R Notebook

Code

This is an R Markdown Notebook. When you execute code within the notebook, the results appear beneath the code.

Try executing this chunk by clicking the *Run* button within the chunk or by placing your cursor inside it and pressing *Cmd+Shift+Enter*.

— Naive Bayes —

Let's use Naive Bayes. This method makes sense, becuase in our correlation matrix plots, we showed that there is a very weak relationship between (most of the) predictors, and Naive Bayes assumes that each predictor is independent from each other.

Hide

```
library(naivebayes)
```

```
naivebayes 0.9.7 loaded
```

Hide

```
red <- read.csv('winequality-red.csv', header = TRUE, sep=";")
red <- na.omit(red)
red.quality <- red$quality
red[,-12] <- scale(red[,-12])

# Red dataset
set.seed(1)
train_red_idx <- sample(nrow(red) * 0.8) # 80-20 train-test split
train.red <- red[train_red_idx,]
train.red$quality <- as.factor(train.red$quality)
train.red.quality <- train.red$quality

test.red <- red[-train_red_idx,]
test.red$quality <- as.factor(test.red$quality)
test.red.quality <- test.red$quality
test.red <- test.red[,-12]

# Fit model:
start <- proc.time()
model <- naive_bayes(quality~., data=train.red)
red_fit_time <- proc.time() - start
red_fit_time
```

```
   user  system elapsed 
  0.097   0.005   0.221 
```

Hide

```
red_acc <- mean(as.character(test.red.quality) == as.character(predict(model, test.red)))
red_acc
```

```
[1] 0.5875
```

Hide

```
white <- read.csv('winequality-white.csv', header = TRUE, sep=";")
white <- na.omit(white)
white.quality <- white$quality
white[,-12] <- scale(white[,-12])

# White dataset
set.seed(1)
train_white_idx <- sample(nrow(white) * 0.8) # 80-20 train-test split
train.white <- white[train_white_idx,]
train.white$quality <- as.factor(train.white$quality)
train.white.quality <- train.white$quality

test.white <- white[-train_white_idx,]
test.white$quality <- as.factor(test.white$quality)
test.white.quality <- test.white$quality
test.white <- test.white[,-12]

# Fit model:
start <- proc.time()
model <- naive_bayes(quality~., data=train.white)
white_fit_time <- proc.time() - start
white_fit_time
```

```
   user  system elapsed 
  0.106   0.004   0.230 
```

Hide

```
white_acc <- mean(as.character(test.white.quality) == as.character(predict(model, test.white)))
white_acc
```

```
[1] 0.3938776
```

Weighted average test accuracy:

Hide

```
# Weighted accuracies:
(red_acc * nrow(test.red) + white_acc * nrow(test.white)) / (nrow(test.red) + nrow(test.white))
```

```
[1] 0.4415385
```

Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by pressing *Cmd+Option+I*.

When you save the notebook, an HTML file containing the code and output will be saved alongside it (click the *Preview* button or press *Cmd+Shift+K* to preview the HTML file).

The preview shows you a rendered HTML copy of the contents of the editor. Consequently, unlike *Knit*, *Preview* does not run any R code chunks. Instead, the output of the chunk when it was last run in the editor is displayed.