An Introduction to Statistical Learning with Applications in R

—

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

—

Chapter 2: Exercises

Graham Strickland

October 28, 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 $p = 4$.

Question 3

(a) We have the following R plot of various error curves: Arbitrary polynomi-

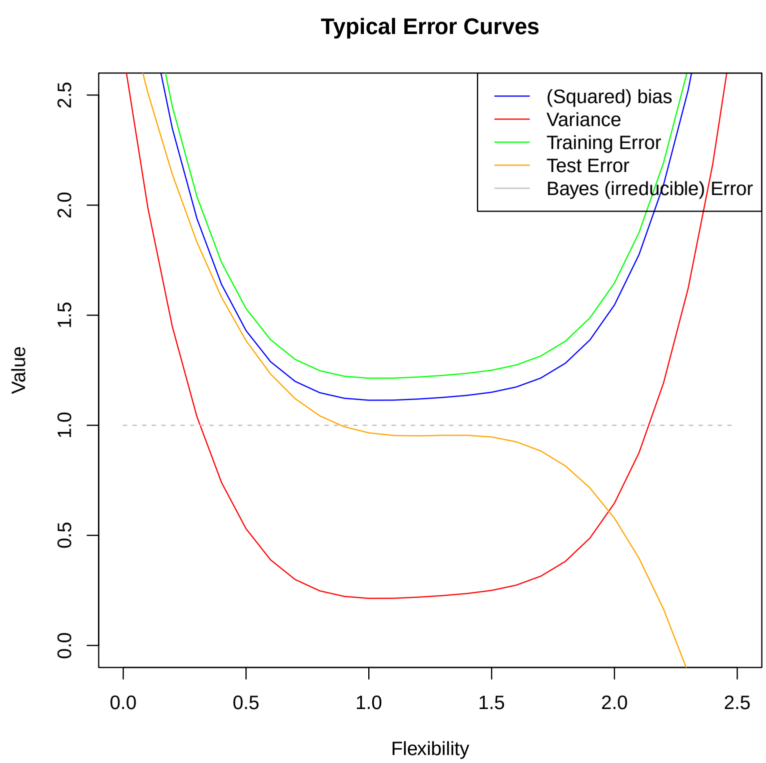

Figure 1: Typical (squared) bias, variance, training error, test error, and Bayes (irreducible) error

als of the appropriate degree were used to generate these plots, since they are approximations, using the following script:

```
# Exercise 2.3(a) - Plots of error curves

squared_bias <- function(x) {
  return((x - 1.25)^4 + 0.01 * x^3 + 0.05 * x^2 - 0.1 * x + 1.15)
```

```r
}

variance <- function(x) {
  return((x - 1.25)^4 + 0.01 * x^3 + 0.05 * x^2 - 0.1 * x + 0.25)
}

training_err <- function(x) {
  return((x - 1.25)^4 + 0.01 * x^3 + 0.05 * x^2 - 0.1 * x + 1.25)
}

test_err <- function(x) {
  return(-(x - 1.25)^3 + 0.05 * x^2 - 0.1 * x + 1.0)
}

bayes_err <- function(x) {
  return(rep(1.0, length(x)))
}

x <- seq(0, 2.5, by = 0.1)

y1 <- squared_bias(x)
y2 <- variance(x)
y3 <- training_err(x)
y4 <- test_err(x)
y5 <- bayes_err(x)

pdf("ex2_3_a.pdf")

plot(x, y1,
  type = "l", col = "blue", lwd = 1, ylim = c(0, 2.5), xlab = "Flexibility",
  ylab = "Value", main = "Typical Error Curves"
)

lines(x, y2, col = "red", lwd = 1)
lines(x, y3, col = "green", lwd = 1)
lines(x, y4, col = "orange", lwd = 1)
lines(x, y5, col = "gray", lwd = 1, lty = 2)

legend("topright",
  legend = c(
    "(Squared) bias", "Variance", "Training Error", "Test Error",
    "Bayes (irreducible) Error"
  ),
  col = c("blue", "red", "green", "orange", "gray"), lwd = 1
)
```

```
dev.off()
```

(b) We are given that the training error will always be greater than the (squared) bias, and that both will be u-shaped. The variance will always be less than the (squared) bias, while the test error will decrease as the model becomes more flexible. The Bayes error is irreducible, so it remains constant.

Question 4

(a) Three real-life applications in which classification might be useful include:

   (i) Determining a voter's party preference based upon the results of a census. The response would be a classification of a voter as being likely to vote for one of the possible political parties and the predictors would be features such as age, demographic, or location, which could be quantitative or qualitative. The goal in this case would be prediction, since we would most likely want to determine their vote in the next election.

   (ii)

   (iii)

(b)

(c)