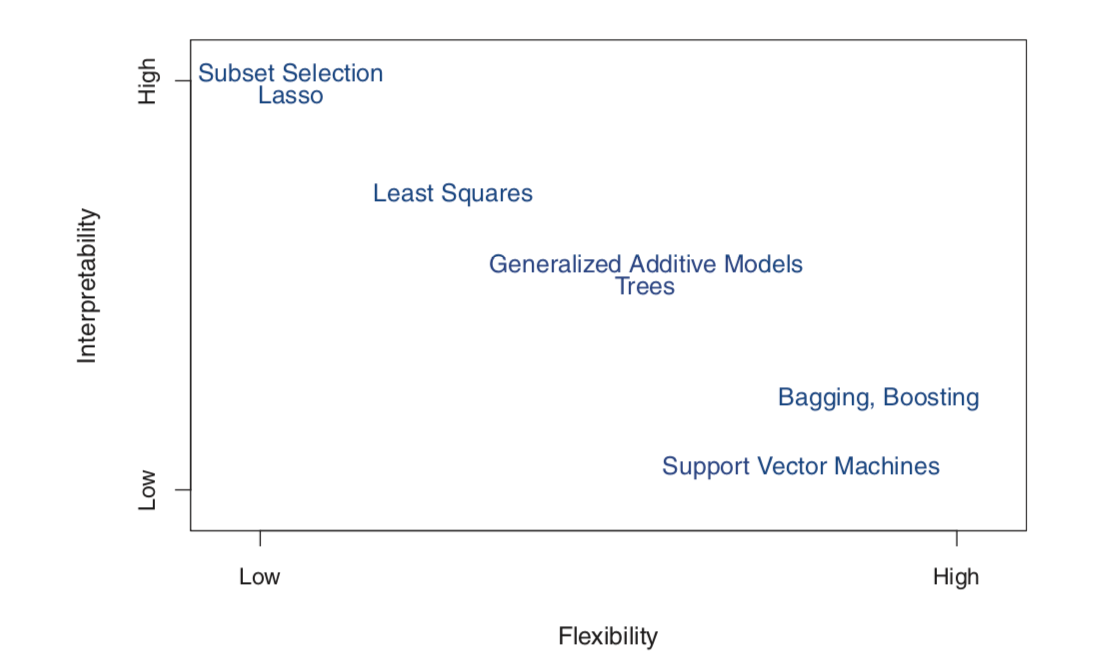
**Chapter 2 Exercises:**

**Conceptual**

**1. For each of parts (a) through (d), indicate whether we would generally expect the performance of a flexible statistical learning method to be better or worse than an inflexible method. Justify your answer.**



1. **The sample size n is extremely large, and the number of predictors p is small.**

An inflexible model will be better because an inflexible model needs a lot of data to produce an accurate model.

1. **The number of predictors p is extremely large, and the number of observations n is small.**

A flexible model will be better because the model needs to accommodate for those predictor

**(c) The relationship between the predictors and response is highly non-linear.**

A flexible model will be better because a flexible model can handle complexity more than inflexible.

**(d) The variance of the error terms, i.e. σ2 = Var(ε), is extremely high.**

Not enough information to make a judgment call.

**2. Explain whether each scenario is a classification or regression prob- lem, and indicate whether we are most interested in inference or prediction. Finally, provide n and p.**

1. **We collect a set of data on the top 500 firms in the US. For each firm we record profit, number of employees, industry and the CEO salary. We are interested in understanding which factors affect CEO salary**.

This is a regression because the target variable is CEO salary, a number.

This is an inference because we care about how the predictors relate to the target value.

n = 500

p = 3 (record profit, number of employees, industry

1. **We are considering launching a new product and wish to know whether it will be a success or a failure. We collect data on 20 similar products that were previously launched. For each prod- uct we have recorded whether it was a success or failure, price charged for the product, marketing budget, competition price, and ten other variables.**

This is a classification because the target variable is binary (success or failure).

