Machine Learning Final Project

JRP

2/1/2020

R Markdown

Good day. I created this simple KNN (known nearest neighbor) machine learning tool to predict the twenty observations in the test dataset.

Details of the study can be found here:

http://web.archive.org/web/20161224072740/http:/groupware.les.inf.puc-rio.br/har

```
train <- read.csv("https://d396qusza40orc.cloudfront.net/predmachlearn/pml-training.csv", strin
gsAsFactors = FALSE)
test <- read.csv("https://d396qusza40orc.cloudfront.net/predmachlearn/pml-testing.csv", strings
AsFactors = FALSE)

#Six young health participants were asked to perform one set of 10 repetitions
#of the Unilateral Dumbbell Biceps Curl in five different fashions:
# -Exactly according to the specification (Class A),
# -Throwing the elbows to the front (Class B),
# -Lifting the dumbbell only halfway (Class C),
# -Lowering the dumbbell only halfway (Class D)
# -Throwing the hips to the front (Class E).

#Remove first seven columns, which are identifiers I do not need

subtrain <- subset(train, select = -c(X,user_name,raw_timestamp_part_1,raw_timestamp_part_2,cvt
d_timestamp,new_window,num_window, classe))
subtest <- subset(test, select = -c(X,user_name,raw_timestamp_part_1,raw_timestamp_part_2,cvtd_t
imestamp,new_window,num_window))

#Remove non-numeric variables

sub_num <- subtrain[sapply(subtrain, is.numeric)]
sub_cha <- subtrain[!sapply(subtrain, is.numeric)]

#Normalize the remaining values

normalize <- function(x) {
    return ((x - min(x)) / (max(x) - min(x))) }

normtrain <- as.data.frame(lapply(sub_num, normalize))
normtest <- as.data.frame(lapply(subtest, normalize))

set.seed(421)
```

```
#Determine k

k <- (nrow(train))^.5

#Merge characters and numerics from test dataset
#Actually don't

#Deal with NA heavy fields / clean-up
#Replace NA with zero

normtrain <- normtrain %>%
     mutate_all(~replace(., is.na(.),0))

normtest <- normtest %>%
     mutate_all(~replace(., is.na(.),0))

#Use numeric variables from train only in test

normtest <- normtest %>%
     select(which((colnames(normtest) %in% colnames(normtrain))))

#Apply knn

knn <- knn(train=normtrain, test=normtest, cl = train$classe, k=k, prob=TRUE)

(test_labels <- knn[1:20])

##  [1] E A C A A E D B A A B C B A E E A B D B
## Levels: A B C D E

correct <- c(0,1,0,1,1,1,1,1,1,1,1,1,1,1,1,0,1,0,1)
mean(correct)

#We can see this program effectively predicted 80% of the exercise types

## [1] 0.8
```

CONCLUSION: This was a very simple program that took the numeric variables to make the predicti on with k being a function of the original number of observations. Obviously much can be done to i mprove the model and program, but for the time being this succeeded in what it was attempting to do.

Thank you for your time and I appreciate any feedback. Have a good day.

-Ramon