Glass

Charlotte Bahr

6/27/2020

library(kerasR) #DNN  
library(keras)  
library(tensorflow)#DNN  
library(data.table)  
library(dataPreparation)  
library(e1071) #Naive Bayes  
library(randomForest) #random Forest  
library(class) #k-Nearest Neighbor   
library(kernlab) #Support Vector Machine  
library(mlbench)#contains Glass data  
library(keras) #contains MNIST data  
library(caret)  
library(mltest)  
library(dplyr) #wg. select  
use\_session\_with\_seed(5678)

## Set session seed to 5678 (disabled GPU, CPU parallelism)

## DATA PRE PROCESSING  
data("Glass")  
Glass <- Glass  
  
Glass$Type[Glass$Type == 3] <- "5"  
Glass$Type[Glass$Type == 4] <- "5"  
Glass$Type[Glass$Type == 6] <- "5"  
  
set.seed(5678)  
Glass <- as.data.frame(Glass)  
Glass$Type <- droplevels(Glass$Type)  
  
# prep for statistical classifier  
set.seed(1234)  
  
train\_index\_SC <- sample(1:nrow(Glass), 0.7\*nrow(Glass))  
test\_index\_SC <- setdiff(1:nrow(Glass), train\_index\_SC)  
  
Train <- Glass[train\_index\_SC,]  
Test <- Glass[test\_index\_SC,]  
  
True\_Label <- Test$Type  
  
#prep for DNN  
  
X\_train <- Train %>%   
 select(-Type) %>%   
 scale()  
  
Y\_train <- to\_categorical(Train$Type)  
  
X\_test <- Test %>%   
 select(-Type) %>%   
 scale()  
  
Y\_test <- to\_categorical(Test$Type)  
  
##MODELS   
  
#k-Nearest Neighbor  
  
pc <- proc.time()  
model\_KNN <- knn(Train, Test, as.factor(Train$Type), k=15) ##sqrt214 =14,63  
print(proc.time() - pc)

## user system elapsed   
## 0.006 0.000 0.006

#Naive Bayes  
  
pc <- proc.time()   
model\_NB <- naiveBayes(as.factor(Train$Type) ~. , Train)  
print(proc.time() - pc)

## user system elapsed   
## 0.005 0.001 0.006

#Random Forest  
  
pc <- proc.time()   
model\_RF <- randomForest(as.factor(Train$Type) ~. , Train)  
print(proc.time() - pc)

## user system elapsed   
## 0.071 0.004 0.074

#Support Vector Machine  
   
pc <- proc.time()   
model\_SVM <- ksvm(Train$Type ~. , Train, type = "C-svc", C = 1, kernel = "vanilladot" )

## Setting default kernel parameters

print(proc.time() - pc)

## user system elapsed   
## 0.013 0.001 0.014

#Deep Neural Network  
  
  
pc <- proc.time()   
  
model\_DNN <- Sequential()  
  
model\_DNN$add(Dense(units=256, input\_shape = dim(X\_train)[2]))  
model\_DNN$add(LeakyReLU())  
model\_DNN$add(Dropout(0.4))  
  
model\_DNN$add(Dense(units=256))  
model\_DNN$add(LeakyReLU())  
model\_DNN$add(Dropout(0.3))  
  
model\_DNN$add(Dense(units=256))  
model\_DNN$add(LeakyReLU())  
model\_DNN$add(Dropout(0.2))  
  
model\_DNN$add(Dense(8))  
model\_DNN$add(Activation("softmax"))  
  
# compile  
keras\_compile(model\_DNN, loss ="categorical\_crossentropy", optimizer = RMSprop(), metrics = "accuracy")  
keras\_fit(model\_DNN, X\_train, Y\_train, batch\_size = 128, epochs = 32, verbose= 1, validation\_split = 0.2)  
  
print(proc.time() - pc)

## user system elapsed   
## 1.453 0.255 1.724

##EVALUATION METRICS  
  
#predictions  
  
pred\_KNN <- #has no prediction value  
  
pred\_NB <- as.factor(predict(model\_NB, Test))  
  
pred\_RF <- as.factor(predict(model\_RF, Test))  
  
pred\_SVM <- as.factor(predict(model\_SVM, Test))  
  
pred\_DNN <- as.factor(keras\_predict\_classes(model\_DNN, X\_test))  
  
  
CF\_KNN <- table(model\_KNN, True\_Label)  
   
CF\_NB <- table(pred\_NB, True\_Label)  
   
CF\_RF <- table(pred\_RF, True\_Label)  
   
CF\_SVM <- table(pred\_SVM, True\_Label)  
  
CF\_DNN <- table(pred\_DNN, True\_Label)  
  
  
print(CF\_KNN)

