

## Q39

Rahul Atre

2023-11-23

### Q39 [Classification]

Do exercises 5 and 6, from DUDADS, chapter 21, using the wine.csv dataset.

**Exercise 5:** Construct and evaluate ANN 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

train_indices = sample(c(1:nrow(wine_df)), 0.80 * nrow(wine_df))

training_set = wine_df[train_indices,]
test_set = wine_df[-train_indices,]

model = neuralnet(Class~., data=training_set, hidden=c(4,2), linear.output = FALSE)

wine_prediction = predict(model, test_set)

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

## [1] 0.3333333
```

Therefore, when running the ANN model on the Wine dataset, we obtain an accuracy rate of 33.33%.

**Exercise 6:** Re-run the ANN models incorporating 10 hidden layers with 30 nodes. How much more time does it take to run a “bigger” neural network on 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

train_indices = sample(c(1:nrow(wine_df)), 0.80 * nrow(wine_df))

training_set = wine_df[train_indices,]
```

```
test_set = wine_df[-train_indices,]

model = neuralnet(Class~., data=training_set, hidden=c(10,30), linear.output = FALSE)

wine_prediction = predict(model, test_set)

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

```
## [1] 0.02777778
```

The larger neural network takes slightly more time to run, at about 3-4 seconds instead of instantaneous like the previous model. However, the accuracy rate is significantly lower, at about 2.78%, in comparison to 33.33% above.

Therefore, a larger model with more layers and nodes does not necessarily imply better performance in accuracy. For this case in particular, it was a computational expense for lower accuracy.