Chapter 7

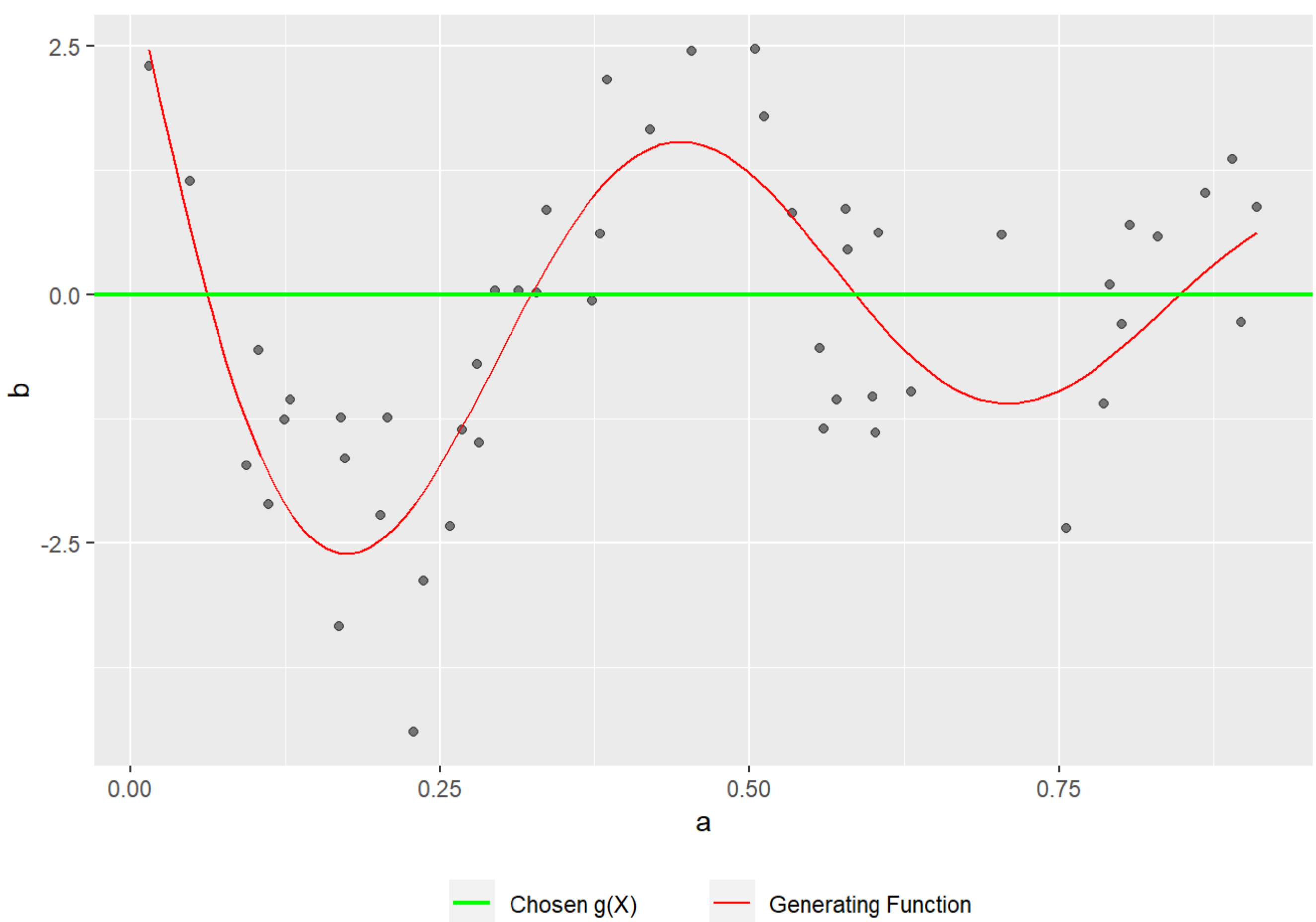
Question 2

We need to generate some data before sketching g_hat under different conditions!

```
library(ggplot2)
set.seed(3)

a <- runif(50)
eps <- rnorm(50)
b <- sin(12*(a + 0.2)) / (a + 0.2) + eps
generating_fn <- function(a) {sin(12*(a + 0.2)) / (a + 0.2)}
df <- data.frame(a, b)
```

