ASSESSING MODEL ACCURACY

Chapter 02 – Part II

Outline

- Assessing Model Accuracy
  - Measuring the Quality of Fit
  - The Bias-Variance Trade-off
  - The Classification Setting

Measuring Quality of Fit

- Suppose we have a regression problem.
- One common measure of accuracy is the mean squared error (MSE) i.e.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

- Where $\hat{y}_i$ is the prediction our method gives for the observation in our training data.

A Problem

- In either case our method has generally been designed to make MSE small on the training data we are looking at e.g. with linear regression we choose the line such that MSE is minimized.

- What we really care about is how well the method works on new data. We call this new data "**Test Data**".

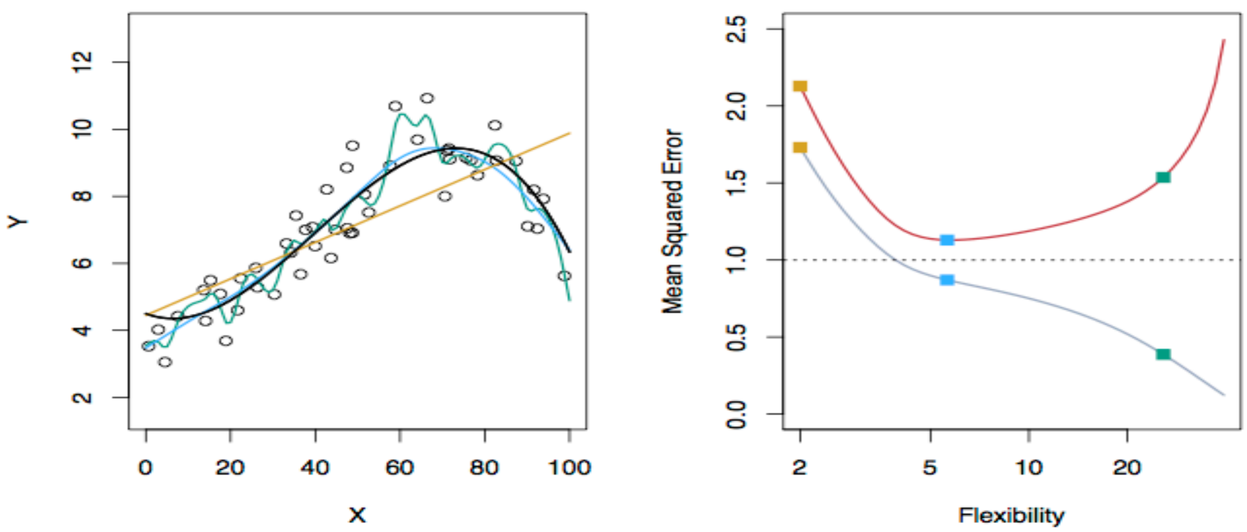
- There is no guarantee that the method with the smallest training MSE will have the smallest test (i.e. new data) MSE.

Training vs. Test MSE's

- In general the more flexible a method is the lower its training MSE will be i.e. it will "fit" or explain the training data very well.
  - Side Note: More Flexible methods (such as splines) can generate a wider range of possible shapes to estimate f as compared to less flexible and more restrictive methods (such as linear regression). The less flexible the method, the easier to interpret the model. Thus, there is a trade-off between flexibility and model interpretability.

- However, the test MSE may in fact be higher for a more flexible method than for a simple approach like linear regression.

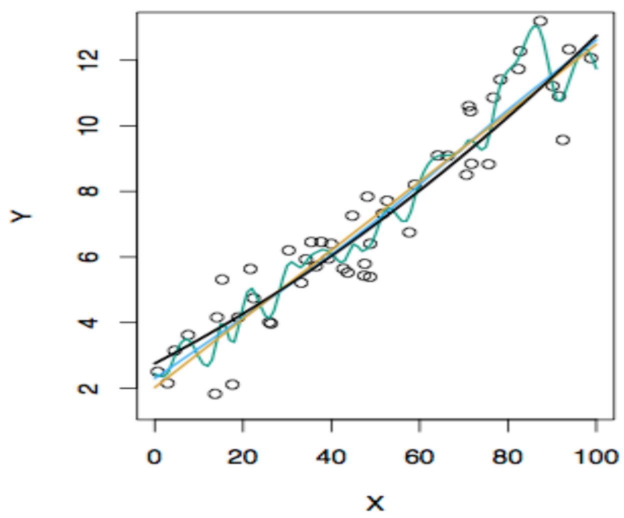
Examples with Different Levels of Flexibility: Example 1

