

DS-GA 3001.009 Applied Statistics: Homework #8

Due on Thursday, December 14, 2023

Please hand in your homework via Gradescope (entry code: RKXJN2) before 11:59 PM.

1. Find the natural cubic spline $f(x)$ with $f(0) = 0$, $f(1) = 2$, and $f(2) = 3$. Specifically, you should specify the coefficients $(a_0, a_1, a_2, a_3, b_0, b_1, b_2, b_3)$ such that

$$f(x) = \begin{cases} g(x) = a_3x^3 + a_2x^2 + a_1x + a_0 & \text{if } x \in [0, 1], \\ h(x) = b_3(x-2)^3 + b_2(x-2)^2 + b_1(x-2) + b_0 & \text{if } x \in [1, 2], \end{cases}$$

where $g(1) = h(1)$, $g'(1) = h'(1)$, $g''(1) = h''(1)$, and $f''(0) = f''(2) = 0$ (boundary conditions for the natural spline). Provide details of how you arrive at your answer.

2. In wavelet shrinkage, the soft and hard thresholding estimator aim to mimic the *ideal truncated estimator*. Consider a one-dimensional Gaussian location model $y \sim \mathcal{N}(\theta, \sigma^2)$ with known σ ; there is an (unknown) upper bound τ of $|\theta|$, i.e. $|\theta| \leq \tau$.

- (a) For the MLE $\hat{\theta}_1(y) = y$, compute the worst-case MSE $\max_{|\theta| \leq \tau} \mathbb{E}_\theta[(\hat{\theta}_1(y) - \theta)^2]$.
- (b) For the zero estimator $\hat{\theta}_2(y) \equiv 0$, compute the worst-case MSE $\max_{|\theta| \leq \tau} \mathbb{E}_\theta[(\hat{\theta}_2(y) - \theta)^2]$.
- (c) The ideal truncated estimator assumes that θ is known, but forces the learner to use either $\hat{\theta}_1(y)$ or $\hat{\theta}_2(y)$. In other words, the learner finds a subset $R = R(\sigma) \subseteq \mathbb{R}$ based on the knowledge of σ , and uses

$$\hat{\theta}(y) = \begin{cases} \hat{\theta}_1(y) & \text{if } \theta \in R, \\ \hat{\theta}_2(y) & \text{if } \theta \notin R. \end{cases}$$

Which choice of R minimizes the worst-case MSE $\max_{|\theta| \leq \tau} \mathbb{E}_\theta[(\hat{\theta}(y) - \theta)^2]$? The resulting estimator is known as the ideal truncated estimator. What is the worst-case MSE for the ideal truncated estimator?

3. In class we have shown that the local polynomial fit $\hat{f}_k(x_0)$ of degree k for $f(x_0)$ takes the form $\hat{f}_k(x_0) = \sum_{i=1}^n w_k(x_0, x_i) y_i$. In this problem we aim to show that

$$\sum_{i=1}^n w_k(x_0, x_i)^2 \leq \sum_{i=1}^n w_{k+1}(x_0, x_i)^2,$$

and therefore the variance of the fit becomes larger when one increases the polynomial degree. The proof relies on the following result in linear algebra:

Lemma 1. For any positive definite matrix M with a block-wise form

$$M = \begin{bmatrix} A & B \\ B^\top & C \end{bmatrix},$$

where A and C are symmetric square matrices, the matrix

$$M^{-1} - \begin{bmatrix} A^{-1} & \\ & O \end{bmatrix}$$

is positive semi-definite, where O is the all-zero matrix.

- (a) Use the lemma and the matrix-form expression of $w_k(x_0, x_i)$ derived in class, prove that for the box kernel $K(x) = \mathbb{1}(|x| \leq 1/2)$ and any bandwidth parameter $h > 0$, it holds that $\sum_{i=1}^n w_k(x_0, x_i)^2 \leq \sum_{i=1}^n w_{k+1}(x_0, x_i)^2$.
 - (b) (*Bonus 5 points*) Prove the lemma.
4. Coding: we will use the Doppler example in “Donoho, D.L. and Johnstone, I.M. (1994) Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81, 425–455” to visualize and test the performances of the nonparametric estimators we learned in class. In this problem we will implement the following estimators:
- (a) Nadaraya–Watson estimator, with a data-driven bandwidth;
 - (b) local polynomial regressors, with $d \in \{1, 20\}$;
 - (c) cubic smoothing and regression splines;
 - (d) Fourier projection estimator;
 - (e) wavelet (soft and hard) thresholding estimators.

Based on inline instructions, fill in the missing codes in <https://tinyurl.com/2aesjazzk>. Be sure to submit a pdf with your codes, outputs, and colab link.