STAT 5372 Assignment 5

(Lectures 17–21)

Mathematical concepts and derivations

1. Show that the linear weights $\{l_i(x_0)\}$ for local linear regression (Lecture 17, slide 13), satisfy (a)–(c) below. It suffices to show this for the case when $n = 2$. (Note that the kernel K_h is defined on Slide 3.)

   (a) $\sum_{i=1}^{n} l_i(x_0) = 1$, for all x_0.

   (b) $\sum_{i=1}^{n} l_i(x_0)(x - x_0) = 0$, for all x_0.

   (c) If $K_h(x_i, x_0) = 0$, then $l_i(x_0) = 0$.

2. (Lecture 18, slide 19). Suppose $\hat{\boldsymbol{f}} = \boldsymbol{L}\boldsymbol{y}$ is a linear smoother based on $\{(x_1, y_1), \ldots, (x_n, y_n)\}$, where $\hat{\boldsymbol{f}} = (\hat{f}_1, \ldots, \hat{f}_n)^T$, $\boldsymbol{y} = (y_1, \ldots, y_n)^T$, and the (i, j) element of the smoothing matrix $\boldsymbol{L}$ is $l_{ij} = l_j(x_i)$. Denote by $\hat{\boldsymbol{f}}^{(-i)} = \tilde{\boldsymbol{L}}\boldsymbol{y}$ the corresponding smooth that results by leaving out the i-th pair $\{(x_i, y_i)\}$. This implies that we have the expressions:

$$\hat{f}_i \equiv \hat{f}(x_i) = \sum_{j=1}^{n} l_{ij} y_j, \qquad \text{and} \qquad \hat{f}_i^{(-i)} \equiv \hat{f}^{(-i)}(x_i) = \sum_{i \neq j=1}^{n} \tilde{l}_{ij} y_j.$$

   (a) If the smoothing matrix satisfies the (rows sum to unity) property in Question 1(a), deduce that $\tilde{l}_{ij} = l_{ij}/(1 - l_{ii})$. (Assume smoothing matrices are unique to the problem at hand.)

   (b) Show that the CV score (defined in the first equality below) can be written in the simpler form given by the second equality:

$$CV = \frac{1}{n} \sum_{i=1}^{n} \left\{ y_i - \hat{f}_{(-i)}(x_i) \right\}^2 = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{y_i - \hat{f}(x_i)}{1 - l_{ii}} \right)^2.$$

3. Show that the Nadaraya-Watson kernel estimator (Lecture 18, Slide 3), is a special case of Loess (Lecture 18, Slide 12) when the local polynomial is a constant. That is, show that the solution of:

$$\hat{\alpha} = \arg\min_{\alpha \in \mathbb{R}} \sum_i K_h(x_0, x_i)(y_i - \alpha)^2, \qquad \text{implies that} \qquad \hat{f}(x_0) = \frac{\sum_i K_h(x_0, x_i) y_i}{\sum_i K_h(x_0, x_i)}.$$

4. (Lecture 21). For $x \in [0,1]$ and equally spaced knots, consider two sets of spline basis functions: the truncated spline basis functions (Slide 14), and the B-spline basis functions (Slide 19). (Note: we did not cover the construction of B-splines, which are more complicated, but they can be plotted with the `splines` package.)

   (a) Write down the set of truncated spline basis functions for representing a cubic spline ($m = 4$) with three knots ($K = 3$). Plot these basis functions (exclude the intercept term).

   (b) For the same knots as in (a), plot the B-spline basis functions (again, exclude the intercept).

5. Consider the smoothing spline problem defined in Lecture 21 (Slides 34-36): $f(x) = \sum_{j=1}^{n} N_j(x)\beta_j$, $f(\boldsymbol{x}) = N\boldsymbol{\beta}$, with $N_{ij} = N_j(x_i)$ the (i,j) element of N, and where the vector $\boldsymbol{\beta}$ minimizes the objective function:

$$S(\boldsymbol{\beta}) = \sum_{i=1}^{n}(y_i - f(x_i))^2 + \lambda \int f''(u)^2 du.$$

   (a) If Ω is defined as the matrix whose (j,k) element is $\Omega_{jk} = \int N_j''(u)N_k''(u)du$, show that the objective function can be written as:

$$S(\boldsymbol{\beta}) = (\boldsymbol{y} - N\boldsymbol{\beta})^T(\boldsymbol{y} - N\boldsymbol{\beta}) + \lambda\boldsymbol{\beta}^T\Omega\boldsymbol{\beta}.$$

   (b) Derive the fact that the solution to this objective function is as follows:

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} S(\boldsymbol{\beta}) = (N^TN + \lambda\Omega)^{-1}N^T\boldsymbol{y}.$$

Simulation

Some questions in this section will have you execute the following algorithm.