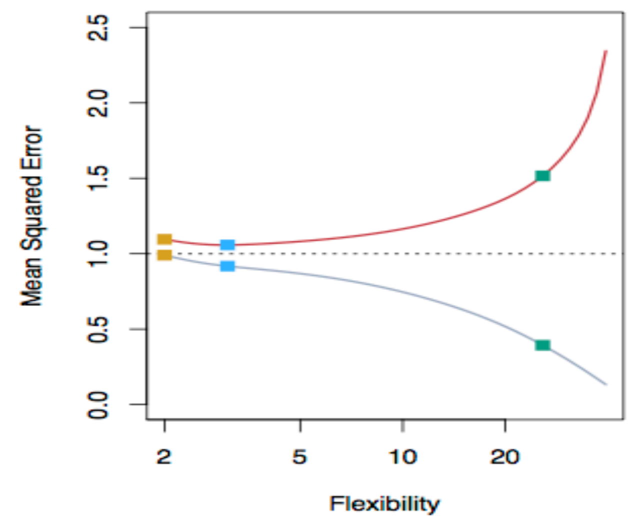


<u>LEFT</u>
Black: Truth
Orange: Linear Estimate
Blue: smoothing spline
Green: smoothing spline (more flexible)

<u>RIGHT</u>
RED: Test MES
Grey: Training MSE
Dashed: Minimum possible test MSE (irreducible error)

Examples with Different Levels of Flexibility: Example 2

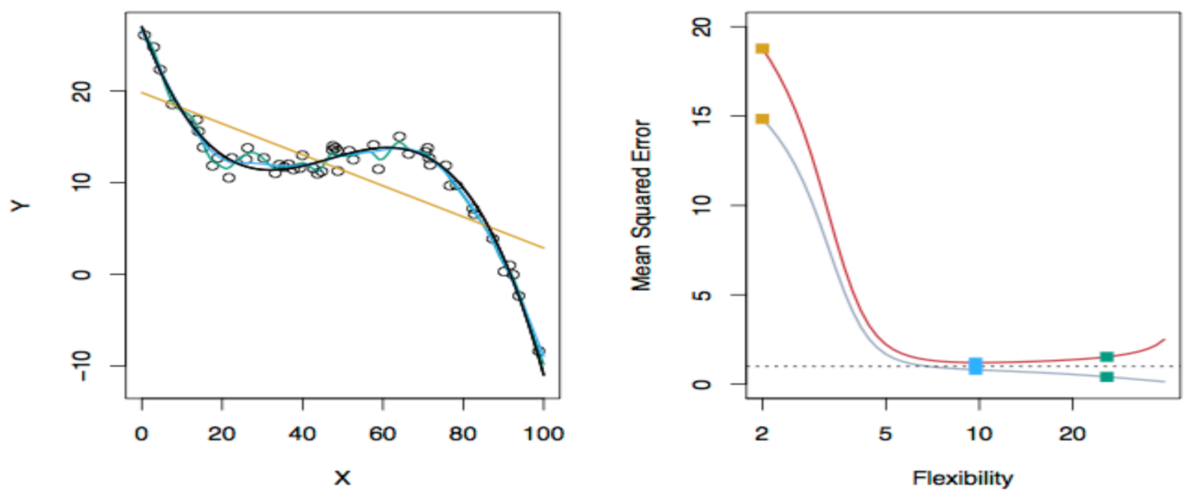




LEFT
Black: Truth
Orange: Linear Estimate
Blue: smoothing spline
Green: smoothing spline (more flexible)

RIGHT
RED: Test MES
Grey: Training MSE
Dashed: Minimum possible test MSE (irreducible error)

Examples with Different Levels of Flexibility: Example 3



<u>LEFT</u>
Black: Truth
Orange: Linear Estimate
Blue: smoothing spline
Green: smoothing spline (more flexible)

<u>RIGHT</u>
RED: Test MES
Grey: Training MSE
Dashed: Minimum possible test MSE (irreducible error)

Bias/ Variance Tradeoff

- The previous graphs of test versus training MSE's illustrates a very important tradeoff that governs the choice of statistical learning methods.

- There are always two competing forces that govern the choice of learning method i.e. bias and variance.

Bias of Learning Methods

- Bias refers to the error that is introduced by modeling a real life problem (that is usually extremely complicated) by a much simpler problem.

- For example, linear regression assumes that there is a linear relationship between Y and X. It is unlikely that, in real life, the relationship is exactly linear so some bias will be present.

- The more flexible/complex a method is the less bias it will generally have.

Variance of Learning Methods

- Variance refers to how much your estimate for f would change by if you had a different training data set.

- Generally, the more flexible a method is the more variance it has.

