Exercise 3.14

Michael Jagdharry

Part (a)

What is the value of $g(x)$ that minimizes $E[(Y - g(X))^2]$?

$$\begin{aligned} E[(Y - g(X))^2] &= \\ E[((Y - E(Y|X)) - (g(X) - E(Y|X)))^2] &= \\ E[(Y - E(Y|X)^2 + (g(X) - E(Y|X))^2 - 2(Y - E(Y|X)(g(X) - E(Y|X))] \end{aligned}$$

By the law of iterated expectation, and since $E(Y|X)$ and $g(X)$ are constants given X, the rightmost term is

$$\begin{aligned} -2(Y - E(Y|X)) * (g(X) - E(Y|X)) &= \\ -2E[E[(Y - E(Y|X)) * (g(X) - E(Y|X))]|X] &= \\ -2E[(g(X) - E(Y|X)) * E[(Y - E(Y|X))|X]] &= \\ -2(g(X) - E(Y|X)) * [E(Y|X) - E(Y|X)] &= \\ -2(g(X) - E(Y|X)) * 0 &= \\ 0 \end{aligned}$$

We have now shown

$$\begin{aligned} E[(Y - g(X))^2] &= \\ E[(Y - E(Y|X)^2] + E[(g(X) - E(Y|X))^2] \end{aligned}$$

$(g(X) - E(Y|X))^2$ is a nonnegative function of $g(X)$ which is minimized to 0 when $g(X) = E(Y|X)$. Consequently $E[(g(X) - E(Y|X))^2]$ is also minimized to 0 when $g(X) = E(Y|X)$ for the same reasons. We can do nothing about the term $E[(Y - E(Y|X)^2]$, since it does not contain $g(X)$. Thus, $E[(Y - g(X))^2]$ is minimized when $g(X) = E(Y|X)$.

Part (b)

We are given these two statements:

$$X, Z \overset{iid}{\sim} N(0, 1)$$
$$Y = X^2 + Z$$

It follows that

$$\begin{aligned}
g(X) = E(Y|X) &= \\
E(X^2 + Z|X) &= \\
E(X^2|X) + E(Z|X) &= \\
X^2 &
\end{aligned}$$

$$\begin{aligned}
MSE &= \\
E[(Y - g(X))^2] &= \\
E[(X^2 + Z - X^2)^2] &= \\
E[Z^2] &= \\
Var(Z) + E(Z)^2 &= \\
1 &
\end{aligned}$$

So the $MSE = 1$.

Part (c)

Suppose $g(X) = a + bX$. Then the a and b which minimize the MSE are found by setting to 0 the partial derivatives of the MSE with respect to each of a and b.

$$\begin{aligned}
\frac{\partial}{\partial a} MSE &= \\
\frac{\partial}{\partial a} E[(Y - a - bX)^2] &= \\
E[\frac{\partial}{\partial a}(Y - a - bX)^2] &= \\
E[-2(Y - a - bX)] &= 0 \\
E(Y) &= E(a + bX) \\
E(Y) &= a + bE(X) \\
E(Y) - bE(X) &= a
\end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial b} MSE &= \\
\frac{\partial}{\partial b} E[(Y - a - bX)^2] &= \\
E[\frac{\partial}{\partial b}(Y - a - bX)^2] &= \\
E[-2X(Y - a - bX)] &= 0 \\
E[X(Y - a - bX)] &= 0 \\
E[XY - aX - bX^2)] &= 0 \\
E(XY) &= aE(X) + bE(X^2) \\
\text{Substituting in a from } \frac{\partial}{\partial a} MSE \text{ yields} \\
E(XY) &= (E(Y) - bE(X))E(X) + bE(X^2) \\
E(XY) &= E(Y)E(X) - bE(X)^2 + bE(X^2) \\
E(XY) - E(Y)E(X) &= bE(X^2) - bE(X)^2 \\
Cov(X, Y) &= bVar(x)
\end{aligned}$$

$$\begin{aligned}
\frac{Cov(X, Y)}{Var(X)} &= b \\
E(Y) - \frac{Cov(X, Y)}{Var(X)} E(X) &= a
\end{aligned}$$

In our case,

$$
\begin{aligned}
E(X) = E(Z) &= 0 \\
E(X^2) = Var(X) &= 1 \\
E(Y) = E(X^2 + Z) = E(X^2) + E(Z) &= 1
\end{aligned}
$$

$$
\begin{aligned}
Cov(X,Y) = E(XY) - E(X)E(Y) &= \\
E(XY) = E(X(X^2 + Z)) &= \\
E(X^3 + XZ) = E(X^3) + E(X)E(Z) &= \\
E(X^3) &= 0
\end{aligned}
$$

Because we know odd moments of a standard normal random variable are 0. Thus,

$$
\begin{aligned}
a = E(Y) - bE(X) = E(Y) &= 1 \\
b = \frac{Cov(X,Y)}{Var(X)} = \frac{E(XY)}{E(X^2)} &= 0
\end{aligned}
$$

$$
\begin{aligned}
MSE = E[(Y - a - bX)^2] &= \\
E[(Y-1)^2] &= \\
E[(Y^2 - 2Y + 1] &= \\
E[(X^2 + Z)^2 - 2(X^2 + Z) + 1] &= \\
E[(X^4 + 2X^2Z + Z^2 - 2X^2 - 2Z + 1] &= \\
E(X^4) + 2E(X^2)E(Z) + E(Z^2) - 2E(X^2) - 2E(Z) + 1 &= \\
E(X^4) + E(Z^2) - 2E(X^2) + 1 &= \\
3 + 1 - 2 + 1 &= \\
&3
\end{aligned}
$$

So the $MSE = 3$ for the linear predictor, and $MSE = 1$ for the conditional expectation predictor.

This means that predictors of the form $a + bX$ perform worse than predictors of the form $E(Y|X)$, because the minimum squared error is higher for $g(X) = a + bX$ than it is for $g(x) = E(Y|X)$.