4.4.3 Quadratic Discriminant Analysis

Dr. Lauren Perry

Multivariate Normal Distribution

We now assume the predictors $X = (X_1, X_2, \ldots, X_p)$ are drawn from a *multivariate normal distribution*.

$$X \sim N(\mu, \Sigma)$$

- Each individual predictor follows a one-dimensional normal distribution.
  - The vector μ contains all p means.
- Each pair of predictors is allowed to be correlated.
  - We represent this correlation with a $p \times p$ covariance matrix Σ that contains each variable's variance and all pairwise covariances.

QDA vs LDA

- Quadratic discriminant analysis is similar to linear discriminant analysis.
  - We assume predictors from the kth class are of the form

$$X \sim N(\mu_k, \Sigma)$$

- However, QDA allows each class to have its own covariance matrix.
  - Now, assume predictors from the kth class are of the form

$$X \sim N(\mu_k, \Sigma_k)$$

Quadratic Discriminant Analysis

Under this assumption, the Bayes classifier assigns an observation $X = x$ to the class for which

$$\begin{aligned}\delta_k(x) &= -\frac{1}{2}(x-\mu_k)^T\Sigma_k^{-1}(x-\mu_k) - \frac{1}{2}\log|\Sigma_k| + \log\pi_k \\ &= -\frac{1}{2}x^T\Sigma_k^{-1}x + x^T\Sigma_k^{-1}\mu_k - \frac{1}{2}\mu_k^T\Sigma_k^{-1}\mu_k - \frac{1}{2}\log|\Sigma_k| + \log\pi_k\end{aligned}$$

is largest.

- (Notice that x now appears as a *quadratic* in the classifier, hence the name.)

QDA vs LDA

Why choose one over the other? The bias-variance trade-off!

- LDA requires estimating significantly fewer parameters.
  - With p parameters, a covariance matrix requires estimating $p(p+1)/2$ parameters.
  - QDA requires estimating K covariance matrices, or $Kp(p+1)/2$ parameters.
  - For 50 predictors, this is some multiple of *1275*!
- QDA is much more flexible (and so much more variable).
  - LDA can therefore result in better prediction performance.
- LDA has strong assumptions about covariance matrix.
  - Violated assumption can result in high bias.

QDA vs LDA

When should we choose one over the other?

In general,

- Use LDA when there are relatively few training observations (and so reducing variance is important).
- Use QDA if the training set is very large, so that the variance is not a major concern.
- Use QDA if the common covariance matrix assumption is clearly violated.

Example: Default (10,000 observations)

```r
library(ISLR2)
data(Default)
set.seed(1)

train.ind <- sample(1:nrow(Default), floor(0.8*nrow(Default)), replace=F)
def.train <- Default[train.ind,]
def.test <- Default[-train.ind,]
```

Example: Default

```r
require(MASS)
```

```
## Loading required package: MASS

##
## Attaching package: 'MASS'

## The following object is masked from 'package:ISLR2':
##
##     Boston
```

```r
mod1 <- qda(default ~ ., def.train)
predval <- predict(mod1, def.test)$class
actual <- def.test$default
mean(predval==actual)
```

```
## [1] 0.9715
```

Note: the corresponding LDA classifier had 0.9705

Example: Default Data Confusion Matrix

```r
table(predval, actual)
```

```
##         actual
## predval   No  Yes
##      No  1929   56
##      Yes    1   14
```

Marginally better than the corresponding LDA classifier.