2a) $\lambda=\infty, m=0$

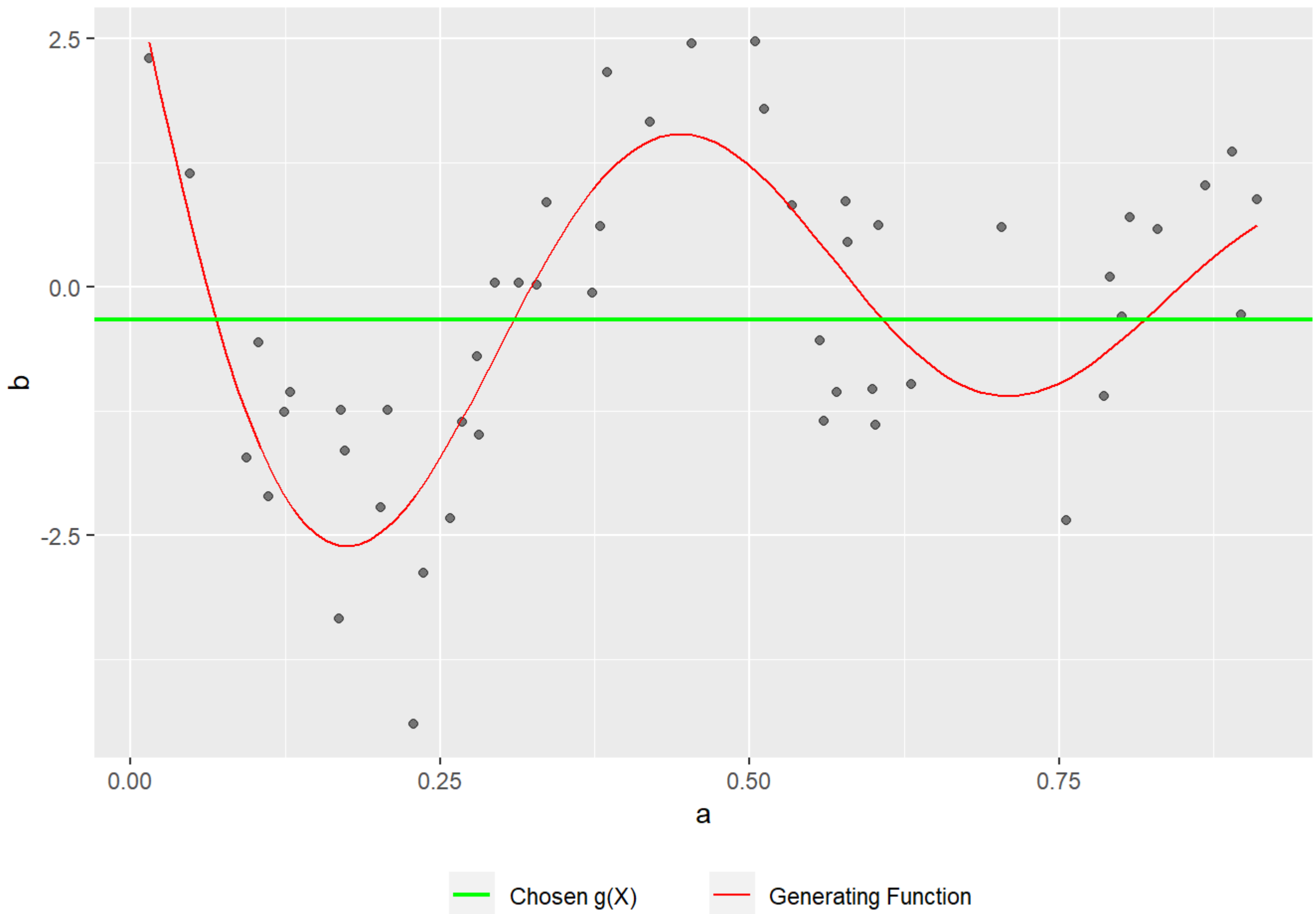
```
ggplot(df, aes(x = a, y = b)) +
  geom_point(alpha = 0.5) +
  stat_function(fun = generating_fn, aes(col = "Generating Function")) +
  geom_hline(aes(yintercept = 0, linetype = "Chosen g(X)"), col = "green", size = 0.8) +
  scale_color_manual(values = "red") +
  theme(legend.position = "bottom", legend.title = element_blank())
```