This is a prediction because we only care about the result and not the relationship between predictors and target.

n = 20

p = 13 (price charged, marketing budget, competition price, ten other variables

1. **We are interested in predicting the % change in the US dollar in relation to the weekly changes in the world stock markets. Hence we collect weekly data for all of 2012. For each week we record the % change in the dollar, the % change in the US market, the % change in the British market, and the % change in the German market.**

This is a prediction because we only care about the % change

n= 52

p = 3 (% change in US market, % change in British market, % change in German market)

**3. We now revisit the bias-variance decomposition.**

**Provide a sketch of typical (squared) bias, variance, training error, test error, and Bayes (or irreducible) error curves, on a single plot, as we go from less flexible statistical learning methods towards more flexible approaches. The x-axis should represent the amount of flexibility in the method, and the y-axis should represent the values for each curve. There should be five curves. Make sure to label each one.**

**As flexibility increases, training MSE will decrease but test MSE will increase. Variance increases and bias decreases but levels off eventually.Bayes irreducible error stays the same**

**4. You will now think of some real-life applications for statistical learning**

1. **Describe three real-life applications in which classification might be useful. Describe the response, as well as the predictors. Is the goal of each application inference or prediction? Explain your answer.**
   1. Spam Mail: prediction problem, binary response, predictors (email address, words, subject, etc)
   2. Cancer: prediction problem, binary (benign or malignant), predictors (X-ray images)
   3. Digits: prediction problem, multiclass response, predictors (images of numbers)
2. **Describe three real-life applications in which regression might be useful. Describe the response, as well as the predictors. Is the goal of each application inference or prediction? Explain your answer.**
   1. Salary: inference problem, what features will give you the highest salary, Response as a number, predictors: education, years worked, age, last name, gender
3. **Describe three real-life applications in which cluster analysis might be useful.**
   1. Customer segment: different marketing to different demographics
   2. Clustering animals into different species
   3. Insurance: figuring out the different types of groups

**5**.  **What are the advantages and disadvantages of a very flexible (versus a less flexible) approach for regression or classification? Under what circumstances might a more flexible approach be preferred to a less flexible approach? When might a less flexible approach be preferred?**

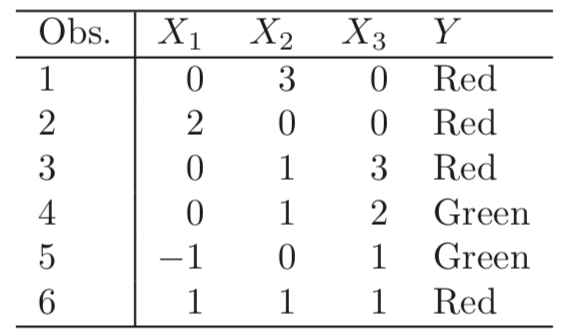
A more flexible approach can provide a better accuracy but takes away interpretability. A less flexible approach will be better if understanding the relationship between predictors and response matters.

**6. Describe the differences between a parametric and a non-parametric statistical learning approach. What are the advantages of a para- metric approach to regression or classification (as opposed to a non- parametric approach)? What are its disadvantages?**

Parametric approaches assume a functional form of the f, which means we have to estimate the parameters. ex. if we assume a linear form, then we have to estimate B0, B1....Bp (for each parameters /predictors)

Non parametric approach does NOT assume any shape, therefore can fit a wider range of functions. The only thing is that it needs a lot of data (specifically observations).

**7. The table below provides a training data set containing six observa- tions, three predictors, and one qualitative response variable.**



**Suppose we wish to use this data set to make a prediction for Y when X1 = X2 = X3 = 0 using K-nearest neighbors.**

1. **Compute the Euclidean distance between each observation and the test point, X1 =X2 =X3 =0.**
   1. 1: 3
   2. 2: 2
   3. 3: sqrt(10)
   4. 4: sqrt(5)
   5. 5: sqrt(2)
   6. 6: sqrt(3)
2. **What is our prediction with K = 1? Why?**

Green, because the nearest neighbor is observation 5 with distance of sqrt(2)

1. **What is our prediction with K = 3? Why?**

Red, because the nearest neighbors are Observation 5, 6, 2. which leads to RED

1. **If the Bayes decision boundary in this problem is highly non linear, then would we expect the best value for K to be large or small? Why?**

Small because the higher the K, the more linear it gets.