Statistical learning assignment 2 - chapter 2

孫浩哲   M072040002

September 27, 2018

1.

(a) better, because a more flexible way will fit the large sample size better.

(b) worse, it might overfitting.

(c) better, the more parameters, the more degree of freedom, a flexible model can fit the data well.

(d) worse, if we use more flexible model on the situation, it might increase variance.

2.

(a) regression, inference.

Because the CEO salary is quantitative output and we want to know the relationship between two features, so we use regression to make the inference.

n - 500 firms in the US.
p - profit, number of employees, industry.

(b) classification, prediction.

The "success" and "failure" is a qualitative output and we want to predict new product's success or failure, so we use classification to make the prediction.

n - 20 similar products previously launched.
p - price charged, marketing budget, comp. price, ten other variables.
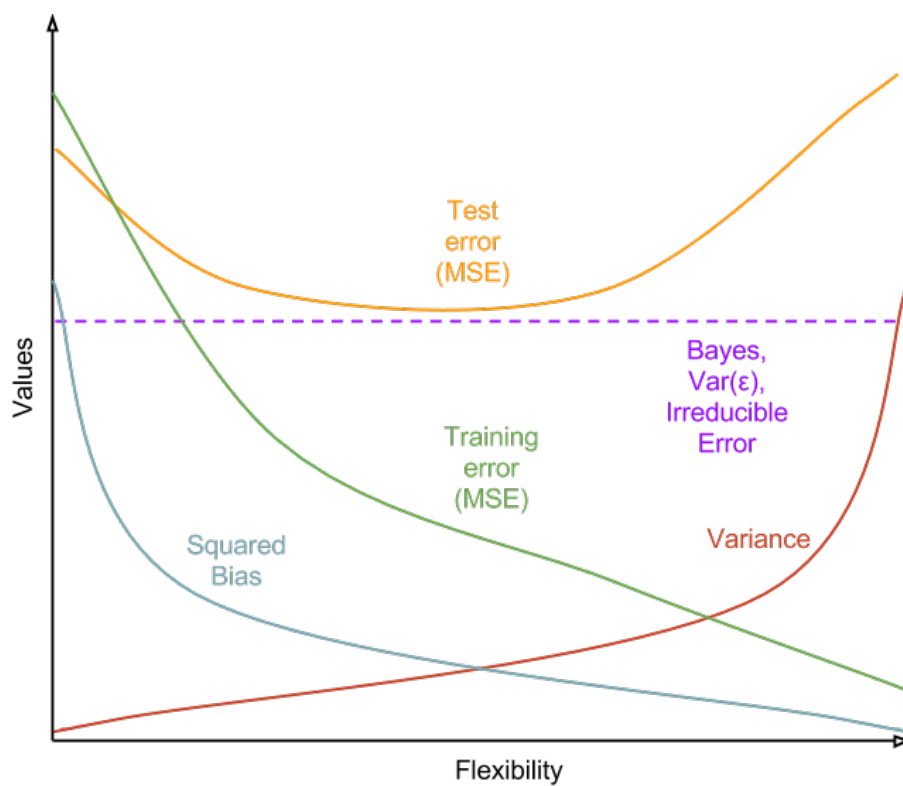
(c) regression, prediction.

The percentage of change in US dollar is quantitative output and we want to make a prediction, so we use regression to make the prediction.

n - 52 weeks of 2012 weekly data.
p - % change in US market, % change in British market, % change in German market.

3.

(a)

(b) Because each index is larger than 0, so each of five curves has the shape displayed in the figure.

4.

(a) i. Iris dataset
Response：Setosa, Versicolor, Virginica

Predictors：length and width of sepal and petal.
● Prediction.Because we want to predict the species of iris.

ii. Race
Response：European, Asian, American, African
Predictors：height, weight, complexion.
● Prediction.We want to classify the race.

iii. Stock market price direction
Response：up, down
Predictors：price movement of each day in last week
● Prediction.We just predict the Stock market price direction is up or down.

(b) i. Salary
Response：Salary
Predictors：years of education, industry experience, Nature of the work
● Inference.We build the function to explore the relationship between the features.

ii. House price
Response：House price
Predictors：region, population, GDP
● Inference.Reason is the same as (i.).

iii. Test score
Response：Test score
Predictors：reading time
● Inference.Reason is the same as (i.).

(c) i. Cancer type.
ii. City or country.
iii. Youtube video recommendations.

5.

i.

Advantages：Obtaining a better fit for non-linear models, decreasing bias.

Disadvantages：Large number of parameters, overfitting, increasing variance.

ii.

When we are interested in prediction and not the interpretability of the results.

iii.

When we are interested in inference and the interpretability of the results.