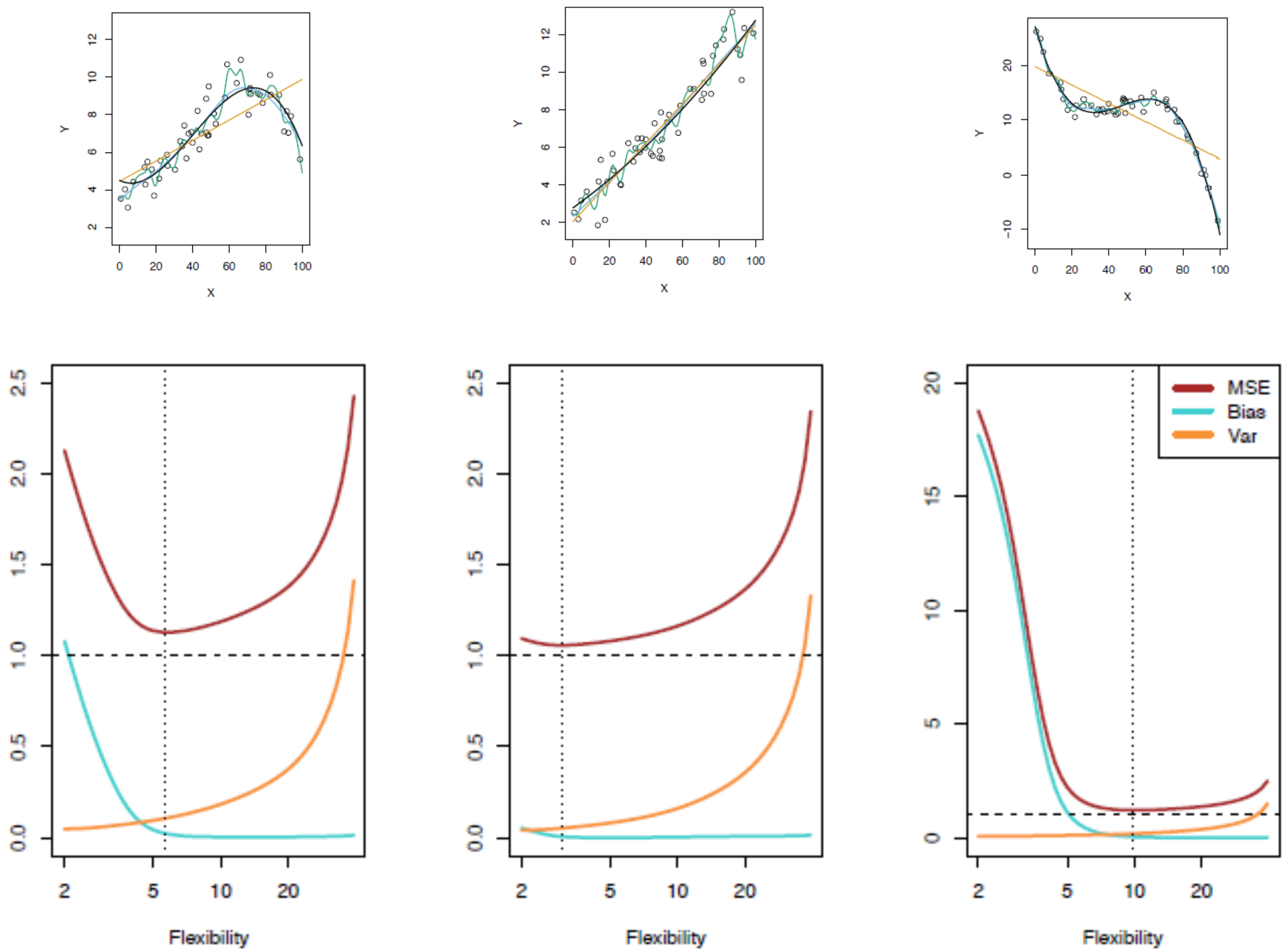
The Trade-off

- It can be shown that for any given, $X=x_0$, the expected test MSE for a new Y at x_0 will be equal to

$$Expected\,Test\,MSE = E\left(Y - f(x_0)\right)^2 = Bias^2 + Var + \underbrace{\sigma^2}_{\text{Irreducible Error}}$$

- What this means is that as a method gets more complex the bias will decrease and the variance will increase but expected test MSE may go up or down!

Test MSE, Bias and Variance



The Classification Setting

- For a regression problem, we used the MSE to assess the accuracy of the statistical learning method
- For a classification problem we can use the error rate i.e.

$$Error\ Rate = \sum_{i=1}^{n} I(y_i \neq \hat{y}_i) / n$$

- $I(y_i \neq \hat{y}_i)$ is an indicator function, which will give 1 if the condition $(y_i \neq \hat{y}_i)$ is correct, otherwise it gives a 0.

