Analysis of Neural Network Performance

Overview of the Analysis

The purpose of this analysis is to evaluate the performance of a deep learning model developed to predict the success of organizations applying for funding from the Alphabet Soup Foundation. The dataset contains information about over 34,000 organizations that have previously received funding. The foundation wants to select future applicants who are most likely to succeed in their ventures based on this model.

Data Preprocessing

- **Target Variable**:
  The target variable for the model is IS_SUCCESSFUL, which indicates whether an organization was successful in securing funding.

- **Feature Variables**:

  The following features were used in the model:

  APPLICATION_TYPE

  AFFILIATION

  CLASSIFICATION

  USE_CASE

ORGANIZATION

STATUS

INCOME_AMT

SPECIAL_CONSIDERATIONS

ASK_AMT

- **Variables to Remove**:
  The columns EIN and NAME were removed from the dataset as they do not provide meaningful information for predicting the target variable.

Compiling, Training, and Evaluating the Model

First Model

Number of layers: 2

Number of nodes: 50 and 40

Activation function: ReLU for hidden layers, Sigmoid for output

Epochs: 100

Accuracy: 0.7296

We can see that we could not achieve an accuracy of 75%.We would be exploring different models by changing the number of input layers ,the number of nodes and the activation hyperparameter to see if we could reach an accuracy of 75%.

Second Model(Increasing the input layers)

Number of input layers: 3

Number of nodes: 80, 40, and 30

Activation function: ReLU for hidden layers, Sigmoid for output

Epochs: 150

Accuracy: 0.7287

This model also performed similarly to the first one, failing to reach the target accuracy.

Third Model(Increasing the nodes)

Number of input layers: 2

Number of nodes: 100 and 40

Activation function: ReLU for hidden layers, Sigmoid for output

Epochs: 50

Accuracy: 0.7293

The third model did not improve performance either.

Fourth Model(Changing the activation hyperparameter and increasing nodes)

Number of layers: 2

Number of nodes: 300 and 80

Activation function: Tanh for hidden layers, Sigmoid for output

Epochs: 50

Accuracy: 0.7282

Again, this model also failed to reach the desired 75% accuracy.

Model Evaluation and Correlation Analysis

The model's performance was evaluated by calculating the correlation between the target variable (IS_SUCCESSFUL) and the features. Some of the most significant features include:

AFFILIATION_Independent: 0.368

ORGANIZATION_Trust: 0.162

CLASSIFICATION_Other: 0.111

APPLICATION_TYPE_T5: 0.088

APPLICATION_TYPE_T10: 0.086

On the other hand, the least significant features included:

AFFILIATION_CompanySponsored: -0.369

APPLICATION_TYPE_T19: -0.123

INCOME_AMT_50M+: -0.0248

After evaluating the correlations,I felt that no specific feature could be dropped to significantly improve model performance.

This suggests that the deep learning model might require further adjustments.

Summary

The deep learning model successfully predicted organizational success based on the provided features, but it did not meet the target accuracy of 75%. Various models with different layers, nodes, and activation functions were tested, but there was no significant improvement.

Recommendation

I felt that using a Random Forest or Gradient Boosting model could be a better choice instead of neural networks.These models are better at handling complex relationships and correlations within the dataset. Random Forests, for example, are less sensitive to overfitting and can perform better with the combination of categorical and numerical features.