Assignment 5: Solutions

CS 750/850 Machine Learning

Problem 1 [33%]

It is mentioned in Chapter 7 of ISL that a cubic regression spline with one knot at ξ can be obtained using a basis of the form x, x^2, x^3, $[x-\xi]_+^3$, where $[x-\xi]_+^3 = (x-\xi)^3$ if $x > \xi$ and equals 0 otherwise. We will now show that a function of the form

$$f(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 [x-\xi]_+^3$$

is indeed a cubic regression spline, regardless of the values of β_0,β_1,β_2, β_3,β_4.

1. Find a cubic polynomial
$$f_1(x) = a_1 + b_1 x + c_1 x^2 + d_1 x^3$$
such that $f(x) = f_1(x)$ for all $x \leq \xi$. Express a_1,b_1,c_1,d_1 in terms of β_0,β_1,β_2,β_3,β_4.
2. Find a cubic polynomial
$$f_2(x) = a_2 + b_2 x + c_2 x^2 + d_2 x^3$$
such that $f(x) = f_2(x)$ for all $x > \xi$. Express a_2,b_2,c_2,d_2 in terms of β_0,β_1,β_2,β_3,β_4. We have now established that $f(x)$ is a piece-wise polynomial.
3. Show that $f_1(\xi) = f_2(\xi)$. That is, $f(x)$ is continuous at ξ.

Solution

1. The values are:

$$a_1 = \beta_0,\ b_1 = \beta_1,\ c_1 = \beta_2,\ d_1 = \beta_3$$

2. Recall that:

$$(x-\xi)^3 = x^3 - \xi^3 + 3x\xi^2 - 3x^2\xi$$

Then, the values are:

$$\begin{aligned} a_2 &= (\beta_0 - \beta_4 \xi^3), \\ b_2 &= (\beta_1 + 3\beta_4 \xi^2), \\ c_2 &= (\beta_2 - 3\beta_4 \xi), \\ d_2 &= \beta_3 + \beta_4 \end{aligned}$$

3. Follows by algebraic manipulation, or simply from the fact that $[x-\xi]^3 = 0$ when $x = \xi$.

It is mentioned in Chapter 7 of ISL that a cubic regression spline with one knot at ξ can be obtained using a basis of the form x, x^2, x^3, $[x-\xi]_+^3$, where $[x-\xi]_+^3 = (x-\xi)^3$ if $x > \xi$ and equals 0 otherwise. We will now show that a function of the form

$$f(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 [x-\xi]_+^3$$

is indeed a cubic regression spline, regardless of the values of β_0,β_1,β_2, β_3,β_4.

1. Find a cubic polynomial
$$f_1(x) = a_1 + b_1 x + c_1 x^2 + d_1 x^3$$
such that $f(x) = f_1(x)$ for all $x \leq \xi$. Express a_1,b_1,c_1,d_1 in terms of β_0,β_1,β_2,β_3,β_4.

2. Find a cubic polynomial

$$f_2(x) = a_2 + b_2 x + c_2 x^2 + d_2 x^3$$

such that $f(x) = f_2(x)$ for all $x > \xi$. Express a_2,b_2,c_2,d_2 in terms of β_0,β_1,β_2,β_3,β_4. We have now established that $f(x)$ is a piece-wise polynomial.

3. Show that $f_1(\xi) = f_2(\xi)$. That is, $f(x)$ is continuous at ξ.

Solution

1. The values are:

$$a_1 = \beta_0,\ b_1 = \beta_1,\ c_1 = \beta_2,\ d_1 = \beta_3$$

2. Recall that:

$$(x - \xi)^3 = x^3 - \xi^3 + 3x\xi^2 - 3x^2\xi$$

Then, the values are:

$$\begin{aligned} a_2 &= (\beta_0 - \beta_4\xi^3), \\ b_2 &= (\beta_1 + 3\beta_4\xi^2), \\ c_2 &= (\beta_2 - 3\beta_4\xi), \\ d_2 &= \beta_3 + \beta_4 \end{aligned}$$

3. Follows by algebraic manipulation, or simply from the fact that $[x - \xi]^3 = 0$ when $x = \xi$.

Problem 3

You will now derive the Bayesian connection to the ridge regression discussed in Section 6.2.2. of ISL.

1. Suppose that $y_i = \beta_0 + \sum_{j=1}^{p} x_{ij}\beta_j + \epsilon_i$ where $\epsilon_1, \ldots, \epsilon_n$ are independent and identically distributed from a normal distribution $N(0,1)$. Write out the likelihood for the data as a function of values β.

2. Assume the following prior for $\beta : \beta_1, \ldots, \beta_p$ are independent and identically distributed according to a normal distribution with mean zero and variance c. Write out the posterior for β in this setting using Bayes theorem.

3. Argue that the ridge regression estimate is the value of β with maximal probability under this posterior distribution. Compute log of the probability in order to make this point. *Hint: The denominator (= the probability of data) can be ignored in computing the maximum probability.*

Solution

1. The likelihood function is:

$$\frac{1}{(2\pi)^{n/2}} \prod_{i=1}^{n} e^{-\frac{\epsilon_i^2}{2}} = \frac{1}{(2\pi)^{n/2}} \prod_{i=1}^{n} e^{-\frac{(y_i - \beta_0 + \sum_{j=1}^{p} x_{ij}\beta_j)^2}{2}}$$

2. To simplify the notation, I will use y to denote the random variable representing data. The posterior is:

$$P(\beta|y) = \frac{P(y|\beta)P(\beta)}{P(y)} = \frac{1}{P(y)} \frac{1}{(2\pi)^{n/2}} \prod_{i=1}^{n} e^{-\frac{(y_i - \beta_0 + \sum_{j=1}^{p} x_{ij}\beta_j)^2}{2}} \frac{1}{(2\pi c)^{n/2}} \prod_{j=1}^{p} e^{-\frac{\beta_j^2}{2c}}$$

3. Take the log of the posterior probability to get:

$$-\frac{1}{2}\sum_{i=1}^{n}\left(y_i-\beta_0+\sum_{j=1}^{p}x_{ij}\beta_j\right)^2-\frac{1}{2c}\sum_{j=1}^{p}\beta_j^2+O(1)$$

Here, $O(1)$ denotes all the terms that do not depend on β. So to maximize this expression, we need to solve the following minimization problem:

$$\min_{\beta}\sum_{i=1}^{n}\left(y_i-\beta_0+\sum_{j=1}^{p}x_{ij}\beta_j\right)^2+\frac{1}{c}\sum_{j=1}^{p}\beta_j^2=\min_{\beta}\sum_{i=1}^{n}\mathrm{RSS}+\frac{1}{c}\sum_{j=1}^{p}\beta_j^2\ ,$$

which is the same as ridge regression.