4 Bias-Variance Dilemma

Exercise 4.1

While you fit a Linear Model to your data set. You are thinking about changing the Linear Model to a Quadratic one (i.e., a Linear Model with quadratic features $\phi(x) = [1,\ x,\ x^2]$). Which of the following is most likely true:

1. Using the Quadratic Model will decrease your Irreducible Error;
2. Using the Quadratic Model will decrease the Bias of your model;
3. Using the Quadratic Model will decrease the Variance of your model;
4. Using the Quadratic Model will decrease your Reducible Error.

Provide motivations to your answers.

Exercise 4.2

Which of the following is/are the benefits of the sparsity imposed by the Lasso?

1. Sparse models are generally more easy to interpret;
2. The Lasso does variable selection by default;
3. Using the Lasso penalty helps to decrease the bias of the fits;
4. Using the Lasso penalty helps to decrease the variance of the fits.

Provide motivation for your answer.

Exercise 4.3

We estimate the regression coefficients in a linear regression model by minimizing ridge regression for a particular value of λ. For each of the following, describe the behaviour of the following elements as we increase λ from 0 (e.g., remains constant, increases, decreases, increase and then decrease):

1. The training RSS;

2. The test RSS;
3. The variance;
4. The squared bias;
5. The irreducible error.

Exercise 4.4

Suppose that Figure 4.1 is showing the Test error of K-NN obtained by using different values for K.

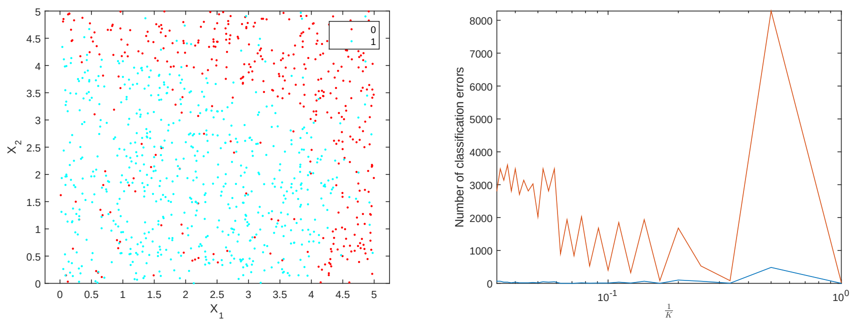

Figure 4.1: Dataset and corresponding error for different K in the K-NN classifier.

Which of the following is most likely true of what would happen to the Test Error curve as we move $\frac{1}{K}$ further above 1?

1. The Test Errors will increase;
2. The Test Errors will decrease;
3. Not enough information is given to decide;
4. It does not make sense to have $\frac{1}{K} > 1$;

Exercise 4.5

Comment on advantages and drawbacks of the following choices:

1. Increase the model complexity and fix number of samples;
2. Increase the number of the samples and fix model complexity.

Exercise 4.6

Assume to have two different linear models working on the same dataset of $N = 100$ samples.

- The first model has $k_1 = 2$ input, considers linear features and has a residual sum of squares of $RSS_1 = 0.5$ on a validation set;
- The second model has $k_2 = 8$ input, considers only quadratic features and has a residual sum of squares of $RSS_2 = 0.3$ on a validation set;

Which one would you choose? Why? Recall that the F-test for statistics for distinguish between linear models is:

$$\hat{F} = \frac{N - p_2}{p_2 - p_1} \frac{RSS_1 - RSS_2}{RSS_2} \sim F(p_2 - p_1, N - p_2),$$

where p_1 and p_2 are the two parameters of the two models and $F(a, b)$ is the Fisher distribution with a and b degrees of freedom.

Exercise 4.7

Which techniques would you consider for evaluate the performance of a set of different models in the case we have:

1. A small dataset and a set of simple models;
2. A small dataset and a set of complex models;
3. A large dataset and a set of simple models;
4. A large dataset and a trainer with parallel computing abilities.

Justify you choices.

Exercise 4.8

Suppose you have a dataset and you decided to use all the samples to train your model, including the selection of the parameters of your model and the features you want to consider. What are the problems and issues arising if you use this methodology?

Which procedure a ML scientist should follow?