- Thus the error rate represents the fraction of incorrect classifications, or misclassifications

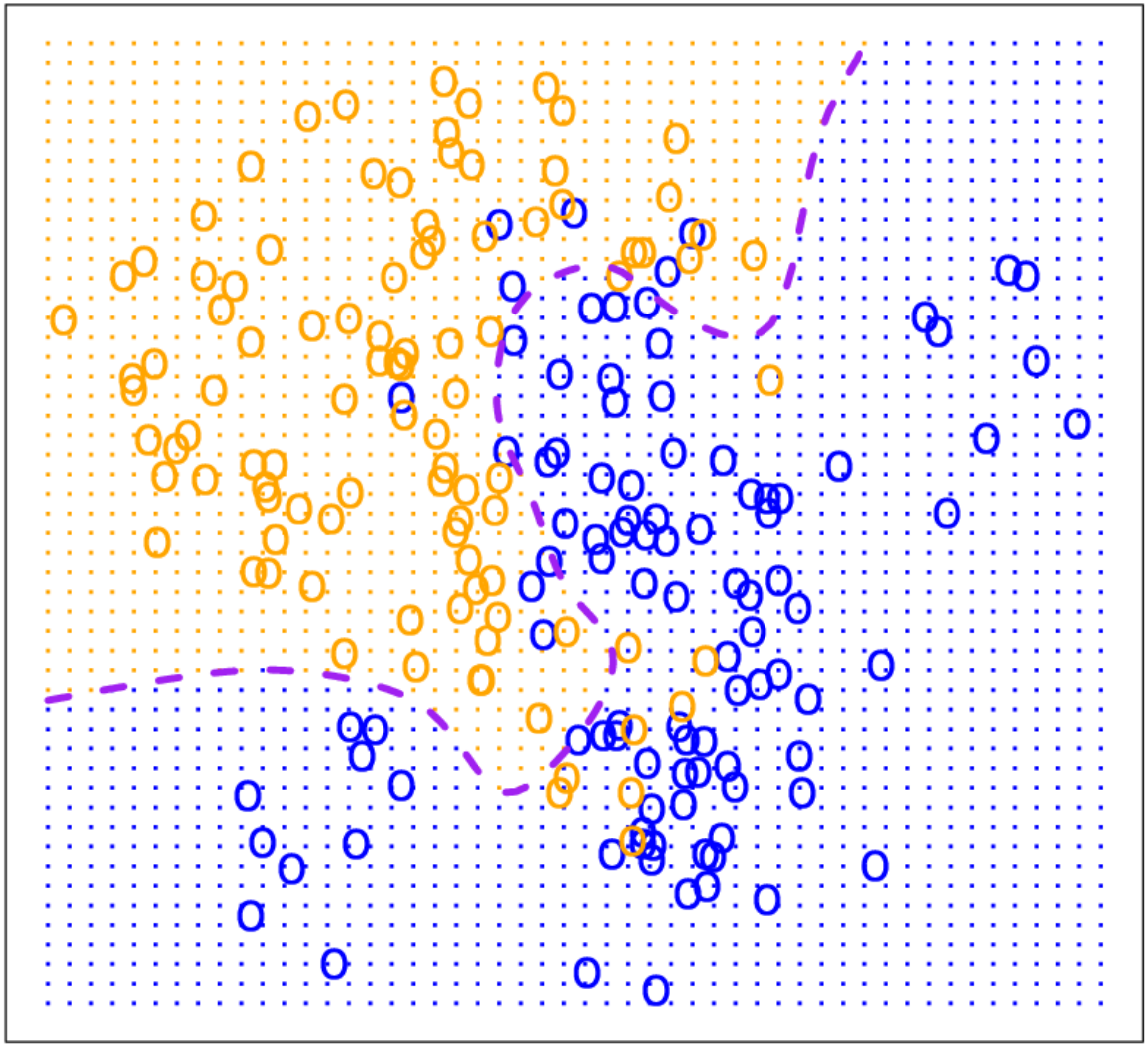
Bayes Error Rate

- The Bayes error rate refers to the lowest possible error rate that could be achieved if somehow we knew exactly what the “true” probability distribution of the data looked like.

- On test data, no classifier (or stat. learning method) can get lower error rates than the Bayes error rate.

- Of course in real life problems the Bayes error rate can’t be calculated exactly.