As λ increases, the penalty term becomes more and more important in the equation. As $\lambda \rightarrow \infty$, this forces $g(x) \rightarrow 0$. We therefore get ghat(x)=0

2b) λ=∞,m=1

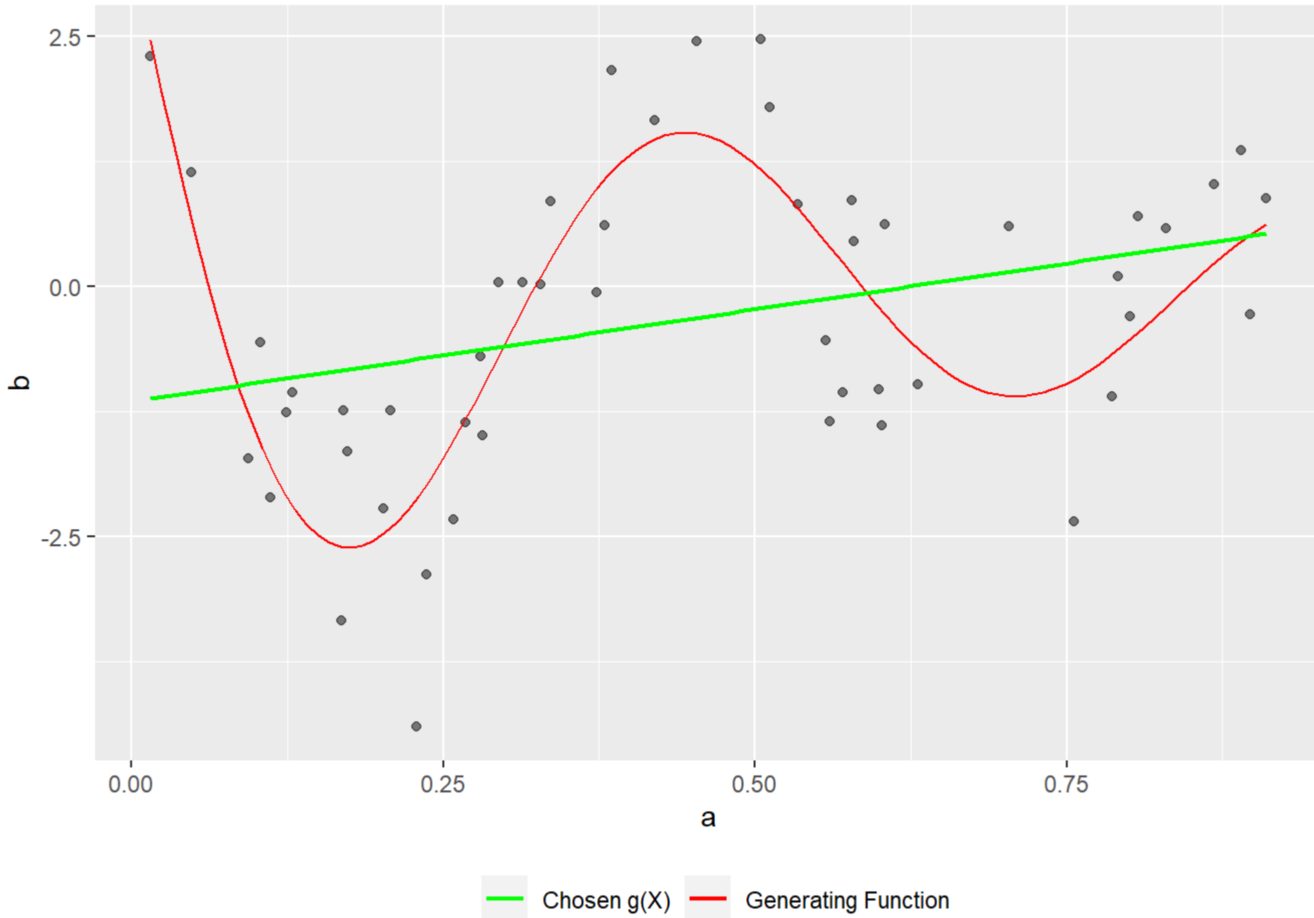
```
ggplot(df, aes(x = a, y = b)) +
  geom_point(alpha = 0.5) +
  stat_function(fun = generating_fn, aes(col = "Generating Function")) +
  geom_hline(aes(yintercept = mean(b), linetype = "Chosen g(X)"), col = "green", size = 0.8) +
  scale_color_manual(values = "red") +
  theme(legend.position = "bottom", legend.title = element_blank())
```



