Module 21 - Report on the Neural Network Model

OVERVIEW

The neural network was testing and predicting if an organization funded by Alphabet Soup would be successful or not. This report is examining the neural network and describing the variables, model, and results from the original test run and 3 attempts to optimize the model.

DATA PREPROCESSING

The variables that were factored into the models were Application Type, Affiliation, Classification, Use Case, Organization, Status, Income Amount, Special Considerations, and Ask Amount. There were all features in the model except the last optimization. The 3rd optimization removed the Status (only 5 rows had a different status amount) and Special Considerations (only 26 rows had a special consideration). Those two variables should have been removed from all the attempts.

COMPILING, TRAINING, & EVALUATING THE MODEL

Model	Neurons	Layers	Activation
Original	86,43,1	3	Relu, relu, sigmoid
1st Optimization	74,37,1	3	Relu, relu, sigmoid
2nd Optimization	68, 34, 10, 1	4	Relu, relu, relu, sigmoid
3rd Optimization	80, 40, 15, 1	4	Relu, relu, relu, sigmoid

I was not able to achieve target model performance. 1st and 2nd attempts at optimization ended up being worse than the original model.

Original model was 72.5% accurate with a loss 56.7%.

1st Optimization added to the original model. Grouped the 'ASK_AMT' column in hopes it would help the accuracy. The 'ASK_AMT' had a lot of different unique values but grouping into just 2 parts did not help accuracy. It was better to not group the column and leave it the way it was because the accuracy dropped to 70.9% and a loss of 58.8%.

2nd Optimization added to the original model. Grouped the 'INCOME_AMT' and added one more layer. 'INCOME_AMT' was over 70% the value of $5,000. Grouping into 70/30 split was not more efficient than the original model even with an extra layer (71.4% accuracy, 57.7% loss).

3rd Optimization was very simple. This should have been the 1st optimization and then had 2nd and 3rd work on this one and then maybe I could have hit the 75% and over mark. Data was cleaned just a little and then 2 columns were removed. 'STATUS' column all rows, except 5, had the same value. I removed those 5 rows and then remove the 'STATUS' column. 'SPECIAL_CONSIDERATIONS' column all rows had the same value except 26 of them. Those 26 rows were dropped and then the 'SPECIAL_CONSIDERATION' column was dropped as well. In addition, an additional layer was added. This slightly improved the accuracy over the first model (73.1% accuracy, 55.9% loss)

SUMMARY

Overall the Optimization attempts were failures or just a small improvement. I do believe removing the Status and Special_Considerations columns were correct but it was just affecting a small piece of the data. I do think the ASK_AMT with over 80 unique values should be segmented better and would suggest using unsupervised learning to find the optimal segmentation number. I also think the very highly funded applicants should be taken out and be treated as different and in their own dataset. Even though 70% accuracy is not bad, but money for a nonprofit can be tight and model should be worked on more and further optimized to find a prediction model that is more accurate and can give Alphabet Soup a higher chance of success for each applicant it funds.