

Q40

Rahul Atre

2023-11-24

Q40 [Classification]

Do exercises 7 and 8, from DUDADS, chapter 21, using the wine.csv dataset.

Exercise 7: Build bagging models for the Wine dataset.

```
set.seed(144) #Setting seed for reproducibility

wine_df = read.csv("wine.csv")
wine_df$Class = as.factor(wine_df$Class) #Convert to a factor

#Let us assume that the response variable we would like to predict is Class

n = nrow(wine_df) #number of rows

#Let us do a 70/30 split on the training/test data

training_set = wine_df[sample(n, 0.7 * n), ]
test_set = wine_df[setdiff(1:n, rownames(training_set)), ]

wine_bagging_model = bagging(Class ~ ., data = training_set, nbagg = 10, coob = TRUE, control = rpart.c

wine_prediction = predict(wine_bagging_model, test_set)

table(wine_prediction, test_set$Class) #Confusion matrix

##
## wine_prediction  1  2  3
##                1 19  2  0
##                2  0 21  0
##                3  0  1 11

wine_accuracy = sum(wine_prediction == test_set$Class) / nrow(test_set)
wine_accuracy
```

```
## [1] 0.9444444
```

Therefore, with the bagging model for the Wine dataset, we have obtained a prediction accuracy rate of 94.44%.

Exercise 9: Build random forest models for the Wine dataset.

```

set.seed(144) #Setting seed for reproducibility

wine_df = read.csv("wine.csv")
wine_df$Class = as.factor(wine_df$Class) #Convert to a factor

#Let us assume that the response variable we would like to predict is Class

n = nrow(wine_df) #number of rows

#Let us do a 70/30 split on the training/test data

training_set = wine_df[sample(n, 0.7 * n), ]
test_set = wine_df[setdiff(1:n, rownames(training_set)), ]

wine_randomForest_model = randomForest(Class ~ ., data = training_set)

wine_prediction = predict(wine_randomForest_model, test_set)

table(wine_prediction, test_set$Class) #Confusion matrix

```

```

##
## wine_prediction  1  2  3
##                1 19  1  0
##                2  0 22  0
##                3  0  1 11

```

```

wine_accuracy = sum(wine_prediction == test_set$Class) / nrow(test_set)
wine_accuracy

```

```
## [1] 0.962963
```

Therefore, with the random forest model for the Wine dataset, we achieve a prediction accuracy rate of 96.29%. This is slightly better than the bagging model that we fitted in Exercise 8.