As $\lambda \to \infty$, this forces $g'(x) \to 0$. This means we would get ghat(x)=c.

2c) λ=∞,m=2

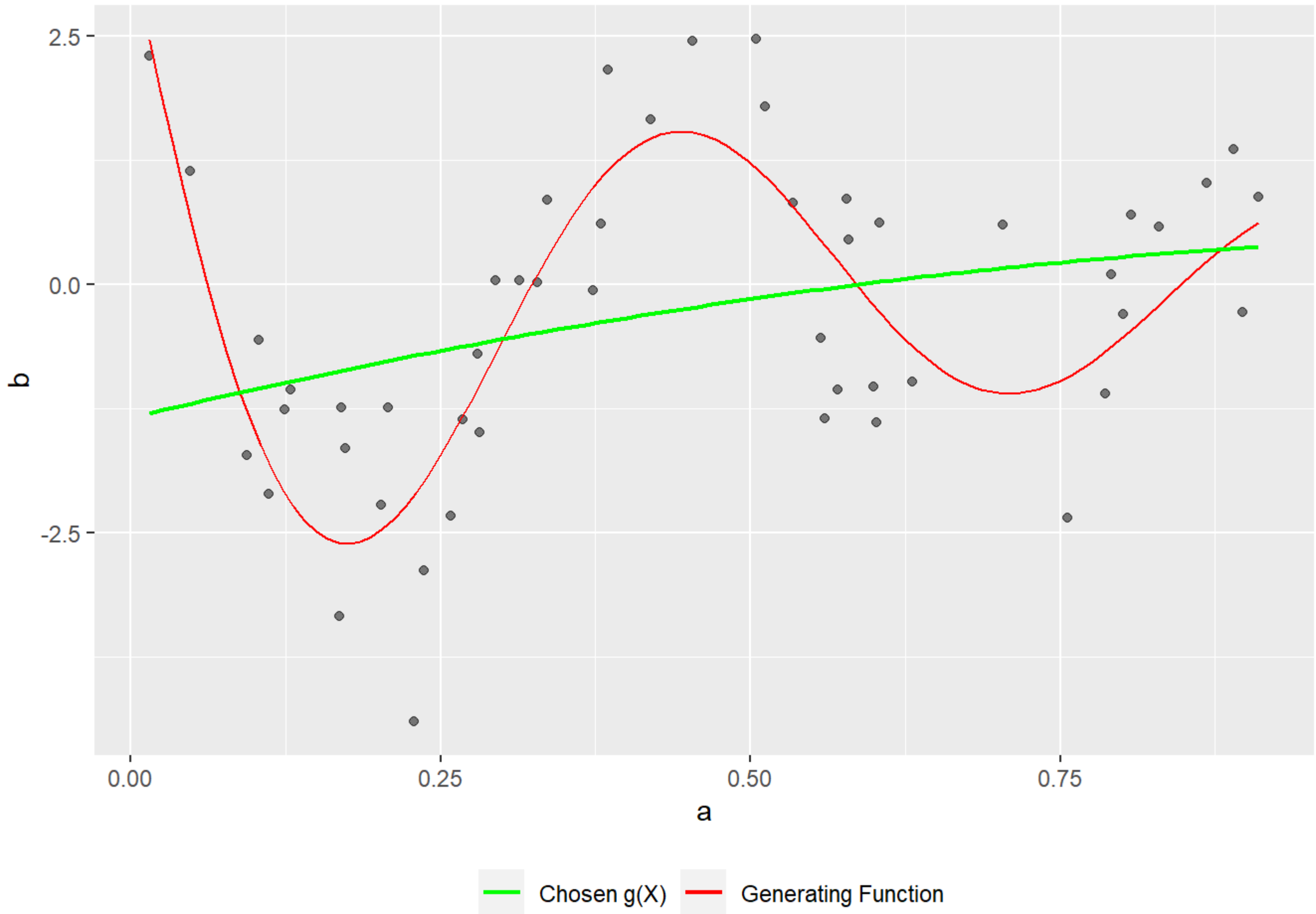
```
ggplot(df, aes(x = a, y = b)) +
  geom_point(alpha = 0.5) +
  stat_function(fun = generating_fn, aes(col = "Generating Function")) +
  geom_smooth(method = "lm", formula = "y ~ x", se = F, size = 0.8, aes(col = "Chosen g(X)")) +
  scale_color_manual(values = c("green", "red")) +
  theme(legend.position = "bottom", legend.title = element_blank())
```



