STAT 3690 Lecture 33

zhiyanggeezhou.github.io

Zhiyang Zhou (zhiyang.zhou@umanitoba.ca)

Apr 20, 2022

A joint application of LDA/QDA & PCA

- Revisit the dataset of handwritten digits in Lecture 23: `mnist` is a list with two components: `train` and `test`. Each of these is a list with two components: images and labels.
  - The `images` component is a matrix with each row for one image consisting of 28*28 = 784 entries (pixels). Their value are integers between 0 and 255 representing grey scale.
  - The `labels` components is a vector representing the digit shown in the image.
  - Uninvertible $\mathbf{S}_k$ because of the shared blank on canvas

```r
options(digits = 4)
mnist = dslabs::read_mnist()
Xtrain = mnist$train$images
Ytrain = mnist$train$labels
Xtest = mnist$test$images
Ytest = mnist$test$labels

# The 3690th image in the training set
i0 = 3690
Ytrain[i0]
image(
  matrix(Xtrain[i0,], ncol = 28),
  col = gray.colors(12, rev = TRUE), axes = FALSE, main = "3690th image in the training set")

# Build classifiers according to PC scores
decompXtrain = prcomp(Xtrain)
s = which(cumsum((decompXtrain$sdev)^2)/sum((decompXtrain$sdev)^2)>=.9)[1]
PCscoresXtrain = decompXtrain$x[,1:s]
objLda = MASS::lda(PCscoresXtrain, Ytrain, method = "moment")
objQda = MASS::qda(PCscoresXtrain, Ytrain, method   = "moment")

# Label prediction according to PC scores
xbarXtrain = colMeans(Xtrain)
PCscoresXtest = sweep(Xtest, 2, xbarXtrain) %*% decompXtrain$rotation[,1:s]
resLda = predict(objLda, PCscoresXtest)$class
resQda = predict(objQda, PCscoresXtest)$class
mean(resLda != Ytest)
mean(resQda != Ytest)
```

Alternative methods (in the view of regression)

- (Multinomial) logistic regression
- k-nearest neighbors (k-NN)
- Tree-based
    - Decision tree/classification and regression tree (CART)
    - Random forest