

Chapter 2.4 Exercise Solutions

Jason Moreau

January 14, 2020

Conceptual

Question 1a.

A flexible statistical learning method would be **worse** if the sample size n is extremely large, and the number of predictors p is small because a flexible method requires more variables (parameters) to reduce the errors.

Question 1b.

A flexible statistical learning method would be **better** if the number of predictors p is extremely large, and the number of observations n is small because it would provide a better fit.

Question 1c.

A flexible statistical learning method would be **better** if the predictors and response is highly non-linear because the flexible method would provide a better fit of the observations.

Question 1d.

A flexible statistical learning method would be **worse** if the variance of the error terms, i.e. $\sigma^2 = \text{Var}(\varepsilon)$, is extremely high because it would overfit the model and provide an inaccurate reading for the analyst.

Question 2a.

This scenario is a regression problem and we are most interested in inference. We are trying to determine the relationship between the dependent and independent variables/predictors.

$n = 500$ firms in the U.S.

$p =$ profit, number of employees, industry, and the CEO salary

Question 2b.

This scenario is a classification problem and we are most interested in prediction. We are trying to determine whether or not the product will be a *success* or *failure*.

$n = 20$ similar products that were previously launched

$p =$ success or failure, price charged, marketing budget, competition price, and ten other variables

Question 2c.

This scenario is a regression problem and we are most interested in prediction. We are trying to determine the relationship between the dependent (% change in the USD/Euro exchange rate) and independent variables.

$n =$ Weekly data for all of 2012

$p =$ % change each week in USD/Euro, the % change in the US market, the % change in the British market, and the % change in the German market

Question 4a.

Classification

- To find how many dogs vs. cats are in animal shelters in the United States
- To find which Japanese vehicle brand is purchased more often – Honda or Toyota
- To find which college major is most popular

Regression

- Find relationship between stock prices and bond rates
- Find relationship between life expectancy and income
- Find relationship between engine cylinders and miles per gallon

Cluster

- Find ethnicities within a city
- Find the gender demographic of a university
- Find the income demographic of a town