Andreas Hill

07/26/2023

**Deep Learning Project: Alphabet Soup Charity Funding Predictor**

Deep Learning and Neural Networks are used to determine if applicants would be funded successfully by Alphabet Soup.

**Data Processing**

The dataset removed any irrelevant information; therefore, EIN and NAME were dropped from the model. The remaining columns were considered features for the model. Although NAME was added back in the second test. CLASSIFICATION and APPLICATION\_TYPE was replaced with ‘Other due to high fluctuation. The data was split into training and testing sets of data. The target variable for the model is “IS\_SUCCESSFUL” and is verified by the value, 1 was considered yes and 0 was no. APPLICATION data was analyzed, and CLASSIFICATION’s value was used for binning. Each unique value used several data point as a cutoff point to bin “rare” categorical variables together in a new value, ‘Other’. Afterwards checked to see if binning was successful. Categorical variables were encoded by ‘pd.get\_dummies().

**Compiling, Training, and Evaluation the Model**

A screenshot of a computer program

Description automatically generatedNeural Network was applied on each model multiple layers, three in total. The number of features dictated the number of hidden nodes. As seen below, a three-node model generated 477 parameters and the first attempt came to 73% accuracy.

A close-up of a computer

Description automatically generated

A screenshot of a computer code

Description automatically generated

**Optimizing the Model**

A screenshot of a computer program

Description automatically generatedFor the second attempt the “NAME” column was included in the data frame and resulted in an increase in total parameters at 3,298. Three hidden nodes was used again and the outcome was an accuracy of 79%.

**Possible Alternative Models**

Trying a Random Forest classifier may be better as it is robust to outliers and can handle both numerical and categorical features well. By ensemble learning and decision tree-based approaches, it can capture complex relationships in the data. Additionally, for efficiency, it could be better to reduce the number of epochs to a range between 20-50, as this could still achieve satisfactory results while potentially reducing training time.