CMSE 820 HW3

This HW is due on Sep 29st at 11:59 pm.

Question 1: Consider the truncated power series representation for cubic splines with K interior knots. Let

$$f(X) = \sum_{j=0}^{3} \beta_j X^j + \sum_{k=1}^{K} \theta_k (X - \xi_k)_+^3,$$

where $X \in \mathbb{R}$.

1. Prove that the natural boundary conditions for natural cubic splines (linear when $X \leq \xi_1$ and $X \geq \xi_K$) imply the following linear constraints on the coefficients.

$$\beta_2 = 0; \quad \beta_3 = 0;$$
$$\sum_{k=1}^{K} \theta_k = 0; \quad \sum_{k=1}^{K} \xi_k \theta_k = 0;$$

2. Then derive one set of the basis functions for natural cubic spline can be

$$N_1(X) = 1; \quad N_2(X) = X,$$
$$N_{k+2}(X) = d_k(X) - d_{K-1}(X),$$

where

$$d_k(X) = \frac{(X - \xi_k)_+^3 - (X - \xi_K)_+^3}{\xi_K - \xi_k}.$$

3. Verify these basis functions satisfy the requirement of Natural Cubic spline.

Question 2: Suppose we have n sample pairs $\{x_i, y_i\}_1^n$, with $a < x_1 < \ldots < x_n < b$. The smoothing cubic spline estimate $\hat{f}$ is defined as

$$\hat{f} = \arg\min_f \sum_{i=1}^{n} \big(y_i - f(x_i)\big)^2 + \lambda \int_a^b \big(f''(t)\big)^2 dt.$$

In the class, we mentioned that it happens that the minimizer of above problem is unique and is a natural cubic spline with knots at the input point $x_1, \ldots, x_n$. Here we will prove it. First, we assume $\tilde{f}$ is any twice differentiable funciton on $[a, b]$.

1. Show that there exists a natural cubic spline f with knots at $x_1, \ldots, x_n$ (in the form of linear combination of those basis functions) such that $f(x_i) = \tilde{f}(x_i), i = 1, \ldots, n$.

2. Define
$$h(X) = \tilde{f}(X) - f(X).$$
Prove the following claim
$$\int_a^b f''(x)h''(x)dx = 0.$$
Hint: you may need to use integration by parts.

3. Now, we can show
$$\int_a^b f''(x)^2 dx \leq \int_a^b \tilde{f}''(x)^2 dx,$$
with equality if and only if $h''(x) = 0$ for all $x \in [a, b]$. Note that $h''(x) = 0$ implies that h must be linear, and since we already know that $h(x_i) = 0$ for all x_i, this is equivalent to $h(x) = 0$. And we finished the proof.

Question 3: Use the data Q3_X_train.csv and Q3_Y_train.csv to build a Ordinary Least squares, ridge regression and LASSO regression models. To select tuning parameters for Ridge and LASSO, please use 5-fold cross validation. Then evaluate the model performances using prediction errors with Q3_X_test.csv and Q3_Y_test.csv.

Question 4: The data Q4_X.csv and Q4_Y.csv are generated from
$$Y = \sin(X) + \epsilon.$$
Based on these two data, let's do the following

- Write down the one set of basis functions for cubic spline and natural cubic spline with knots at $\frac{\pi}{4}, \frac{\pi}{2}, \pi, \frac{3\pi}{2}, \frac{7\pi}{4}$. Plot those basis function together with the original data.
- Fit a cubic spline and a natural cubic spline models base on the two dataset. Plot the fitted model using 100 x's evenly spaced between $(-1, 7)$.

to fit a cubic spline and a natural cubic spline model with knots at $\frac{\pi}{4}, \frac{\pi}{2}, \pi, \frac{3\pi}{2}, \frac{7\pi}{4}$.

Question 5: Given a response vector $\mathbf{y} \in R^n$, predictor mvatrix $\mathbf{X} \in \mathbb{R}^{p \times n}$, and tuning parameter $\lambda > 0$, recall the ridge regression estimate (assume we have already center $\mathbf{y}$ and $\mathbf{X}$)
$$\hat{\beta}^{\text{ridge}} = \arg\min_{\beta} \|\mathbf{y} - \mathbf{X}^T \beta\|^2 + \lambda \|\beta\|^2,$$

a. Show that $\hat{\beta}^{\text{ridge}}$ is simply the vector of ordinary regression coefficients from regressing the response $\tilde{\mathbf{y}} = (\mathbf{y}^T, 0^T)^T \in \mathbb{R}^{n+p}$ onto the predictor matrix $\tilde{\mathbf{X}} = \begin{bmatrix} \mathbf{X} & \sqrt{\lambda} I \end{bmatrix} \in \mathbb{R}^{(n+p) \times p}$, where $0 \in \mathbb{R}^p$, and $I \in \mathbb{R}^{p \times p}$ is the identity matrix.

b. Show that the matrix $\tilde{\mathbf{X}}$ always has full row-rank, i.e, its rows are always linearly independent, regardless of the row of $\mathbf{X}$. Hence argue that the ridge regression estimate is always unique, for any matrix of predictors $\mathbf{X}$.

c. Write out an explicit formula for $\hat{\beta}^{\text{ridge}}$ involving $\mathbf{X}, \mathbf{y}, \lambda$. Conclude that for any $a^T \in \mathbb{R}^p$, the estimate $a^T \hat{\beta}^{\text{ridge}}$ is a linear function of $\mathbf{y}$.

d. Now consider the estimation of $a^T \hat{\beta}^{\text{ridge}}$, with β^* being the true coefficient vector. Based on what we've seen in lecture, ridge regression can have a lower MSE than linear regression. But we have proved that the ordinary linear regression estimate is the BLUE. Given that it is indeed linear (from part (c)), what does this imply about the ridge regression estimate, $a^T \hat{\beta}^{\text{ridge}}$?

e. Let $\mathbf{X}$ have singular value decomposition $\mathbf{X} = UDV^T$, where $U \in \mathbb{R}^{p\times r}, D \in \mathbb{R}^{r\times r}, V \in \mathbb{R}^{n\times r}, U, V$ have orthonormal columns, and D is diagonal with elements $d_1 \geq \ldots \geq dr \geq 0$. Rewrite your formula for the ridge regression solution $\hat{\beta}^{\text{ridge}}$ from (c) by replacing $\mathbf{X}$ with UDV^T , and simplifying the expression as much as possible.

f. Assume that

$$\mathbf{y} = \mathbf{X}^T \beta^* + \epsilon, \quad \text{with } \mathbb{E}[\epsilon] = 0, \quad \text{Cov}(\epsilon) = \sigma^2 I,$$

and let $a \in \mathbb{R}^p$. Prove that $a^T \hat{\beta}^{\text{ridge}}$ is indeed a biased estimate of $a^T \beta^*$, for any $\lambda > 0$.