

Lab9: 判别与分类

1. 内容: 练习判别与分类方法的使用
 2. 作业提交: 完成后面的作业, 现场演示给助教并解释结果.
-

手写数字的识别

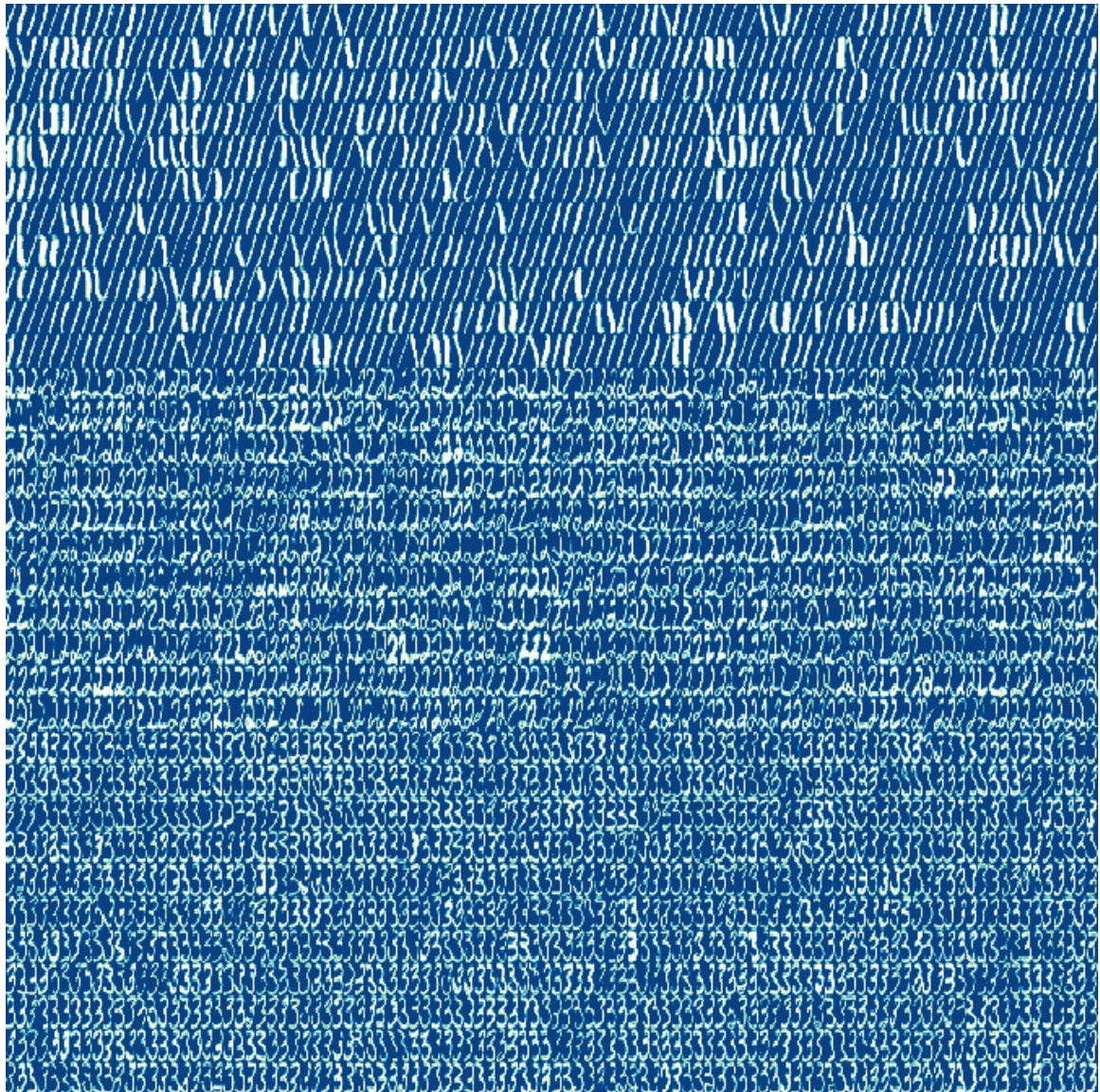
```
digits<-read.table("digits.txt")
digits<-as.matrix(digits)

library(RColorBrewer)
showMatrix <- function(x, ...)
  image(t(x[nrow(x):1,]), xaxt = 'none', yaxt = 'none',
        col = rev(colorRampPalette(brewer.pal(9, 'GnBu'))(100)), ...)

par(mfrow=c(1,3),mar=rep(0,4))
showMatrix(matrix(digits[1,],16,16))
showMatrix(matrix(digits[1101,],16,16))
showMatrix(matrix(digits[2201,],16,16))

#let's start with lda classifier.
library(MASS)
digits<-data.frame(cl=rep(1:3,each=1100),digits)
#take 75% samples as training set, you can change this
idx<-c(1:825,1100+1:825,2200+1:825)
#you should take the training set randomly.
train<-digits[idx,]
## record the running time
ptm <- proc.time()
z<-lda(cl~.,data=train)
proc.time() - ptm

#Now it's your turn to complete...
```



练习 1. 数据文件 *digits.txt* 为 USPS¹ 的手写数字数据的一部分, 其为 3300×256 的一个矩阵, 每一行为一个 16×16 分辨率的手写数据, 1-1100 行为数字“1”, 1101-2200 行为数字“2”, 2201-3300 行为数字“3”.

- (1) 将数据随机划分为训练集和检验集
- (2) 使用训练集, 选择不同的判别分类方法 (*FLDA*, *SVM* 和 *kNN*), 对分类器进行训练
- (3) 使用检验集计算错误分类率, 对比这些方法 (错误分类率和时间成本).
- (4) 由于每个手写数字为 256 维, 使用主成分方法进行降维, 重复上面 1-3 过程, 对比两类做法.