STAT380: Assignment 2

Vivienne Crowe ID:40071153

4/2/21

1

1.1 A.4

(a) Let d denote the number of degree of each spline, k is the number of knots, and n is the number of constraints.

$$(d+1)(k+1) - nk$$

In this case, we constrain the 0-th, 1st and 2nd derivative, and so $n = 3$. The splines are cubic, so $d = 3$:

$$4(k+1) - 3k = k + 4$$

(b) A cubic spline is a continuous piece-wise function, whose first and second derivatives are also continuous.

1. Continuity
   $\lim_{x \to c_i^-} f(x) = \lim_{x \to c_i^+} f(x)$

2. Continuity of 1st derivative
   $\lim_{x \to c_i^-} f'(x) = \lim_{x \to c_i^+} f'(x)$

3. Continuity of 2nd derivative
   $\lim_{x \to c_i^-} f''(x) = \lim_{x \to c_i^+} f''(x)$

To show (1), first consider the boundary between c_1 and c_2:

$$\lim_{x \to c_1^-} f(x) = \sum_{j=0}^{4} \beta_j c_1^j + 0$$

$$\begin{aligned}
\lim_{x \to c_1^+} f(x) &= \sum_{j=0}^{4} \beta_j c_1^j + \lim_{x \to c_1^+} \beta_4 b_4(x) \\
&= \sum_{j=0}^{4} \beta_j c_1^j + 0
\end{aligned}$$

Now for the general case, between c_i and c_{i+1}:

$$\begin{aligned}
\lim_{x \to c_i^-} f(x) &= \sum_{j=0}^{4} \beta_j c_i^j + \sum_{j=4}^{i+2} \beta_j b_j(x) + \lim_{x \to c_i^-} \beta_{i+3} \cdot (x - c_i)_+^3 \\
&= \sum_{j=0}^{4} \beta_j c_i^j + \sum_{j=4}^{i+2} \beta_j b_j(x) \\
\lim_{x \to c_i^+} f(x) &= \sum_{j=0}^{4} \beta_j c_i^j + \sum_{j=4}^{i+2} \beta_j b_j(x) + \lim_{x \to c_i^+} \beta_{i+3} \cdot (x - c_i)_+^3 \\
&= \sum_{j=0}^{4} \beta_j c_i^j + \sum_{j=4}^{i+2} \beta_j b_j(x)
\end{aligned}$$

To show (2), first note $f'(x)$:

$$f'(x) = \sum_{j=1}^{4} \beta_j j x^{j-1} + \sum_{j=4}^{K+3} 3\beta_j (x - c_{j-3})_+^2$$

Now to check the continuity at $x = c_i$:

$$\begin{aligned}
\lim_{x \to c_i^-} f'(x) &= \sum_{j=1}^{4} \beta_j j c_i^{j-1} + \sum_{j=4}^{i+2} 3\beta_{i+3} (x - c_i)_+^2 \\
\lim_{x \to c_i^+} f'(x) &= \sum_{j=1}^{4} \beta_j j c_i^{j-1} + \sum_{j=4}^{i+2} 3\beta_{i+3} (x - c_i)_+^2 + \lim_{x \to c_i^+} 3\beta_{i+3} (x - c_i)_+^2 \\
&= \sum_{j=1}^{4} \beta_j j c_i^{j-1} + \sum_{j=4}^{i+2} 3\beta_{i+3} (x - c_i)_+^2
\end{aligned}$$

For (3)

$$f''(x) = \sum_{j=2}^{4} \beta_j j (j-1) x^{j-2} + \sum_{j=4}^{K+3} 6\beta_j (x - c_{j-3})_+$$

$$\begin{aligned}
\lim_{x \to c_i^-} f'(x) &= \sum_{j=1}^{4} \beta_j j c_i^{j-1} + \sum_{j=4}^{i+2} 6\beta_{i+3} (x - c_i)_+ \\
\lim_{x \to c_i^+} f'(x) &= \sum_{j=1}^{4} \beta_j j c_i^{j-1} + \sum_{j=4}^{i+2} 6\beta_{i+3} (x - c_i)_+ + \lim_{x \to c_i^+} 6\beta_{i+3} (x - c_i)_+ \\
&= \sum_{j=1}^{4} \beta_j j c_i^{j-1} + \sum_{j=4}^{i+2} 6\beta_{i+3} (x - c_i)_+
\end{aligned}$$

(c) Being about to express a cubic spline model in this way demonstrates that it is a type of linear model since we are able to express $f(x)$ as a linear combination of features. In this case, the features are a non-linear transformation of the original predictor variables.

1.2 A.5

First note that for ordinary least squares:

$$\hat{\beta} = (\mathbf{X^TX})^{-1}\mathbf{Xy}$$

Since $\mathbf{X}$ is square and full rank, it is therefore inveritiable, and so $(\mathbf{X^TX})^{-1} = \mathbf{X^{-1}}(\mathbf{X^T})^{-1}$. So it follows that:

$$\hat{\beta} = \mathbf{X^{-1}}$$

Now I'll show that $\mathbf{X^{-1}} = \mathbf{V\Sigma^{-1}U^T}$.

$$\begin{aligned}
\mathbf{X}(\mathbf{V\Sigma^{-1}U^T}) &= (\mathbf{V\Sigma^{-1}U^T})(\mathbf{U\Sigma V^T}) \\
&= (\mathbf{V\Sigma^{-1}\Sigma V^T}) \\
&= (\mathbf{VV^T}) \\
&= \mathbf{I}
\end{aligned}$$

Now onto the solution for ridge regression, the problem statement, in vector notation is:

$$\begin{aligned}
&\min_{\beta} \quad (\mathbf{y} - \mathbf{X}\beta)^{\mathbf{T}}(\mathbf{y} - \mathbf{X}\beta) - \lambda\beta^{\mathbf{T}}\beta \\
&\min_{\beta} \quad \mathbf{y^Ty} - \mathbf{2y^TX}\beta + \beta^{\mathbf{T}}\mathbf{X^TX}\beta) - \lambda\beta^{\mathbf{T}}\beta
\end{aligned}$$

Now to solve for where the derivative is equal to zero.

$$\begin{aligned}
0 &= \frac{\partial}{\partial\beta}\mathbf{y^Ty} - \mathbf{2y^TX}\beta + \beta^{\mathbf{T}}\mathbf{X^TX}\beta - \lambda\beta^{\mathbf{T}}\beta \\
0 &= -2\mathbf{y^TX} + (\mathbf{X^TX} + \mathbf{X^TX})\beta - \mathbf{2}\lambda\beta \\
\mathbf{y^TX} &= (\mathbf{X^TX} + \lambda\mathbf{I})\beta \\
\implies \hat{\beta} &= (\mathbf{X^TX} + \lambda\mathbf{I})^{-1}\mathbf{y^TX}
\end{aligned}$$