STT 811

In-Class Assignment 9

This problem will use the OJ dataset. Note that you will use Purchase as the target (no need to convert to 0/1)

1. Split the data into training and test datasets (with a 75/25 split).

```
oj <-  ISLR2::OJ
split_pct <- 0.75
n <- length(oj$Purchase)*split_pct # train size
row_samp <- sample(1:length(oj$Purchase), n, replace = FALSE)
train <- oj[row_samp,]
test <- oj[-row_samp,]
```

2. Build a Naïve Bayes model for your target based on PriceDiff and LoyalCH (using the naivebayes command). Compute the confusion matrix for both the train and test datasets. How do they compare?

```
naiveBayesModel <- naiveBayes(Purchase ~ PriceDiff+LoyalCH, data = train)
test_pred <- predict(naiveBayesModel,test)
train_cm <- confusionMatrix(as.factor(as.integer(2*naiveBayesModel$fitted.values)),
reference = as.factor(train$Purchase))
test_cm <- confusionMatrix(as.factor(as.integer(2*test_pred)), reference =
as.factor(test$Purchase))
print("train CM")
train_cm$table
print("test CM")
test_cm$table
```

3. Build another naïve Bayes model using Store as well (make sure it is categorical). Compare how the confusion matrix does (on test) vs. (2)

```
naiveBayesModel2 <- naiveBayes(Purchase ~ PriceDiff+LoyalCH+as.factor(STORE), data =
train)
test_pred <- predict(naiveBayesModel2,test, type = "response")
train_cm <- confusionMatrix(as.factor(as.integer(2*naiveBayesModel$fitted.values)),
reference = as.factor(train$Purchase))
test_cm <- confusionMatrix(as.factor(as.integer(2*test_pred)), reference =
as.factor(test$Purchase))
print("train CM 2")
train_cm$table
print("test CM 2")
test_cm$table
```