LDA

2025-12-22

```r
data(iris)
library(MASS)
y =(iris$Species =="versicolor")
x = iris[,-5]
n=nrow(iris)
ind_train = sample(1:n, size = 0.8*n)

y_train = y[ind_train]
y_test = y[-ind_train]
x_train = x[ind_train,]
x_test = x[-ind_train,]

fit <- lda(y_train ~ ., data = data.frame(x_train))
pred <- predict(fit, newdata = x_test)
pred_class = pred$class
mean(pred_class == y_test)
```

```
## [1] 0.7666667
```

Take home exercise:

1. Implement function cv.lda to obtain the cv score of lda.
2. One assumption that is often used in our group is the Exponential Tilting. That is, we assume that $f_1/f_0 = \exp(\alpha + \beta^T x)$. Show that this assumption is a generalization of LDA assumption. That is, show that the log of ratio of two Normal is a linear function.
3. In class, we talked about binary classification as two sample problem, and classification relies on the estimation of density ratio. We can use a lot of techniques to accomplish this. One method that is often adopted is called Kernel Density Estimation (KDE). That is,

$$f(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x_i - x)$$

   where $K_h(u) = \frac{1}{h}K(u/h)$, and $K(.)$ is a pre-specified function. One common choice of $K(.)$ is the Gaussian Kernel, that is, $K(x) = \frac{1}{\sqrt{2\pi}}\exp(\frac{-x^2}{2})$. Use this idea to do classification. First, obtain the estimate of f_0, f_1, then use CV to select h. Finally, achieve classification. You can refer to page 208-210 of ESL for more on KDE.