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

Figure 4.1 is showing the the training data used to train a K-NN classifier (left) and the training (blue) and test (orange) performances obtained by using different values for K (right).

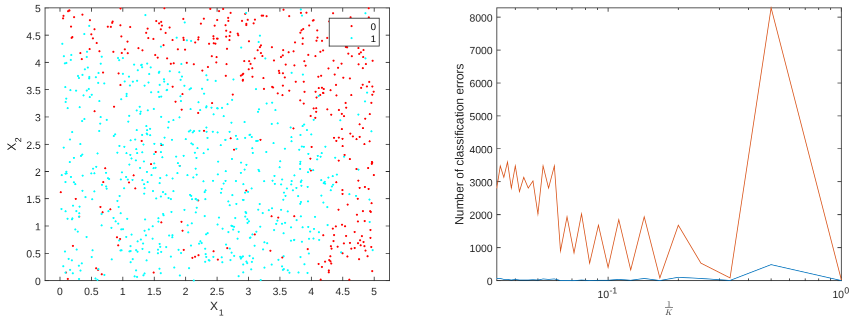

Figure 4.1: Dataset and corresponding error for different K in the K-NN classifier.

Which of the following would most likely 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 the following:

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

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 to evaluate the performances 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.

1. What are the problems and issues arising if you use this methodology?
2. Which procedure a ML scientist should follow?

Exercise 4.9

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

1. Using the Lasso penalty increases the variance of the fits.
2. It does variable selection implicitly.
3. Using the Lasso penalty increases the bias of the fits.
4. Sparse models are generally easier to interpret.

Provide motivation for your answer.

Exercise 4.10

Consider the following statements and tell if they are true or false. Motivate your answers.

1. The computation of the bias-variance decomposition is possible only theoretically. No algorithm provides an explicit decomposition of the twos.
2. An error which is comparable on the training and the test, but larger than what is required by the application, means that the used method has a large variance.
3. The cross-validation error provides slightly larger estimates of the prediction error on newly seen data.
4. If a model results in being too complex, to solve the problem we need to carefully remove some of the input features.

Exercise 4.11

Tell if the following statements about the bias-variance dilemma (and related topics) are true or false. Motivate your answers.

1. If the performance on the training and the test sets are getting almost the same as we are using more and more data for training, but both are worse than desired, the model might be too simple for the task.
2. Given a fixed training set, increasing the complexity of the model always improves its generalization capabilities.
3. It is a good idea to increase the number of samples used for training if we decided to increase the model complexity.
4. Adding new features to the model might help if the model has a bias with respect to the real process generating data.
5. The use of cross-validation decreases the variance of the model.

6. The use of the error on a validation set is suggested when we have a large dataset and we want to discriminate the performance of a set of (computationally) simple models.

Exercise 4.12

Tell if the following statements about the bias-variance dilemma (and related topics) are true or false. Motivate your answers.

1. Regularization techniques are likely to increase the bias of a model.
2. If the performance on the training is matching the desired performance, while the one on a test set is not satisfactory the model might be too complex for the task.
3. Increasing the training set size always improves the model performance.
4. It is a good idea to increase the number of samples used for training if we decided to increase the model complexity.
5. The use of the error on a validation set is suggested when we have a small dataset and we want to discriminate the performance of a set of computationally expensive models.

Exercise 4.13

Tell whether the following statements about training a supervised learning model are true or false. Motivate your answers.

1. The availability of a large dataset might lead to choosing a more complex model.
2. When data is very noisy, it is not a good idea to employ regularization.
3. The larger is the training set, the smaller would be the training error.
4. Increasing the number of features would generally lead to a decrease of the training error.

Exercise 4.14

Tell if the following statements about bias-variance trade-off are true or false. Motivate your answers. Consider a regression problem with input variables x_1, x_2, and x_3, that are linearly independent.

1. In linear regression, if we replace variable x_1 with $x_1 + x_2$, we do not change the bias of the model.

2. In linear regression, if we replace variable x_1 with $x_1^2/100$, we might increase the variance of the model.

3. For an arbitrary model, if we remove variable x_2, we do not increase the variance of the model.

4. For an arbitrary model, if we add variable x_3^2, we cannot increase the bias of the model.