N \sum_{i=1}^{N} \mathbb{E}_{y_0}\Big[\overbrace{(y_i^0 - f(x_i))^2}^{A_1} + \overbrace{(f(x_i) - \mathbb{E}(\hat{f}(x_i)))^2}^{B_1} + \overbrace{(\mathbb{E}(\hat{f}(x_i)) - \hat{f}(x_i))^2}^{C_1} + \overbrace{2(y_i^0 - f(x_i))(f(x_i) - \mathbb{E}(\hat{f}(x_i)))}^{D_1} + \overbrace{2(y_i^0 - f(x_i))(\mathbb{E}(\hat{f}(x_i)) - \hat{f}(x_i))}^{E_1} + \overbrace{2(f(x_i) - \mathbb{E}(\hat{f}(x_i)))(\mathbb{E}(\hat{f}(x_i)) - \hat{f}(x_i))}^{F_1} \Big]\Big]$$

Right expression,

I won't write out the entire right expression, but it would be very similar with corresponding pairs labelled with A_2, B_2, C_2, D_2, E_2 and F_2.

Now we will deal with the cancellation in the entire expression part by part.

$A_1 - A_2 \Rightarrow$ Notice $\mathbb{E}_y \sum_i \mathbb{E}_{y^o} (y_i^o - f(x_i))^2 = \mathbb{E}_y \sum_i (y_i - f(x_i))^2$

$$= N\mathbb{E}[\varepsilon^2] = N\sigma^2$$

Thus $A_1 - A_2 = 0$

$B_1 - B_2 \Rightarrow$ Identical on both sides $\Rightarrow B_1 - B_2 = 0$

$C_1 - C_2 \Rightarrow$ identical on both sides $\Rightarrow C_1 - C_2 = 0$

$D_1 - D_2 \Rightarrow$ Note that in both D's only the y_i and y_i^o Parts are Non-Constant w.r.t $\mathbb{E}_y$. for Example

$$\mathbb{E}(D_1) = 2\mathbb{E}_y[\sum_i (y_i^o - f(x_i))(f(x_i) - \mathbb{E}(\hat{f}(x_i))]$$
$$= 2\sum_i (\mathbb{E}_y(y_i^o) - f(x_i))(f(x_i) - \mathbb{E}(\hat{f}(x_i)))$$

and since $\mathbb{E}_y(y_i^o) = \mathbb{E}_y(y_i) = f(x_i)$

$$= 0$$

By the same logic $\mathbb{E}(D_2) = 0$ also.

$E_1 - E_2 \Rightarrow$ $E_1 = 2\mathbb{E}_y[1/N \sum_i \mathbb{E}_{y^o}(-f(x_i) + y_i^o) \cancel{- \mathbb{E}(\hat{f}(x_i))} (\mathbb{E}(\hat{f}(x_i)) - \hat{f}(x_i))]$

By independance of $y_i^o - f(x_i)$ and $\mathbb{E}(\hat{f}(x_i)) - \hat{f}(x_i)$ we can use the Property $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ and since $\mathbb{E}_{y^o}(y_i^o) = f(x_i)$ we have

$$E_1 = 2\mathbb{E}_y[1/N \sum_i (f(x_i) - f(x_i))(\mathbb{E}(\hat{f}(x_i)) - \hat{f}(x_i))] = 0$$

And $E_2 = \mathbb{E}_y[2\sum_i (y_i - f(x_i))(\mathbb{E}(\hat{f}(x_i)) - \hat{f}(x_i))]$

$$= 2\sum_i \mathbb{E}_y (y_i - \mathbb{E}(y_i))(-(\hat{f}(x_i) - \mathbb{E}(\hat{f}(x_i))))$$

Previous line uses $f(x_i) = \mathbb{E}(y_i)$ and Pulls out minus sign on second expression.

$$= -2\sum_i \mathbb{E}_y (y_i - \mathbb{E}(y_i))(\hat{f}(x_i) - \mathbb{E}(\hat{f}(x_i)))$$
$$= -2\sum_i \operatorname{cov}(y_i, \hat{f}(x_i))$$

Overall $E_1 - E_2 = 2\sum_i \operatorname{cov}(y_i, \hat{f}(x_i))$

$f_1 - f_2 \Rightarrow$ Identical on both sides $\Rightarrow f_1 - f_2 = 0$

Putting this all together and returning to ω we obtain:

$$\omega = \frac{1}{N}\left[0 + 0 + 0 + 0 + 2\sum_i \text{cov}(y_i, \hat{f}(x_i)) + 0\right]$$

and since $\hat{f}(x_i) = \hat{y}_i$

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