Deep Learning Report

Overview

This report will discuss various elements of the Module 21 Neural Networks assignment. It will provide an overview of the assignment's objectives, detail the preprocessing steps undertaken, and present an analysis of the results obtained from testing four different neural network variants. The assignment involved importing a CSV file named "charity_data.csv" in Google Colab, preprocessing the data, and training a neural network to predict factors contributing to the success of charity campaigns. The CSV contained 34,299 rows of data without any null values, primarily consisting of integer and float data types, with other appropriate types used for specific characteristics.

Results

Data Preprocessing

After importing dependencies and reading in the CSV, I began preprocessing. Following instructions, I dropped the unimportant columns “EIN” and “NAME”, then determined the number of unique values in each column. Due to the “APPLICATION_TYPE” column having many unique values with only 1 or 2 entries, they were merged into a new “Other” label with a cutoff of 500. This was repeated for the “CLASSIFICATION” column, this time with a cutoff of 1800.

After consolidating the DataFrame, I used .get_dummies() to encode categorical variables. Then, split the data into a features array, x, and a target array, y. These were used to create the train_test_split function to create the training and testing datasets. Finally, the StandardScaler was applied to the training and testing datasets.

Compiling, Training, and Evaluating the Model

Model 1:

Dropped “EIN” and “NAME” columns
2 Hidden layers; 5 nodes for layer 1, 1 node for layer 2
ReLU (Rectified Linear Unit
100 Epochs
73.04% accuracy with 55.63% loss

Model 2:

Dropped “EIN” and “NAME” columns
3 Hidden layers; 5 nodes for layer 1, 3 nodes for layer 2, 1 node for layer 3
ReLU (Rectified Linear Unit
100 Epochs
72.44% accuracy with 55.97% loss

Model 3:

Dropped “EIN” and “NAME” columns
3 Hidden layers; 7 nodes for layer 1, 5 nodes for layer 2, 3 nodes for layer 3
ReLU (Rectified Linear Unit
100 Epochs
72.29% accuracy with 54.93% loss

Model 4:

Dropped “EIN” and “NAME” columns
4 Hidden layers; 99 nodes for layer 1, 69 nodes for layer 2, 39 nodes for layer 3, 9 nodes for layer 4
ReLU (Rectified Linear Unit
100 Epochs
73.12% accuracy with 56.73% loss

Summary

All models performed about the same. Accuracy was around 73% and loss was around 55%. After the large number of nodes in the final model, it led me to believe more nodes was not going to help. For further refinement of the models, I would want to look at more preprocessing, starting with changing my cutoffs, and possibly dropping other columns.