Bayes Optimal Classifier



K-Nearest Neighbors (KNN)

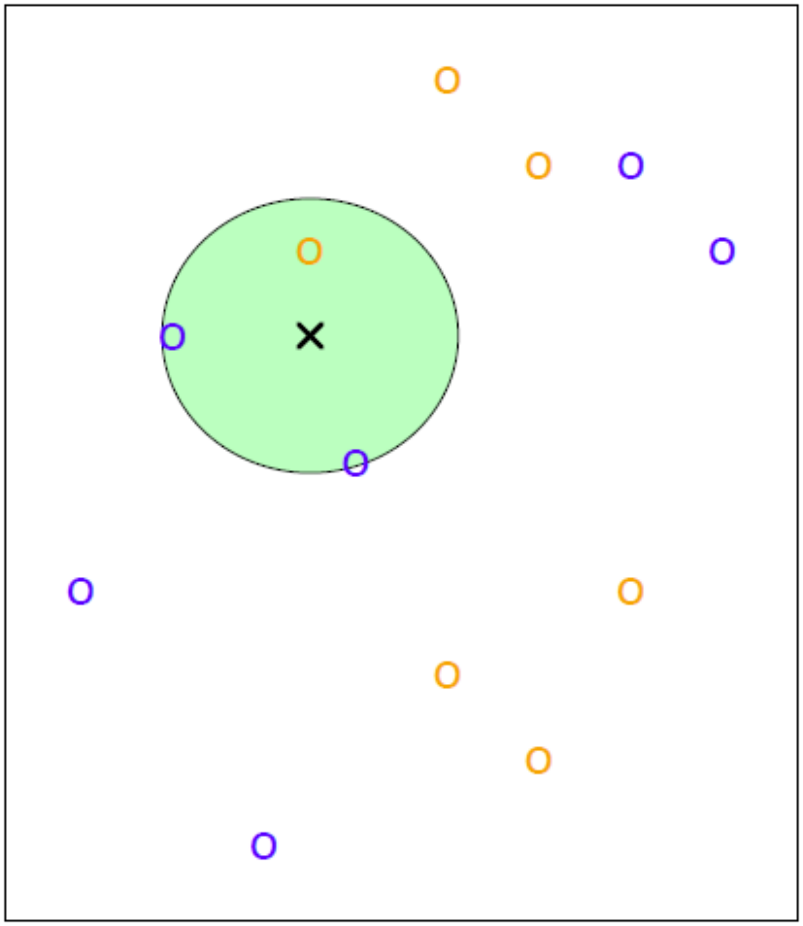
- k Nearest Neighbors is a flexible approach to estimate the Bayes Classifier.

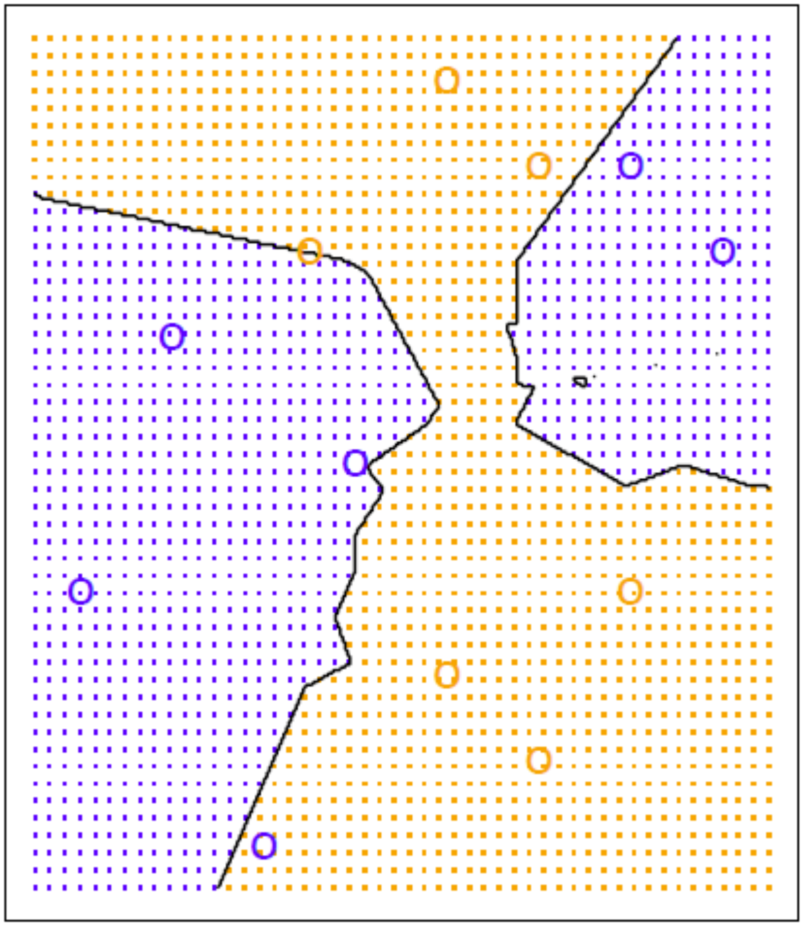
- For any given X we find the k closest neighbors to X in the training data, and examine their corresponding Y.

