

DEEP LEARNING CHALLENGE

Neural Network Model Report

1. Project Overview:

Alphabet Soup (A/S), a non-profit organization, seeks to leverage its historical funding data, which spans over 34,000 organizations they've supported over the years. They aim to develop a tool that will help identify applicants most likely to succeed in future ventures. This tool will be built using machine learning and neural network techniques.

2. Dataset

The dataset for this assignment is provided in a .csv file, accessible online at the following link: https://static.bc-edx.com/data/dl-1-2/m21/lms/starter/charity_data.csv. The Jupyter Notebook contains code that reads the .csv file, making the data available for analysis.

3. Data Cleaning

After reading the .csv file, data preprocessing was performed, which involved the following steps:

- Dropping unnecessary columns
- Identifying unique values
- Recategorizing data, such as creating an "Other" category using a cutoff value
- Converting categorical data into numerical format using `pd.get_dummies`

Next, the data was split into feature and target arrays, followed by scaling using `StandardScaler`

4. Results and Analysis:

a. Data Preprocessing:

(i) What variable(s) are the target(s) for your model?

The target for the model is the column "IS_SUCCESSFUL," which contains binary values (0 or 1) indicating whether the funds were used effectively.

(ii) What variable(s) are the features for your model?

The features for the model include: STATUS, ASK_AMOUNT, APPLICATION_TYPE, AFFILIATION, INCOME_AMOUNT, and SPECIAL_CONSIDERATIONS.

(iii) What variable(s) should be removed from the input data because they are neither targets nor features?

The variables to be removed are: EIN, NAME, CLASSIFICATION, USE_CASE, and ORGANIZATION.

b. Compiling, Training, and Evaluating the Model:

- Initial model has 3 layers (35 neurons, 35 neurons, and 1 neuron) using RELU as activation function for two layers and SIGMOID for the last, with 100 Epochs. This initial model did not meet model performance, as the Accuracy was calculated at 0.7293 and the Loss calculated at 0.5618
- The first modification was to increase the number of Epochs to 200, but this lowered the Accuracy score to 0.7282 and increased the Loss to 0.5935
- The second modification, the number of Epochs was returned to 100 and a fourth layer of 35 neurons (with RELU) was added. The Accuracy score finished 0.7286 and Loss equaling 0.5611.
- Finally, the third modification, returned the model to 3 layers (35 neurons, 35 neurons, and 1 neuron) and changed the activation functions to SIGMOID for all layers. The Accuracy score finished as 0.7321, with Loss equaling 0.5544.

5. Summary

This tool was designed to explore whether insights could be derived from historical data. However, none of the iterations (the original version plus three modifications) achieved the desired threshold of 75%.

While having access to this data is beneficial, it may be worth discussing whether additional information from A/S could further enhance this model. A thorough review of each column's content is also recommended to ensure all relevant data is included in the tool's setup.

In the data preprocessing phase, two columns (APPLICATION_TYPE and CLASSIFICATION) were assigned arbitrary cutoff values, which could affect how the model interprets the data. Re-evaluating these cutoff values may be advantageous in future iterations of the tool.

6. Ethical consideration:

The original spreadsheet contains potentially sensitive corporate financial information, so it's important to handle the dataset with care. After preprocessing, corporate identities have been removed, ensuring that the results, analysis, and output no longer include sensitive information