

Artificial Neural Networks (Deep Learning)

Findings

A bank has been losing a higher number of customers than usual. The company have given me a portion of their client information so that I can determine which clients in the future. The dataset is *Churn_Modelling.csv*.

Several of the columns (RowNumber, CustomerID, and Surname) are irrelevant in determining why the customer left the bank. These columns will be removed. I then encoded the two columns *Geography* and *Gender* so that they were integers rather than strings. I then removed a dummy variable from *Geography* to avoid the dummy variable trap. Lastly, I split the dataset into a training (80% - 8,000) and test set (20% - 2,000). Lastly, I applied feature scaling. This concluded my data preprocessing.

We then begin to create the neural network. We create the first three layers with six nodes and one output layer with a single node. We set an epochs of 100 and a batch size of 10. After running the algorithm we get an accuracy of 86.4% (see *Epoch (ANN).txt* for full epoch).

Afterward we predict the test set result to see if our algorithm was accurate. We can take a look at the confusion matrix.

	0	1
0	1526	69
1	204	201

- 1526 people who left the bank were predicted as leaving the bank.
- 69 people who left the bank were predicted as staying the bank.
- 204 people who staying the bank were predicted as leaving the bank.
- 201 people who stayed with the bank were predicted as staying with the bank.

1,727 out of 2000 were correctly predicted (86.35%). This is a very close result to the prediction of the epoch. Therefore, we know our model is accurate and consistent.

The first following code when executed will show which users are likely to leave the company. The second line of code will set a true and false value if the user is above or below 50% likelihood of leaving the company.

```
# Predict the test set results
y_pred = cl.predict(X_test)
y_pred = (y_pred > 0.5) # True if larger than 0.5
```