Textbook_Quiz_1

Tim Yang

2022-09-04

2.4 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.

(a) The sample size n is extremely large, and the number of predictors p is small.

- The performance of a flexible statistical learning model would be better, as sufficiently large sample size is provided and as flexible models, given the small number of predictors, do not have to estimate large number of parameters .

(b) The number of predictors p is extremely large, and the number of observations n is small.

- The performance of a flexible statistical learning model would be worse, as flexible models require very large number of parameters to be estimated while the number of observations is small.

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

- The performance of a flexible statistical learning model would be better, as flexible models can fit many different possible functional forms for f, not just a linear one.

(d) The variance of the error terms, sigma^2 = Var(E) i.e., is extremely high.

- The performance of a flexible statistical learning model would be worse, as flexible models can over-fit the data, following errors and noise too closely.

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

(a) 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.

- Regression, inference, n = 500, p = 3

(b) 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 product, we have recorded whether it was a success or failure, the price charged for the product, marketing budget, competition price, and ten other variables.

- Classification, prediction, n = 20, p = 13

(c) 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.

- Regression, prediction, n = 52, p = 3