Class Notes

Statistical Computing & Machine Learning

Class 13

Review

The multivariate Gaussian

- Mean is specified by a vector ¯; the mean for each predictor.
- Matrix **Σ**: Covariance matrix among predictors.

Example

Plotting the decision region for LDA and QDA

```
require(ISLR)
require(MASS, quietly = TRUE)
```

```
##
## Attaching package: 'MASS'
```

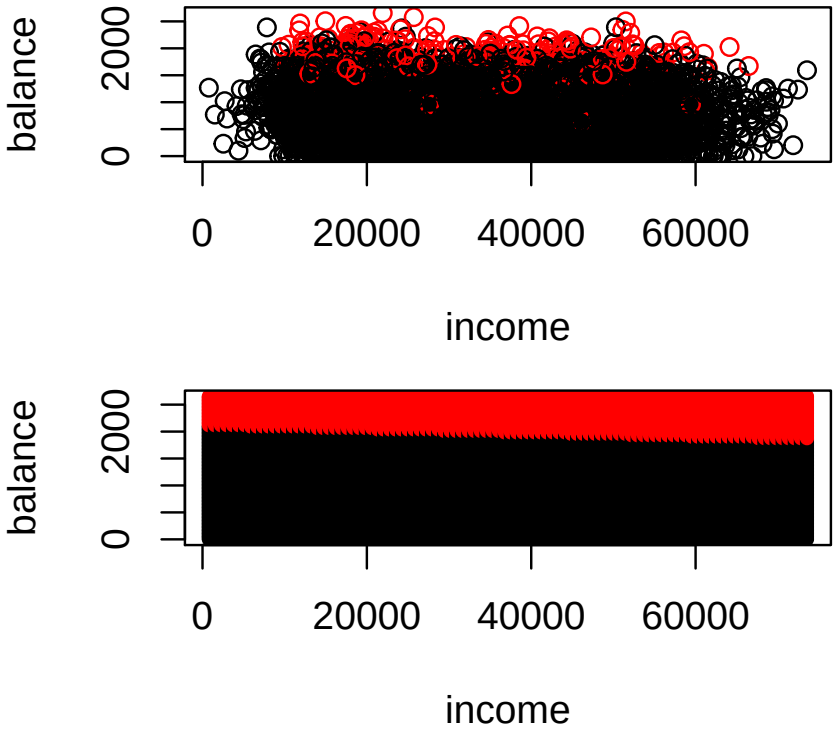
```
## The following object is masked from 'package:dplyr':
##
##     select
```

```
mod1 <- lda(default ~ balance + income, data = Default)
# plot(mod1)

pts <- with(Default,
            expand.grid(
            balance = seq(min(balance), max(balance), length = 100),
            income =  seq(min(income),  max(income),  length = 100)
))
out <- predict(mod1, newdata = pts)$class
pts <- cbind(pts, prediction = out)
plot(balance ~ income, data = Default, col = default)


plot(balance ~ income, data = pts, col = prediction, pch = 20)
```

Now fit the model with `prior = c(.5, .5)`
Show as well for `qda()`

LDA

All classes are treated as having the same **Σ**.



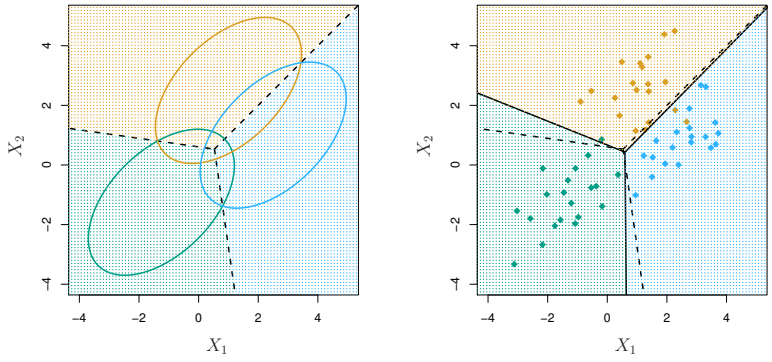


Figure 4.6 from ISL

QDA

Classes are treated with different $\mathbf{\Sigma}_i$.

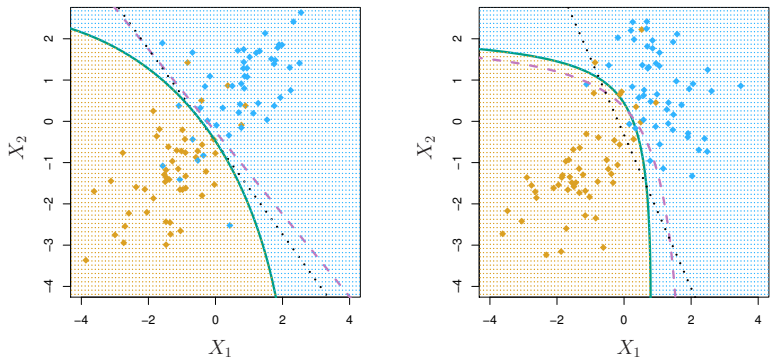


Figure 4.9 from ISL. Left: Bayes (purple dashed), LDA (black dotted), and QDA (green solid)

decision boundaries for a two-class problem with $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$. Right: QDA

Error test rates on various classifiers

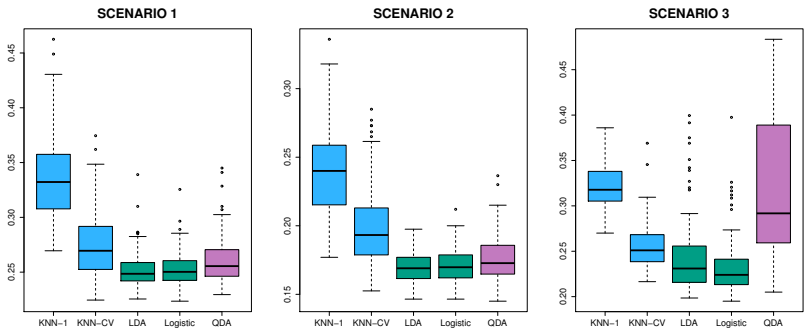


Figure 4.10 from ISL

Scenarios: In all, class means are different.

1. Each class is two uncorrelated Gaussian random vars.
2. Both classes had a correlation of -0.5
3. Uncorrelated, like (1), but the distribution is t(df=?): long tailed to right.

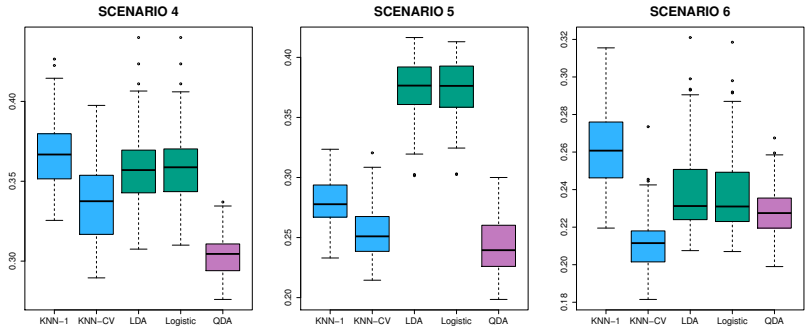


Figure 4.11 from ISL

4. Like (2), but one class has $\rho = 0.5$ and the other $\rho = -0.5$
5. A nonlinear predictor with X_1^2, X_2^2, $X_1 \times X_2$ giving a quadratic decision boundary
6. A decision boundary more complicated than a quadratic.