Algorithm 1. *The model for the pairs (x_i, y_i), $i = 1, \ldots, n$, with $n = 101$, is:*

$$y_i = f(x_i) + \epsilon_i, \qquad x_i = 6\left(\frac{i-1}{100}\right) - 3, \qquad \{\epsilon_i\} \sim iid\ N(0,1),$$

for a specified ***`function`*** *$f(x)$. Note that the design points x_i are uniformly spaced in the range $[-3,3]$. Using a specified* ***`method`****, obtain fitted values $\hat{y}_1, \ldots, \hat{y}_n$. Then proceed as follows.*

(i) Simulate $M = 1000$ realizations from the model. From realization m, obtain the fitted values $\hat{y}_{1,m}, \ldots, \hat{y}_{n,m}$, corresponding to the observed values $y_{1,m}, \ldots, y_{n,m}$.

(ii) For $i = 1, \ldots, n$, let

$$b_i = \frac{1}{M}\sum_{m=1}^{M}\hat{y}_{i,m} - f(x_i), \qquad v_i^2 = \frac{1}{M}\sum_{m=1}^{M}\hat{y}_{i,m}^2 - \left[\frac{1}{M}\sum_{m=1}^{M}\hat{y}_{i,m}\right]^2,$$

be the (estimated) bias and variance at x_i. Similarly, let $MSE_i = b_i^2 + v_i^2$ to be the corresponding MSE at x_i.

(iii) Plot (x_i, b_i), (x_i, v_i^2), and (x_i, MSE_i), for $i = 1, \ldots, n$.

6. Apply Algorithm 1 to produce the plots of bias, variance and MSE, for the following cases.

   (a) The `function` is $f(x) = -x^2$, and the fitting `method` is polynomial regression with linear and quadratic terms, i.e., fit the model:

   $$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon.$$

   (b) The fitting `method` is polynomial regression with up to 5th order terms, but the `function` is:

   $$f(x) = \begin{cases} 0, & \text{if } x \leq 0 \\ x^2, & \text{if } x > 0. \end{cases}$$

7. Apply Algorithm 1 to produce the plots of bias, variance and MSE, for the following cases.

   (a) The `function` is $f(x) = -x^2$, and the fitting `method` is *loess* (locally weighted linear regression).

   (b) The fitting `method` is also *loess* (locally weighted linear regression), but the `function` is:

   $$f(x) = \begin{cases} 0, & \text{if } x \leq 0 \\ x^2, & \text{if } x > 0. \end{cases}$$

8. Apply Algorithm 1 to produce the plots of bias, variance and MSE, for the following cases.

   (a) The `function` is $f(x) = -x^2$, and the fitting `method` is *smoothing splines.*

   (b) The fitting `method` is also *smoothing splines*, but the `function` is:

   $$f(x) = \begin{cases} 0, & \text{if } x \leq 0 \\ x^2, & \text{if } x > 0. \end{cases}$$

Application

9. The course website contains the dataset `motorcycle.txt` from an experiment on the efficacy of crash helmets. The response (Acceleration) is measured by an accelerometer placed inside a helmet. Use *loess* to model acceleration over time, as follows.

   (a) Choose a criterion by which to select the smoothing parameter. Plot this criterion versus the smoothing parameter and choose the optimal value for use in (b) and (c).

   (b) Plot a smooth curve estimating acceleration as a function of time, and include bands to indicate confidence regions.

   (c) Prepare an ANOVA table testing the sequence of models:

   $$\text{Null (intercept only)} \subset \text{Linear} \subset \text{Quadratic.}$$

10. The course website contains the dataset `asthma.txt` from a study of the relationship between childhood asthma and exposure to air pollution from concentrated animal feeding operations. Treat asthma (Yes/No) as following a binomial distribution given exposure, and plot the probability of developing asthma, using the methods outlined below. Present the 4 plots in (a)–(d) together in one figure (2 panels by 2 panels).

   (a) Using the method of *kernel density* (naive Bayes classifier), plot a smooth curve estimating the relationship between exposure and the probability of developing asthma.

   (b) Using the method of *GLM*, plot a smooth curve estimating the relationship between exposure and the probability of developing asthma, with (simultaneous) 95% confidence bands.

   (c) Using the method of *smoothing splines*, plot a smooth curve estimating the relationship between exposure and the probability of developing asthma, with (simultaneous) 95% confidence bands.

   (d) Using the method of *local likelihood*, plot a smooth curve estimating the relationship between exposure and the probability of developing asthma, with (simultaneous) 95% confidence bands.

   (e) In the context of the method of *local likelihood*, prepare an ANOVA table testing the sequence of models:

$$\text{Null (intercept only)} \subset \text{Linear} \subset \text{Local},$$

   and state your conclusion. (Note: if $\theta(x) = \alpha + \beta x$ is the linear predictor, the "Null" model should be a GLM with $\theta(x) = \alpha$, the "Linear" model should be a GLM with $\theta(x) = \alpha + \beta x$, and the the "Local" model should be a local likelihood model also with $\theta(x) = \alpha + \beta x$.)