As λ→∞, this forces g″(x)→0. This means we would get ghat(x)=ax+b

2d) λ=∞,m=3

```
ggplot(df, aes(x = a, y = b)) +
  geom_point(alpha = 0.5) +
  stat_function(fun = generating_fn, aes(col = "Generating Function")) +
  geom_smooth(method = "lm", formula = "y ~ x + I(x^2)", se = F, size = 0.8, aes(col = "Chosen g
(X)")) +
  scale_color_manual(values = c("green", "red")) +
  theme(legend.position = "bottom", legend.title = element_blank())
```



As λ→∞, this forces g(3)(x)→0. This means we would get ghat(x)=ax2+bx+c.

2e) λ=0,m=3

```
interp_spline <- smooth.spline(x = df$a, y = df$b, all.knots = T, lambda = 0.0000000000001)
fitted <- predict(interp_spline, x = seq(min(a) - 0.02, max(a) + 0.02, by = 0.0001))
fitted <- data.frame(x = fitted$x, fitted_y = fitted$y)

ggplot(df, aes(x = a, y = b)) +
  geom_point(alpha = 0.5) +
  stat_function(fun = generating_fn, aes(col = "Generating Function")) +
  geom_line(data = fitted,
            aes(x = x, y = fitted_y, col = "Chosen g(X)"), size = 0.8) +
  scale_color_manual(values = c("green", "red")) +
  theme(legend.position = "bottom", legend.title = element_blank())
```

However, since λ=0, the penalty term no longer plays any role in the selection of ghat(x). For this reason, we can achieve RSS = 0

