Student Number: [redacted] Name: Bryan Hoang

2. (25 points)

(a) **Answer:**

Proof. Let be interpreted as $\hat{Y} = aX + b$ a linear predictor of Y from the random variables $X = X_1$ and $X_0 = 1$ with prediction coefficients a and b. Then, the orthogonality principle states that the linear predictor is optimal in the MSE sense when for $i = 0, 1$,

$$\begin{aligned}
E\Big[(Y - \hat{Y})X_i\Big] &= 0 \\
E\Big[\big(Y - (aX + b)\big)X_i\Big] &= 0 \\
E\left[\left(Y - \left(\frac{\operatorname{Cov}(X,Y)}{\operatorname{Var}(X)}X + E[Y] - \frac{\operatorname{Cov}(X,Y)}{\operatorname{Var}(X)}E[X]\right)\right)X_i\right] &= 0 \\
E[YX_i] - E[Y]E[X_i] - \frac{\operatorname{Cov}(X,Y)}{\operatorname{Var}(X)}E[XX_i] + \frac{\operatorname{Cov}(X,Y)}{\operatorname{Var}(X)}E[X]E[X_i] &= 0
\end{aligned}$$

For $i = 0$,

$$\begin{aligned}
&E[YX_i] - E[Y]E[X_i] - \frac{\operatorname{Cov}(X,Y)}{\operatorname{Var}(X)}E[XX_i] + \frac{\operatorname{Cov}(X,Y)}{\operatorname{Var}(X)}E[X]E[X_i] \\
&= E[Y] - E[Y]E[1] - \frac{\operatorname{Cov}(X,Y)}{\operatorname{Var}(X)}E[X] + \frac{\operatorname{Cov}(X,Y)}{\operatorname{Var}(X)}E[X]E[1] \\
&= E[Y] - E[Y] - \frac{\operatorname{Cov}(X,Y)}{\operatorname{Var}(X)}E[X] + \frac{\operatorname{Cov}(X,Y)}{\operatorname{Var}(X)}E[X] \\
&= 0.
\end{aligned}$$

For $i = 1$,

$$\begin{aligned}
&E[YX_i] - E[Y]E[X_i] - \frac{\operatorname{Cov}(X,Y)}{\operatorname{Var}(X)}E[XX_i] + \frac{\operatorname{Cov}(X,Y)}{\operatorname{Var}(X)}E[X]E[X_i] \\
&= E[YX] - E[Y]E[X] - \frac{\operatorname{Cov}(X,Y)}{\operatorname{Var}(X)}E[X^2] + \frac{\operatorname{Cov}(X,Y)}{\operatorname{Var}(X)}E[X]^2 \\
&= \operatorname{Cov}(X,Y) - \frac{\operatorname{Cov}(X,Y)}{\operatorname{Var}(X)}\big(E[X^2] - E[X]^2\big) \\
&= \operatorname{Cov}(X,Y) - \frac{\operatorname{Cov}(X,Y)}{\operatorname{Var}(X)}\operatorname{Var}(X) \\
&= 0.
\end{aligned}$$

Therefore, the orthogonality principle implies that the choices for a and b are optimal. □

(b) **Answer:**

Proof.

$$\begin{aligned}
\hat{Y} &= aX + b \\
&= \frac{\operatorname{Cov}(X,Y)}{\operatorname{Var}(X)}X + E[Y] - \frac{\operatorname{Cov}(X,Y)}{\operatorname{Var}(X)}E[X] \\
&= E[Y] + \frac{\operatorname{Cov}(X,Y)}{\operatorname{Var}(X)}\big(X - E[X]\big) \\
&= E[Y] + \frac{\operatorname{Cov}(X,Y)\sqrt{\operatorname{Var}(Y)}}{\sqrt{\operatorname{Var}(X)}^2\sqrt{\operatorname{Var}(Y)}}\big(X - E[X]\big) \\
&= m_Y + \frac{\rho\sigma_Y}{\sigma_X}(X - m_X),
\end{aligned}$$

Student Number: [redacted] Name: Bryan Hoang

which is what we wanted to prove. □

(c) **Answer:**

Proof.

$$
\begin{aligned}
& E\big[\big(Y-(aX+b)\big)^2\big] \\
&= E\big[\big(Y-(\hat{Y})\big)^2\big] \\
&= E[Y^2]-E[(\hat{Y})^2] && \text{by question 1} \because \hat{Y} \text{ is optimal} \\
&= E[Y^2]-E\left[\left(m_Y+\frac{\rho\sigma_Y}{\sigma_X}(X-m_X)\right)^2\right] && \text{by part (b)} \\
&= E[Y^2]-E[m_Y]-2E\left[m_Y\frac{\rho\sigma_Y}{\sigma_X}(X-m_X)\right]-E\left[\frac{\rho^2\sigma_Y^2}{\sigma_X^2}(X-m_X)^2\right] \\
&= \underbrace{E[Y^2]-m_Y}_{\mathrm{Var}(Y)=\sigma_Y^2}-2m_Y\frac{\rho\sigma_Y}{\sigma_X}(\underbrace{E[X]-m_X}_{m_X-m_X=0})-\frac{\rho^2\sigma_Y^2}{\sigma_X^2}\underbrace{E\big[(X-m_X)^2\big]}_{\mathrm{Var}(X)=\sigma_X^2} \\
&= \sigma_Y^2-\rho^2\sigma_Y^2 \\
&= \sigma_Y^2(1-\rho^2).
\end{aligned}
$$

□