- If the majority of the Y's are orange we predict orange otherwise guess blue.

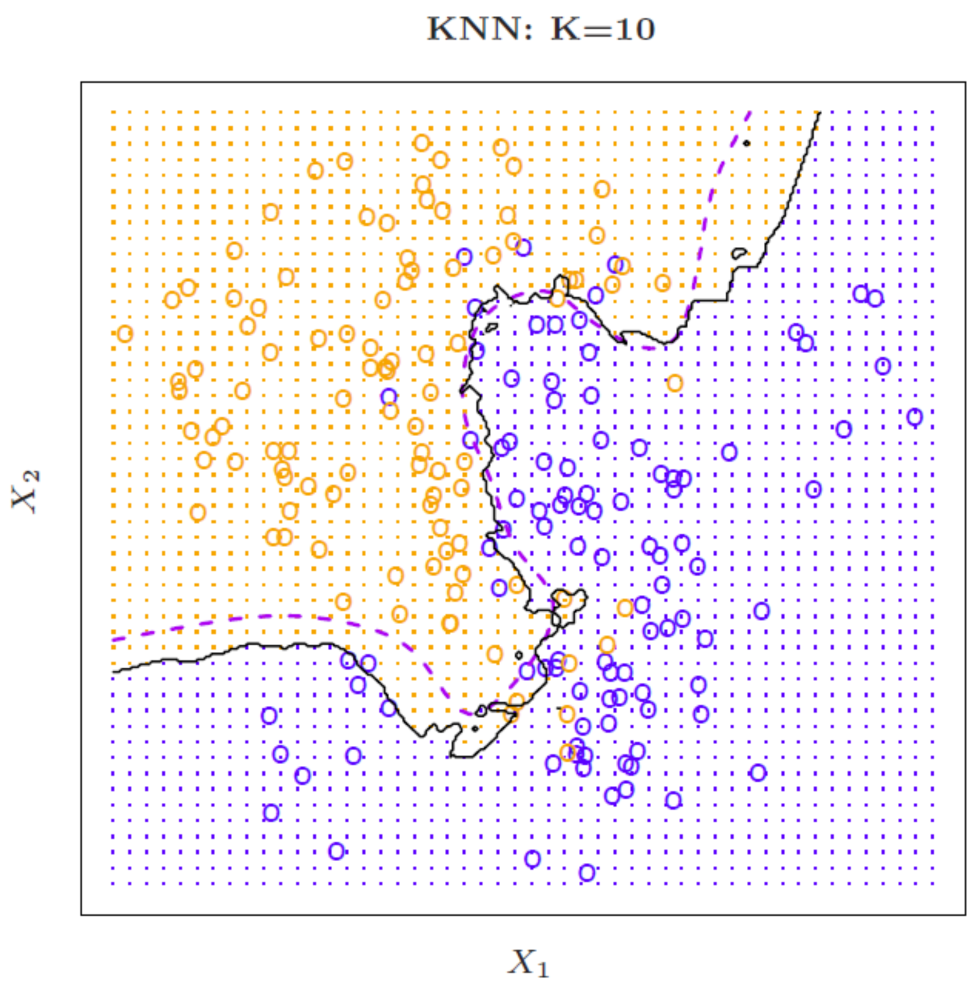
- The smaller that k is the more flexible the method will be.

