Student Number: [redacted] Name: Bryan Hoang

1. (20 points) **Answer:**

Proof. Starting with the LHS of the equality yields

$$E[(Y - \hat{Y})^2] = E[Y^2] - 2E[Y\hat{Y}] + E[\hat{Y}^2]$$

by the linearity of expectation. To try to use the orthogonality principle, let's try rewriting the middle term, $2E[Y\hat{Y}]$, as

$$\begin{aligned} 2E[Y\hat{Y}] &= 2E\Big[(Y - \hat{Y} + \hat{Y})\hat{Y}\Big] \\ &= 2\Big(E\Big[(Y - \hat{Y})\hat{Y}\Big] + E[\hat{Y}^2]\Big) \\ &= 2\left(E\left[(Y - \hat{Y})\sum_{i=1}^{m} a_i X_i\right] + E[\hat{Y}^2]\right) \\ &= 2\left(\sum_{i=1}^{m} a_i E\Big[(Y - \hat{Y})X_i\Big] + E[\hat{Y}^2]\right) \end{aligned}$$

Since $\hat{Y}$ is an optimal linear predictor, then the orthogonality principle says that

$$\begin{aligned} &E\Big[(Y - \hat{Y})X_i\Big] = 0 \quad \forall i = 1, \dots, m \\ \Rightarrow\ &E\Big[(Y - \hat{Y})\hat{Y}\Big] = 0. \end{aligned}$$

Therefore, we have that

$$\begin{aligned} E\Big[(Y - \hat{Y})^2\Big] &= E[Y^2] - 2E[\hat{Y}^2] + E[\hat{Y}^2] \\ &= E[Y^2] - E[\hat{Y}^2]. \end{aligned}$$

□