Homework 1 Part 2

Stat 435, Spring 2020
Due Friday, April 10, 11:59pm

Kernel smoothing: a simplified analysis of estimation error

Training sample $(x_1, y_1, \ldots, (x_n, y_n)$.
Predictor values $x_1, \ldots, x_n$ are considered fixed (not random).
Response values $y_i = f(x_i) + \epsilon_i$, with $\epsilon_1, \ldots, \epsilon_n$ independent, $\mathbf{E}(\epsilon_i) = 0$, $\mathbf{V}(\epsilon_i) = \sigma^2$.

Define $\mathbf{y} = (y_1, \ldots, y_n)$, $\mathbf{f} = (f(x_1), \ldots, f(x_n)$, etc.

$K_\lambda(x)$: Kernel with bandwidth parameter λ controlling the amount of smoothing.
Define $\mathbf{W}$: $n \times n$ weight matrix with $\mathbf{W}_{i,j} = K_\lambda(x_i - x_j)/\sum_j K_\lambda(x_i - x_j)$.
Then $\hat{\mathbf{f}} = \mathbf{W}\,\mathbf{y}$.
Note: $\mathbf{W}$, $\hat{\mathbf{f}}$, etc will depend on λ. We supress this dependence to simplify notation.

We want to calculate the expectation (over ϵ) of the average squared estimation error:

$$D = \mathbf{E}\left(\frac{1}{n}\|\mathbf{f} - \hat{\mathbf{f}}\|^2.\right)$$

(**a**) (20 points) Show that

$$D = \frac{1}{n}\left(\|(W - I)\,\mathbf{f}\|^2 + \sigma^2\,\mathbf{trace}\ (\mathbf{W}^T\mathbf{W})\right)$$

Note: D is the sum of a bias term that depends only on f (not on σ^2 and a variance term that depends only on σ^2 (not of f.

(b) (20 points) Using a Gaussian kernel ϕ_σ, plot squared bias, variance, and their sum for $\sigma =$ `seq(from = 0.01, to = 2, by = 0.01)`. Print the optimal choice for σ.

(c) (10 points) Plot the training sample, $\mathbf{f}$, and $\hat{\mathbf{f}}$ for the optimal choice of σ.