KNN Example with k = 3

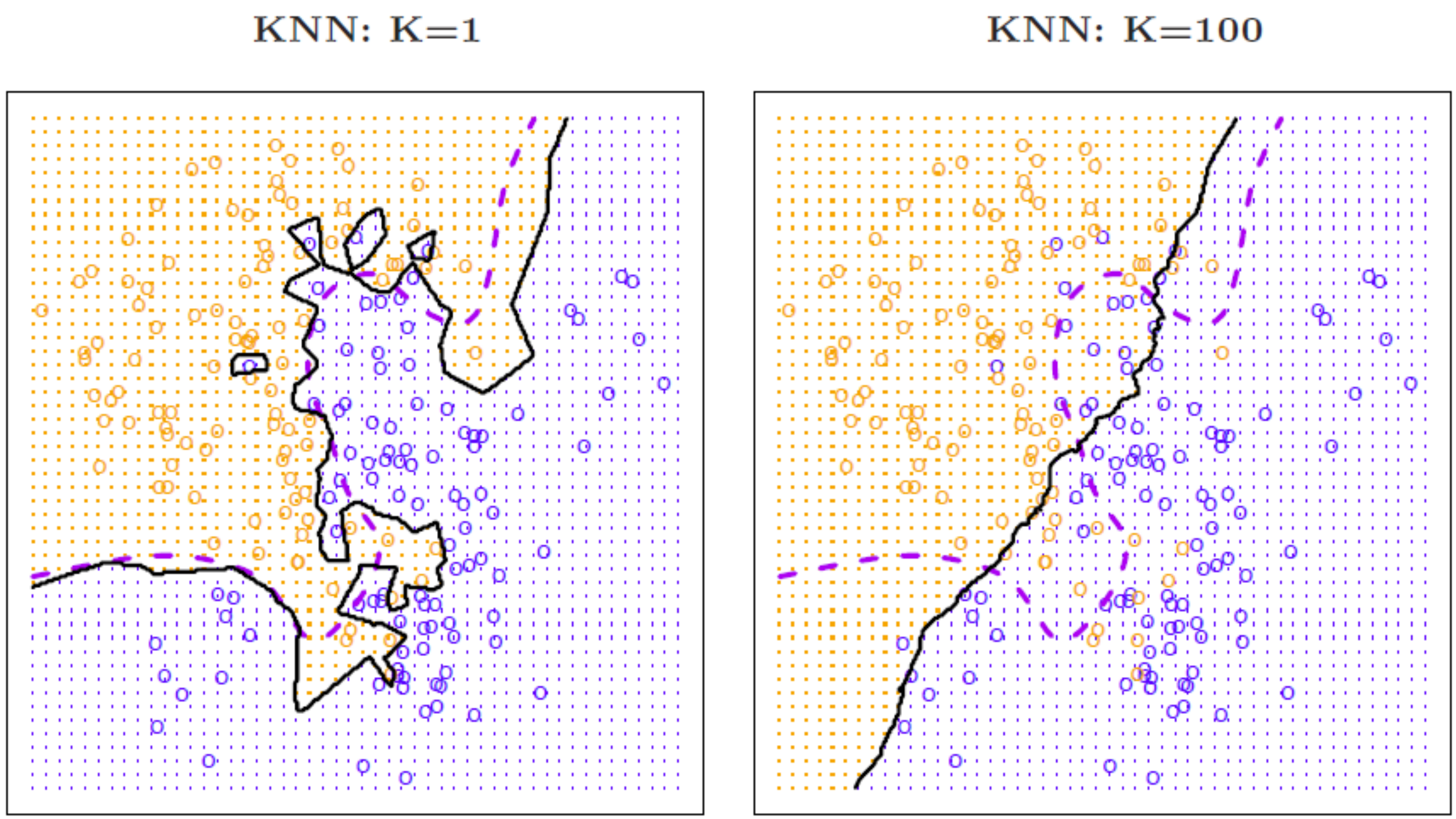




Simulated Data: K = 10



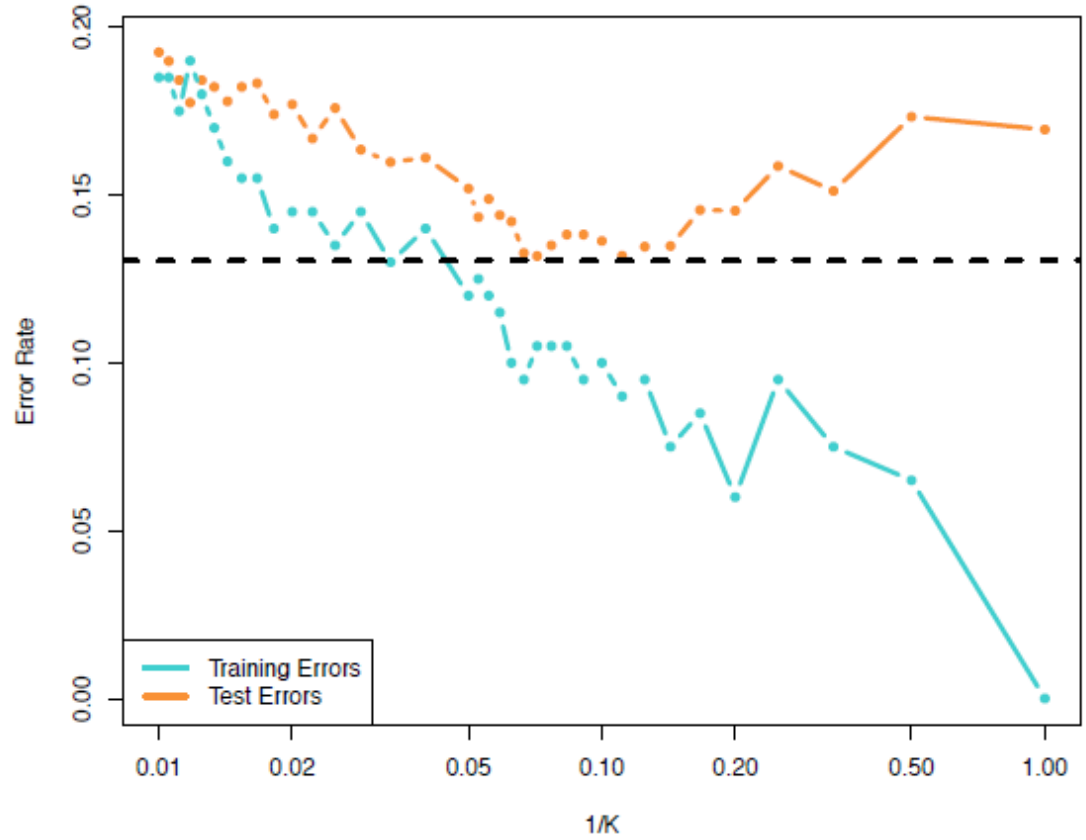
K = 1 and K = 100



Training vs. Test Error Rates on the Simulated Data

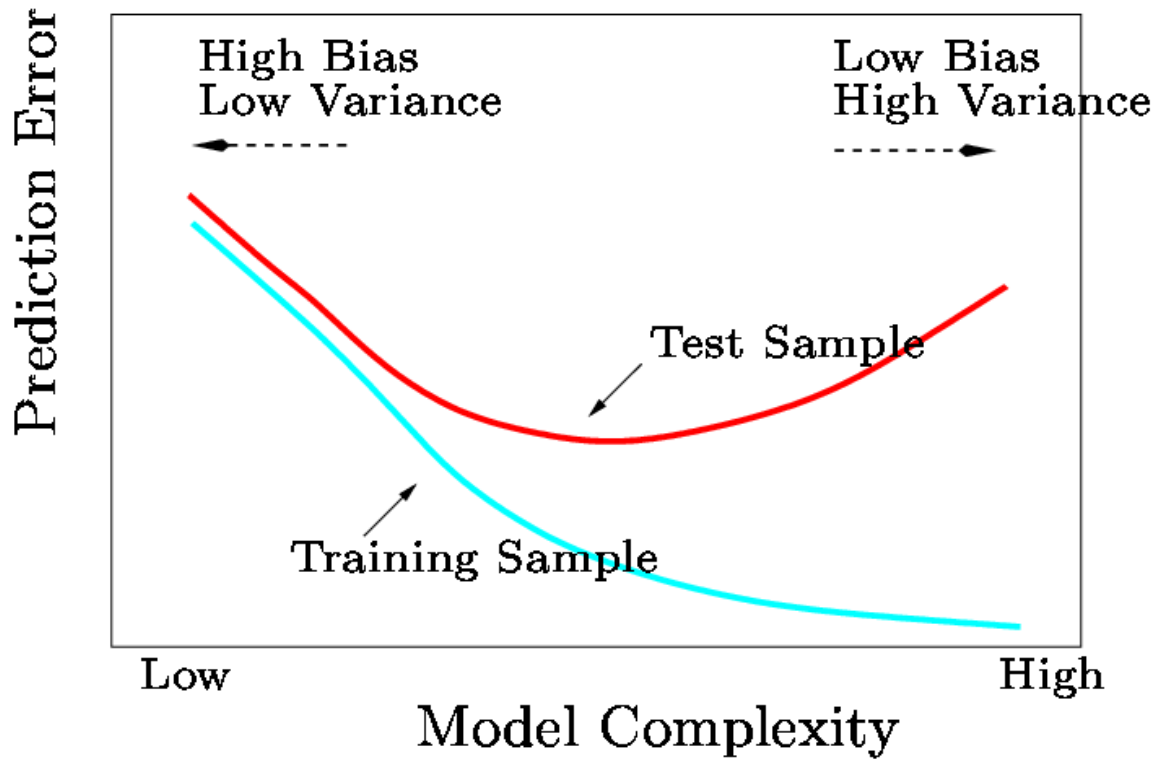
- Notice that training error rates keep going down as k decreases or equivalently as the flexibility increases.

- However, the test error rate at first decreases but then starts to increase again.



A Fundamental Picture

- In general training errors will always decline.
- However, test errors will decline at first (as reductions in bias dominate) but will then start to increase again (as increases in variance dominate).



We must always keep this picture in mind when choosing a learning method. More flexible/complicated is not always better!