Homework 4 - Machine Learning

Felix Nguyen

March 2019

1 Exercise 1

For a regression splines model with n order and K knots, we can calculate total number of coefficients as follow:

$$\text{Possible Coefficients} * \text{Number of knots} - \text{Total constraints}$$

$$= (n+1)(K+1) - nK = K + n + 1$$

Therefore, we have the total number of coefficients of each model:

$$A = 6 + 4 + 1 = 11;$$

$$B = 5 + 5 + 1 = 11;$$

$$C = 8 + 2 + 1 = 11;$$

$$D = 10 + 3 + 1 = 14;$$

With total coefficients of 4 models is $11 + 11 + 11 + 14 = 47$.

2 Exercise 2

- Test MSE is W, since more knots would initially leads to better fit, which decreases Test MSE, then increases Test MSE again when model is overfit.
- Variance is Y, since more knots leads to higher variance.
- Squared Bias is Z, since Test MSE - Variance - Squared Bias = Irreducible Error, thus must be constant. Between Z and X, Z satisfies this.
- Train MSE is X, since Train MSE would obviously decrease with more knots.

3 Exercise 3

Let's call $L(x_0) = [l_1(x_0), .., l_i(x_0)]^T$ as the vector of all $l_i(x_0)$. We have:

$$L(x_0) = b(x_0)^T (B^T W(x_0) B)^{-1} B^T W(x_0)$$

$$= \sum_{i=1}^{N} l_i(x_0)$$

Therefore,

$$L(x_0)B = b(x_0)^T$$

This equal to

$$\sum_{i=1}^{N} l_i(x_0) x_i^j = x_0^j$$

For linear regression:

$$\sum_{i=1}^{N} l_i(x_0)(x_i - x_0) = x_0 - x_0 = 0$$

For case where j = 0, then

$$\sum_{i=1}^{N} l_i(x_0) = b_0 x_0 = 1$$

For case where $j > 0$, we have

$$\sum_{i=1}^{N} l_i(x_0)(x_i - x_0)^j = \sum_{i=1}^{N} l_i(x_0)[\sum_{k=0}^{j} \binom{j}{k} x_i^k (-x_0)^{j-k}]$$

We also have that:

$$\sum_{i=1}^{N} l_i(x_0)x_i^j = x_0^k (x_0)^{j-k} = \sum_{i=1}^{N} l_i(x_0) x_i^k (x_0)^{j-k}$$

So,

$$\sum_{i=1}^{N} l_i(x_0)(x_i - x_0)^j = \sum_{k=0}^{j} \binom{j}{k} \sum_{i=1}^{N} l_i(x_0) x_i^k (-x_0)^{j-k}]$$

$$\sum_{i=1}^{N} l_i(x_0)(x_i - x_0)^j = \sum_{k=0}^{j} \binom{j}{k} (-1)^{j-k} \sum_{i=1}^{N} l_i(x_0) x_i^j$$

For all $j > 0$, $\sum_{k=0}^{j} \binom{j}{k} (-1)^{j-k} = 0$, so

$$\sum_{i=1}^{N} l_i(x_0)(x_i - x_0)^j = 0 \sum_{i=1}^{N} l_i(x_0) x_i^j = 0$$

Therefore $b_j x_0 = 0$ for all $j > 0$.

4 Exercise 4

4.1 Part A

We have:

$$\int_a^b g"(x)h"(x)dx = [g"(x)h'(x)]_a^b - \int_a^b g'''(x)h'(x)dx$$

Since $g''(a) = g''(b) = 0$, we have $[g"(x)h'(x)]_a^b = 0$. For the remaining, $g'''(x) = 0$ for $x < x_1$ and $x > x_N$. The other cases could be defined as:

$$-\sum_{i=1}^{N-1} \int_{x_i}^{x_i+1} g'''(x)h'(x)dx = -\sum_{i=1}^{N-1} ([g'''(x)h(x)]_{x_i}^{x_i+1} - \int_{x_i}^{x_i+1} g''''(x)h(x)dx)$$

The first term is 0 since $h(x_i) = 0$ for all i, and similarly, $g''''(x) = 0$ for all x, so the second term is also 0. Therefore:

$$\int_a^b g"(x)h"(x)dx = 0$$

4.2 Part B

From Part A, we have:

$$\int_a^b \tilde{g}''(t)^2 dt = \int_a^b (g''(t) + h''(t))^2 dt$$

$$= \int_a^b g''(t)^2 dt + \int_a^b h''(t)^2 dt + 2\int_a^b g''(t)h''(t)dt$$

$$= \int_a^b g''(t)^2 dt + \int_a^b h''(t)^2 dt$$

Hence, $\int_a^b \tilde{g}''(t)^2 dt \geq \int_a^b g''(t)^2 dt$, and equality only hold if h remains zero in $[a, b]$.

4.3 Part C

For any f_1, we can construct a natural cubic spline f_2 with the same function values in knots x_i as f_1, and thus:

$$\sum_{i=1}^{N}(y - f_1(x_i))^2 = \sum_{i=1}^{N}(y - f_2(x_i))^2$$

For the second part, from B, we have:

$$\lambda \int_a^b f_1''(t)^2 dt \geq \lambda \int_a^b f_2''(t)^2 dt$$

Therefore, for f_1 to be minimal, we must have $f_1 = f_2$, which means f_1 is a cubic spline.