code for random forest, CART and naive bayes

Zefan Xu

2022-04-01

```
library(tidyverse)
library(dimRed)
library(readr)
library(rpart)
library(rpart.plot)
library(rattle)
```

```
# training and test set
train_url = "https://raw.githubusercontent.com/ashwinpo/503Project/main/tf-train.csv"
test_url = "https://raw.githubusercontent.com/ashwinpo/503Project/main/tf-test.csv"
train = read.csv(url(train_url))
test = read.csv(url(test_url))
```

```
# clean the data
set.seed(503)
train = train %>% select(-reuters)
test = test %>% select(-reuters)
train$label = factor(train$label, levels = c(0,1), labels = c("Fake", "True"))

# subsamples
sub_ind <- sample(seq_len(nrow(train)), size = floor(0.3 * nrow(train)))

sub <- train[sub_ind, ]
rest = train[-sub_ind,]
```

Random Forest

```
# Random Forest model
library(randomForest)
rf_tree = randomForest(as.factor(label) ~ ., data = sub, mtry = 3, importance = TRUE)
# variable importance plot
varImpPlot(rf_tree, n.var = 8)
```

```
# accuracy function
accuracy = function(cm){
    print(paste("Fake Accuracy: ", round(cm[1,1] / (cm[1,1] + cm[2,1]), 4)))
    print(paste("True Accuracy: ", round(cm[2,2] / (cm[1,2] + cm[2,2]),4)))
    print(paste("Overall Accuracy: ", round((sum(diag(cm)) / sum(cm)), 4)))
}
```

```
# Training accuracy for Random Forest
rf_train.pred = rf_tree$predicted
cm = table(rf_train.pred, sub$label)
cm
accuracy(cm)
```

```
# Test accuracy for Random Forest
predictRF = predict(rf_tree, newdata=test)
cm1 = table(test$label, predictRF)
cm1
accuracy(cm1)
```

Classification Tree

```
# Classification Tree
tree = rpart(as.factor(label) ~ ., data = train, parms = list(split = "gini"), method = "class")
fancyRpartPlot(tree)
```

```
printcp(tree)
```

Naive Bayes

```
library(e1071)
```

```
# Fit naive bayes model
naiveB = naiveBayes(label~., data = train)
```

```
# Training accuracy for naive bayes
naiveB_train = predict(naiveB, train)
cm2 = table(naiveB_train, train$label)
cm2
accuracy(cm2)
```

```
# Test accuracy for naive bayes
naiveB_pred = predict(naiveB, newdata = test)
cm3 = table(naiveB_pred, test$label)
cm3
accuracy(cm3)
```