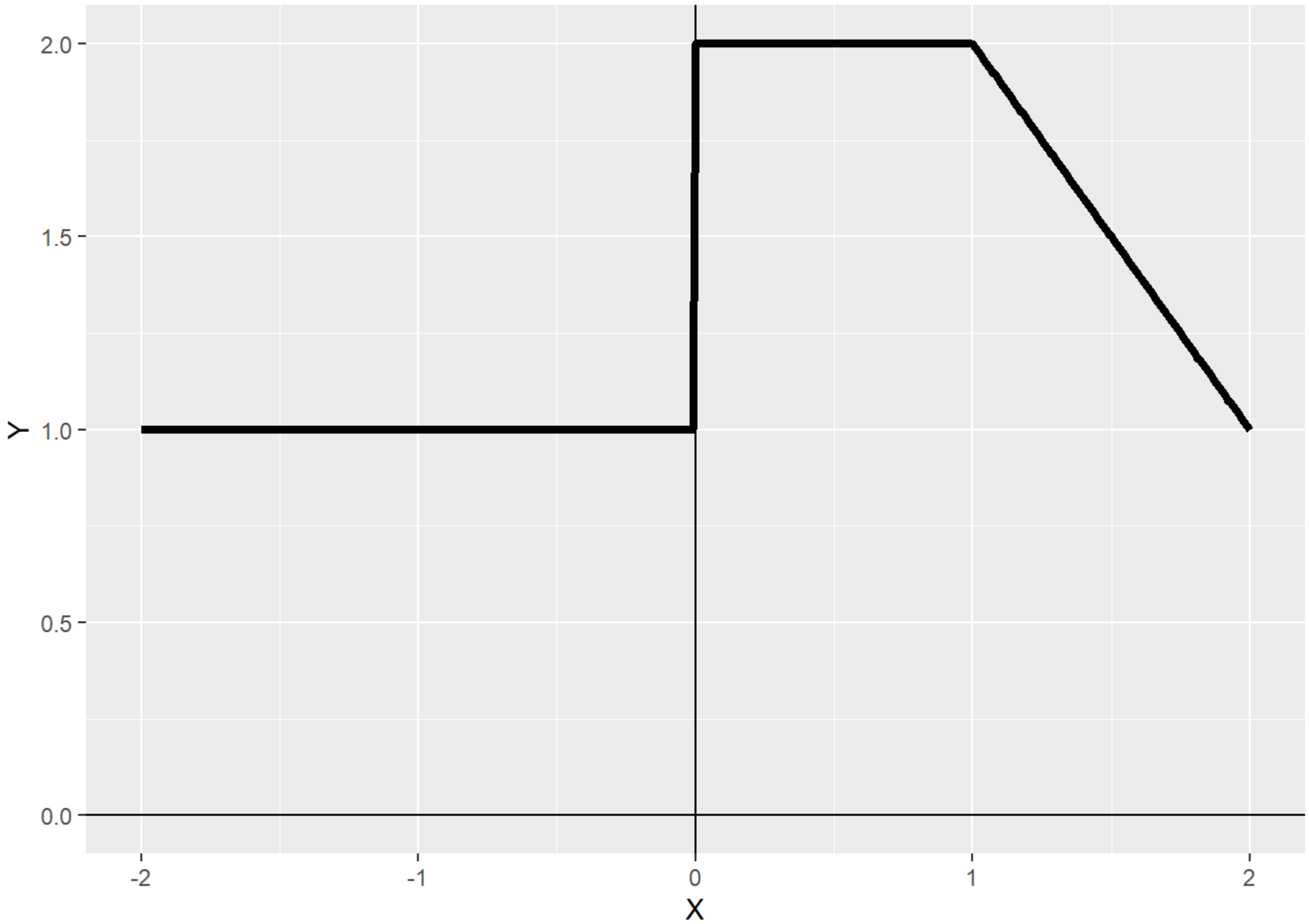
Question 3

```
a = seq(-2, 2, 0.01)
b = 1 + a + -2 * (a - 1)^2 * (a >= 1)
df <- data.frame(a, b)

ggplot(df, aes(x = a, y = b)) +
  geom_vline(xintercept = 0) +
  geom_vline(xintercept = 1, col = "blue") +
  geom_hline(yintercept = 0) +
  geom_line(size = 1.5)
```

The curve is linear between −2 and 1 with y=1+x Quadratic between 1, and 2 with y=1+x−2(x−1)^2

Question 4

```
X = seq(-2, 2, 0.01)
Y = 1 + (X >= 0 & X <= 2) - (X - 1)*(X >= 1 & X <= 2) + 3*(X - 3)*(X >= 3 & X <= 4) + 3*(X > 4 &
X <= 5)
df <- data.frame(X, Y)

ggplot(df, aes(x = X, y = Y)) +
  geom_vline(xintercept = 0) +
  geom_hline(yintercept = 0) +
  geom_line(size = 1.5)
```



The curve is constant between −2 and 0 with y=1 Constant between 0 and 1 with y=2 Linear between 1 and 2 with y=3−x.

Question 5

a. As $\lambda \to \infty$, will g1 or g2 have the smaller training RSS? Answer:- The smoothing spline g2 will most likely have the smaller training RSS since it is a higher order polynomial owing to the penalty term's order (it will be more flexible).

b. As $\lambda \to \infty$, will g1 or g2 have the smaller test RSS? Answer:- The test RSS will depend on the distribution of test data. If we have to provide the behavior of test RSS based on the nature of curve, g2 will have more test RSS as it is more flexible and hence may overfit the data.

c. For $\lambda = 0$, will g1 or g2 have the smaller training and test RSS? Answer:- If $\lambda=0$, we have g1=g2, so they will have the same training and test RSS.