An Introduction to Statistical Learning with Applications in R

—

Gareth James, Daniela Witten, Trevor Hastie, & Robert Tibshirani

—

Chapter 1: Exercises

Graham Strickland

October 21, 2024

Question 1

(a) Given a very large sample size n and a small number of predictors p, an inflexible method would be better than a flexible one, since the risk of overfitting is less.

(b) For the same reasons as (a), a flexible method would yield better results for small n and large p.

(c) A flexible method would yield better results, since non-linear functions cannot be accurately modelled by linear functions.

(d) If there is high variance in the error terms, an inflexible method would be better, since a flexible method would introduce even more variance in the values of $\hat{f}$.

Question 2

(a) This is a classification problem, since we are trying to identify a qualitative trend in the data. It is an inference problem, since we are not trying to estimate future values of f. In this case, we have $n = 500$ and $p = 4$.

(b) This is a classification problem, since we are trying to classify the product as either a success or a failure. It is also a prediction problem, since we are looking to estimate a future output. We have $n = 20$ and $p = 14$.

(c) This is a regression problem since we have quantitative data and assume that it fits some function f, which we are attempting to estimate. Since this is a future estimate, it is a prediction problem. We have $n = 52$ and $p = 4$.