Deep Learning Model Analysis: Alphabet Soup Charity Funding Prediction

Overview of the Analysis

The goal of this project is to build a binary classification model to predict whether organizations funded by Alphabet Soup will be successful. By leveraging deep learning techniques using TensorFlow and Keras, we created and optimized a neural network model to improve prediction accuracy.

Results

Data Preprocessing

* Target Variable: The target variable for the model is IS\_SUCCESSFUL, which indicates whether the funding was effectively utilized.
* Feature Variables:
  + APPLICATION\_TYPE (Application category)
  + AFFILIATION (Industry sector)
  + CLASSIFICATION (Government classification)
  + USE\_CASE (Funding use case)
  + ORGANIZATION (Type of organization)
  + STATUS (Organization’s operational status)
  + INCOME\_AMT (Income classification)
  + SPECIAL\_CONSIDERATIONS (Special factors impacting approval)
  + ASK\_AMT (Funding amount requested)
* Removed Variables: The EIN and NAME columns were dropped since they are identifiers and do not contribute to the model’s predictive power.
* Categorical Encoding:
  + Categorical variables were encoded using pd.get\_dummies() to convert them into numerical values.
  + Rare categorical values were grouped into an "Other" category to simplify feature representation and improve model generalization.
* Data Splitting and Scaling:
  + The data was split into training and testing sets using train\_test\_split().
  + The features were standardized using StandardScaler() to ensure consistent input magnitudes.

Model Compilation, Training, and Evaluation

* Neural Network Architecture:
  + Input Layer: Based on the number of encoded feature variables.
  + Hidden Layers:
    - First Hidden Layer: 80 neurons, ReLU activation function.
    - Second Hidden Layer: 30 neurons, ReLU activation function.
  + Output Layer:
    - 1 neuron with a sigmoid activation function to generate binary predictions (0 or 1).
* Compilation and Training:
  + The model was compiled using the Adam optimizer and binary cross-entropy loss function.
  + It was trained over 50 epochs, with a callback to save weights every five epochs.
* Performance Evaluation:
  + The model achieved an accuracy of 72%, which was below the 75% target.

Model Optimization

To improve accuracy, several optimizations were applied:

1. Increased Neurons: Hidden layer neurons were increased to 100 and 50, respectively.
2. Additional Hidden Layer: A third hidden layer with 20 neurons was added to enhance learning capability.
3. Changed Activation Functions: Experimented with tanh and LeakyReLU to improve feature extraction.
4. Adjusted Training Epochs: Increased to 75 epochs for better convergence.
5. Tuned Dropout Layers: Introduced dropout regularization to prevent overfitting.

After optimization, the model achieved an accuracy of 76.5%, surpassing the initial goal.

Summary and Recommendations

The optimized deep learning model successfully classified organizations with a 76.5% accuracy, making it a valuable tool for Alphabet Soup’s funding decisions. Overall, deep learning proved to be an effective approach for this problem, and further tuning or ensemble techniques may further improve its performance.