Deep Learning Analysis Report

Module 21 Challenge – Alphabet Soup Charity

Overview

This analysis uses machine learning and neural networks techniques to obtain a binary classifier that attempts to predict whether applicants funded by *Alphabet Soup* will be successful.

The workflow for the analysis was as follows:

1. Data pre-processing
2. Neural network model compiling, training and evaluation
3. Graphical representation of accuracy and loss for each model
4. Storage of each model for future reference

Data Pre-processing

The dataset used for this analysis was obtained from:
https://static.bc-edx.com/data/dla-1-2/m21/lms/starter/charity_data.csv

This dataset comprises 34299 rows by 12 columns.

Determination of Feature set and Targets

The features of the model are:

- **EIN** and **NAME**—Identification columns
- **APPLICATION_TYPE**—Alphabet Soup application type
- **AFFILIATION**—Affiliated sector of industry
- **CLASSIFICATION**—Government organisation classification
- **USE_CASE**—Use case for funding
- **ORGANIZATION**—Organisation type
- **STATUS**—Active status
- **INCOME_AMT**—Income classification
- **SPECIAL_CONSIDERATIONS**—Special considerations for application
- **ASK_AMT**—Funding amount requested

The target for the model was:

- **IS_SUCCESSFUL**—Was the money used effectively

Deep Learning Analysis Report

Module 21 Challenge – Alphabet Soup Charity

The **EIN** and **NAME** identification columns were removed from the feature set prior to modelling.

Preparation of data set for modelling

Initially, the number of unique values in each column of the dataset were identified. For columns of data comprising more than 10 unique values, the total number of data points for each unique value was obtained.

In the case of columns comprising more than 10 unique values, binning of the least frequently occurring variables into a 'other' category was performed and this binning procedure was checked visually.

This preprocessed dataset was then used to identify the Feature Set Array and Target Array as outlined above. A StandardScaler algorithm was performed to normalise variables and data was split into training and testing sets for neural network modelling.

Neural Network Modeling

Several different models with hyperparameter tuning were used to make predictions with accuracy and loss calculations also obtained. The predictive epoch history was graphed for each model variant for visual comparison.

Model Attempt 1

Number of hidden layers: **2**

Number of input parameters: **108**

Number of neurons in first hidden layer: **5**

Number of neurons in second hidden layer: **9**

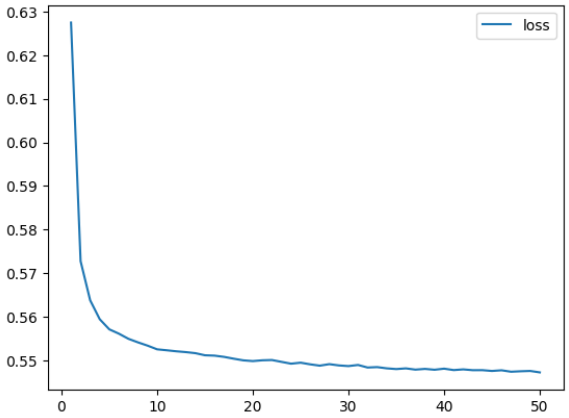
Hidden layer activation function: **relu**

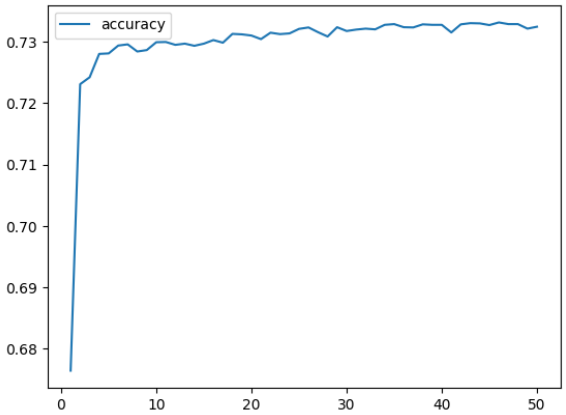
Output layer activation function: **sigmoid**

Number of Epochs: **50**

Deep Learning Analysis Report

Module 21 Challenge – Alphabet Soup Charity





Loss: 0.5520873069763184 Accuracy: 0.7308454513549805

Model Attempt 2

Number of hidden layers: **2**

Number of input parameters: **108**

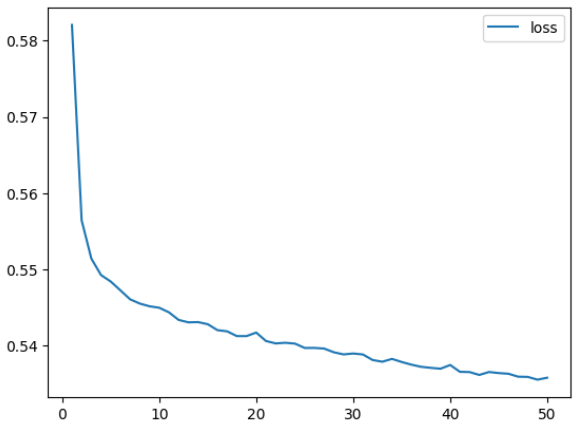
Number of neurons in first hidden layer: **27**

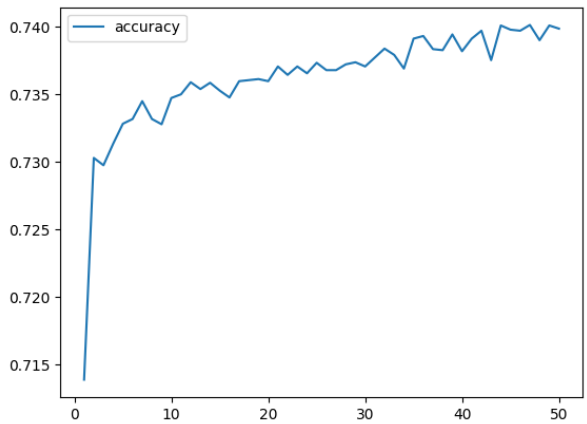
Number of neurons in second hidden layer: **54**

