weekly assignment 02

Xiang Li

2024/2/19

```
library(MASS)
# library(bnlearn) library(mclust)
```

1.

```
gen_data = function(n) {
    p = 3
    n1 = n2 = n/2
    cov_1 = diag(rep(1, p)) + 0.2
    cov_2 = cov_1
    cov_2[1, 2] = cov_2[2, 1] = cov_2[1, 2] + 0.5
    x_class1 = mvrnorm(n1, mu = rep(3, p), Sigma = cov_1)
    x_class2 = mvrnorm(n2, mu = rep(2, p), Sigma = cov_2)
    x = rbind(x_class1, x_class2)
    y = rep(c(1, 2), c(n1, n2))
    df = as.data.frame(cbind(x, y))
    names(df) = c(paste0("x", 1:p), "y")
    return(df)
}
```

I can expect LDA to perform better on small training set and QDA perform better on large training set. Because in small training set the variance of QDA will be largely higher than LDA and in large training set the bias of LDA will be higher than QDA.

2.

```
set.seed(519)
n_rep = 100
acc_mt = matrix(data = 0, nrow = 100, ncol = 4)
colnames(acc_mt) = c("LDA_50", "LDA_10000", "QDA_50", "QDA_10000")
test = gen_data(10000)
for (i in 1:n_rep) {
    train1 = gen_data(50)
    train2 = gen_data(10000)
    lda_50 = lda(y ~ x1 + x2 + x3, data = train1)
    lda_10000 = lda(y ~ x1 + x2 + x3, data = train2)
```

```
    qda_50 = qda(y ~ x1 + x2 + x3, data = train1)
    qda_10000 = qda(y ~ x1 + x2 + x3, data = train2)
    pre_lda_50 = predict(lda_50, newdata = test)
    pre_lda_10000 = predict(lda_10000, newdata = test)
    pre_qda_50 = predict(qda_50, newdata = test)
    pre_qda_10000 = predict(qda_10000, newdata = test)
    acc_mt[i, 1] = mean(pre_lda_50$class == test$y)
    acc_mt[i, 2] = mean(pre_lda_10000$class == test$y)
    acc_mt[i, 3] = mean(pre_qda_50$class == test$y)
    acc_mt[i, 4] = mean(pre_qda_10000$class == test$y)
}
acc = apply(acc_mt, 2, mean)
acc
```

```
##    LDA_50 LDA_10000    QDA_50 QDA_10000
##  0.720099  0.742126  0.716371  0.754132
```

The obtained result is same with my expectation.

3.

The Bayesian classifier should be Bayesian Network.

$$\begin{aligned} P(X_1 = x_1, X_2 = x_2, X_3 = x_3, y = 1) &= P(X_1 = x_1, X_2 = x_2, X_3 = x_3 \mid y = 1)P(y = 1) \\ &= f(x_1, x_2, x_3 \mid y = 1)P(y = 1) \end{aligned} \tag{1}$$

$$\begin{aligned} P(X_1 = x_1, X_2 = x_2, X_3 = x_3, y = 2) &= P(X_1 = x_1, X_2 = x_2, X_3 = x_3 \mid y = 2)P(y = 2) \\ &= f(x_1, x_2, x_3 \mid y = 2)P(y = 2) \end{aligned} \tag{2}$$

4.