

EST171 - APRENDIZADO DE MÁQUINA
Departamento de Estatística
Universidade Federal de Minas Gerais

Lista 3

Henrique Aparecido Laureano Magno Tairone de Freitas Severino

Outubro de 2016

Sumário

Exercício I

2

Exercício I

Seu objetivo é criar classificadores para prever o dígito correspondente a uma imagem. O conjunto de dados está disponível em <https://www.kaggle.com/c/digit-recognizer>. Você deve usar o arquivo `train.csv` para criar os seus classificadores (incluindo validação), e deve fornecer ao site as previsões encontradas para o conjunto `test.csv`. Note que os dígitos correspondentes no conjunto teste não estão indicados! O site irá ranquear seu grupo de acordo com as previsões fornecidas. Como parte do exercício, você deverá:

```
# <code r> ===== #
path <- "C:/Users/henri/Dropbox/Scripts/aprendizado de maquina/list 3/"

train <- read.csv2(paste0(path, "train.csv"),
                  , header = TRUE
                  , sep = ",") ; dim(train)
# </code r> ===== #

[1] 42000    785
```

a) Inscrever seu time Kaggle. Qual o nome dele?

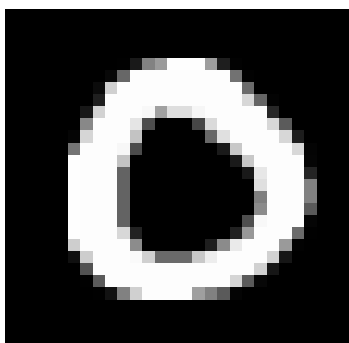
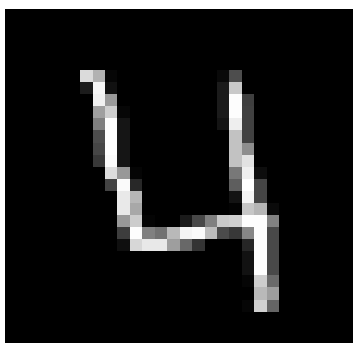
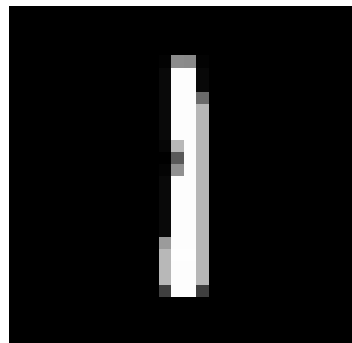
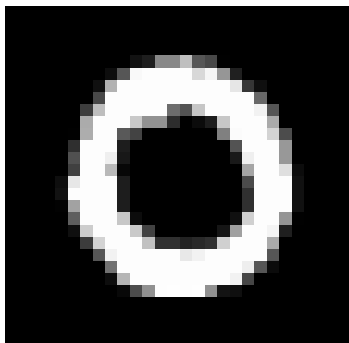
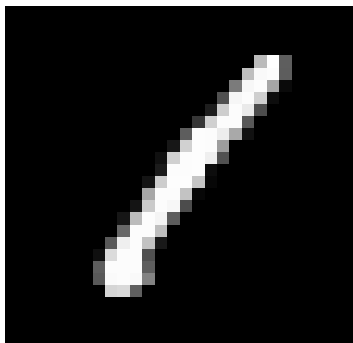
b) Plotar 5 imagens deste banco.

```
# <code r> ===== #
library(jpeg)

view <- function(row){
  c <- 1
  m <- matrix(NA, 28, 28)
  for (i in 1:28){
    for (j in 1:28){
      m[i, j] <- train[row, 1 + c]
      c <- c + 1}}
  return(t(apply(m, 2, rev)))}

par(mfrow = c(2, 3), mar = c(2, 2, 2, 2) + .1)

for (i in 1:5) image(view(i), col = grey(c(0:255)/256), axes = FALSE)
# </code r> ===== #
```



Você também deve implementar os seguintes classificadores, assim como estimar seus riscos via conjunto de teste (usando o conjunto de validação para estimar seus erros), mostrar o resultado de cada um:

c) Bagging

d) Árvore de Classificação

e) Boosting

f) SVM

Escolha um dos classificadores ajustados (o que achar melhor) e submeta suas respostas ao site