## True\_Label  
## model\_KNN 1 2 5 7  
## 1 27 1 0 0  
## 2 0 21 0 0  
## 5 0 0 8 1  
## 7 0 0 1 6

print(CF\_NB)

## True\_Label  
## pred\_NB 1 2 5 7  
## 1 22 18 4 0  
## 2 5 4 2 0  
## 5 0 0 0 1  
## 7 0 0 3 6

print(CF\_RF)

## True\_Label  
## pred\_RF 1 2 5 7  
## 1 18 4 3 0  
## 2 5 16 0 0  
## 5 4 1 5 1  
## 7 0 1 1 6

print(CF\_SVM)

## True\_Label  
## pred\_SVM 1 2 5 7  
## 1 18 12 4 0  
## 2 9 9 2 0  
## 5 0 0 1 1  
## 7 0 1 2 6

print(CF\_DNN)

## True\_Label  
## pred\_DNN 1 2 5 7  
## 1 17 3 2 0  
## 2 9 16 2 1  
## 5 1 2 4 1  
## 7 0 1 1 5

#metrics  
set.seed(2445)  
  
ml\_test\_KNN <- ml\_test(model\_KNN, True\_Label, output.as.table = FALSE)  
  
ml\_test\_NB <- ml\_test(pred\_NB, True\_Label, output.as.table = FALSE)  
  
ml\_test\_RF <- ml\_test(pred\_RF, True\_Label, output.as.table = FALSE)  
  
ml\_test\_SVM <- ml\_test(pred\_SVM, True\_Label, output.as.table = FALSE)  
  
ml\_test\_DNN <- ml\_test(pred\_DNN, True\_Label, output.as.table = FALSE)  
  
#Macro Average Accuracy  
  
MAvA\_KNN <- print((sum(ml\_test\_KNN$balanced.accuracy, na.rm = TRUE))/4)

## [1] 0.9546342

MAvA\_NB <- print((sum(ml\_test\_NB$balanced.accuracy, na.rm = TRUE))/4)

## [1] 0.6040656

MAvA\_RF <- print((sum(ml\_test\_RF$balanced.accuracy, na.rm = TRUE))/4)

## [1] 0.7843102

MAvA\_SVM <- print((sum(ml\_test\_SVM$balanced.accuracy, na.rm = TRUE))/4)

## [1] 0.6390338

MAvA\_DNN <- print((sum(ml\_test\_DNN$balanced.accuracy, na.rm = TRUE))/4)

## [1] 0.735832

#Macro Average F1   
  
MAvF1\_KNN <- print((sum(ml\_test\_KNN$F1, na.rm = TRUE))/4)

## [1] 0.9261485

MAvF1\_NB <- print((sum(ml\_test\_NB$F1, na.rm = TRUE))/4)

## [1] 0.4030356

MAvF1\_RF <- print((sum(ml\_test\_RF$F1, na.rm = TRUE))/4)

## [1] 0.6841234

MAvF1\_SVM <- print((sum(ml\_test\_SVM$F1, na.rm = TRUE))/4)

## [1] 0.4876384

MAvF1\_DNN <- print((sum(ml\_test\_DNN$F1, na.rm = TRUE))/4)

## [1] 0.6296879

#MAvMCC  
  
MAvMCC\_KNN <- print((sum(ml\_test\_KNN$MCC, na.rm = TRUE))/4)

## [1] 0.9109648

MAvMCC\_NB <- print((sum(ml\_test\_NB$MCC, na.rm = TRUE))/4)

## [1] 0.1826387

MAvMCC\_RF <- print((sum(ml\_test\_RF$MCC, na.rm = TRUE))/4)

## [1] 0.5523544

MAvMCC\_SVM <- print((sum(ml\_test\_SVM$MCC, na.rm = TRUE))/4)

## [1] 0.2812175

MAvMCC\_DNN <- print((sum(ml\_test\_DNN$MCC, na.rm = TRUE))/4)

## [1] 0.4753409

#MAvGeometricMean  
  
MAvGM\_KNN <- print((sum(ml\_test\_KNN$geometric.mean, na.rm = TRUE))/4)

## [1] 0.9537208

MAvGM\_NB <- print((sum(ml\_test\_NB$geometric.mean, na.rm = TRUE))/4)

## [1] 0.4406548

MAvGM\_RF <- print((sum(ml\_test\_RF$geometric.mean, na.rm = TRUE))/4)

## [1] 0.7783041

MAvGM\_SVM <- print((sum(ml\_test\_SVM$geometric.mean, na.rm = TRUE))/4)

## [1] 0.5796574

MAvGM\_DNN <- print((sum(ml\_test\_DNN$geometric.mean, na.rm = TRUE))/4)

## [1] 0.7217735