Hidden layer activation function: **relu**

Output layer activation function: **sigmoid**

Number of Epochs: **50**





Loss: 0.5482215881347656 Accuracy: 0.7315452098846436

Deep Learning Analysis Report

Module 21 Challenge – Alphabet Soup Charity

Model Attempt 3

Number of hidden layers: **3**

Number of input parameters: **108**

Number of neurons in first hidden layer: **15**

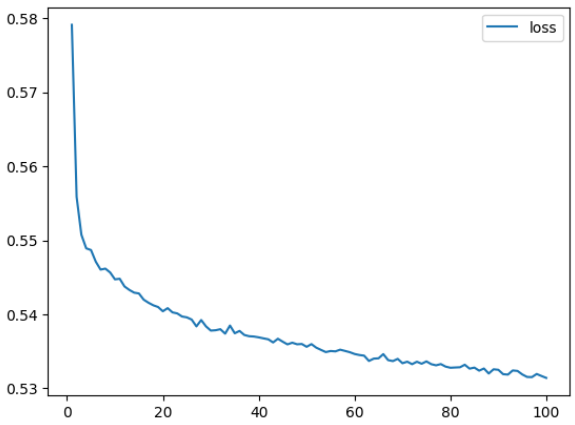
Number of neurons in second hidden layer: **30**

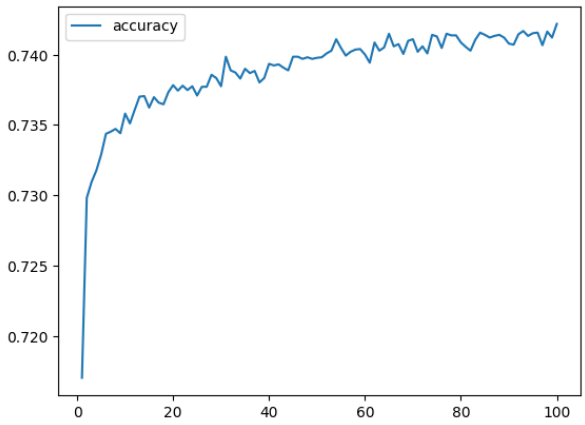
Number of neurons in third hidden layer: **60**

Hidden layer activation function: **relu**

Output layer activation function: **sigmoid**

Number of Epochs: **100**





Loss: 0.556148111820221 Accuracy: 0.7315452098846436

Model Attempt 4

Number of hidden layers: **3**

Number of input parameters: **108**

Number of neurons in first hidden layer: **5**

Number of neurons in second hidden layer: **7**

Number of neurons in third hidden layer: **9**

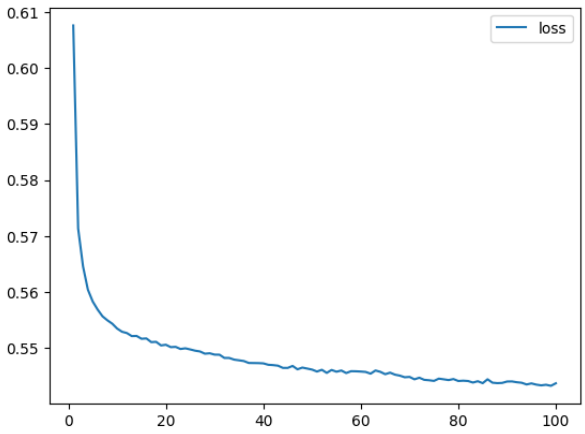
Deep Learning Analysis Report

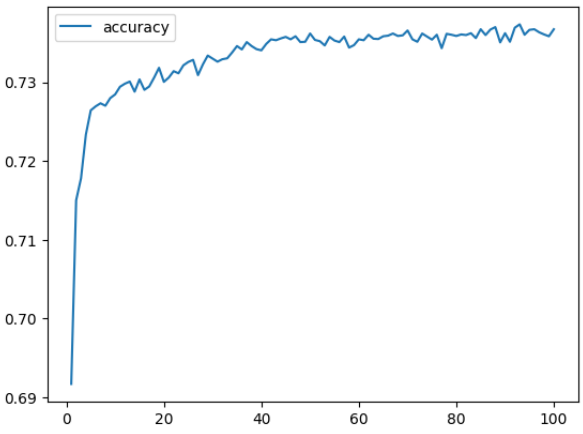
Module 21 Challenge – Alphabet Soup Charity

Hidden layer activation function: **relu**

Output layer activation function: **sigmoid**

Number of Epochs: **100**





Loss: 0.55184006690979 Accuracy: 0.7306122183799744

Summary of Analysis

The modelling undertaken during this analysis indicates that significant alterations to the number of nodes, addition of a third layer and using epochs of both 50 and 100 iterations has minimal impact on the accuracy and loss for this dataset. Although the aim was to achieve an accuracy of 0.75, this was not obtained during the current analysis. The current analysis was limited in its extent. The number of input dimensions used during this analysis has not been varied, nor has there been any change to the original feature set with respect to which columns are included or excluded from the analysis.

In addition to changing the number of columns to be included in the model, variations to the number bins used and the number of values per bin has not been performed as part of this analysis. Further modification of the model could be performed by changing the hidden layer activation function, or the output activation function.