DS 100: Principles and Techniques of Data Science **Date: March 16, 2018**

Discussion #7

Name:

Bias-Variance Tradeoff

1. Let X be a random variable with mean $\mu = \mathbb{E}[X]$. Using the definition $\text{Var}(X) = \mathbb{E}[(X-\mu)^2]$, show that for any constant c,

$$\mathbb{E}[(X-c)^2] = (\mu - c)^2 + \text{Var}(X).$$

2. In the context of question 1, conclude that

- $\text{Var}(X) \leq \mathbb{E}[(X-c)^2]$ for any c
- $\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$

3. Suppose we make **independent** observations $X_1, \ldots, X_n$ with a common density $f(x)$, and we construct a KDE to estimate the density:

$$\widehat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} K_h(x - X_i),$$

where $K_h(y) = K(y/h)/h$.

(a) Write the bias-variance decomposition for the $\mathbb{L}_2$-error $\mathbb{E}[(\widehat{f}(x) - f(x))^2]$ at a point x.

(b) What happens to each term as the number of samples n increases?

(c) What happens to each term as the bandwidth h approaches 0 or ∞?

4. Recall that we can break down squared error into Noise, Bias and Variance:

$$\mathbb{E}\left((y - f(x))^2\right) = \mathbb{E}\left[(y - h(x))^2\right] + (h(x) - \mathbb{E}(f(x)))^2 + \mathbb{E}\left[(\mathbb{E}(f(x)) - f(x))^2\right]$$

   where $y = h(x) + \epsilon$, $\mathbb{E}(\epsilon) = 0$, $\text{Var}(\epsilon) = \sigma^2$

   As we increase model complexity, how are these terms affected? Draw a graph showing how variance, bias and test error change as model complexity increases.

Regularization

5. In a petri dish, yeast populations grow exponentially over time. In order to estimate the growth rate of a certain yeast, you place yeast cells in each of n petri dishes and observe the population y_i at time x_i and collect a dataset $\{(x_1, y_1), \ldots, (x_n, y_n)\}$. Because yeast populations are known to grow exponentially, you propose the following model:

$$\log(y_i) = \beta x_i \tag{1}$$

where β is the growth rate parameter (which you are trying to estimate). We will derive the L_2 regularized estimator least squares estimate.

   (a) Write the *regularized least squares loss function* for β under this model. Use λ as the regularization parameter.

   (b) Solve for the optimal $\widehat{\beta}$ as a function of the data and λ.