

Defined Architecture of ANN Model:

- Specify the architecture of the ANN model, including the input, hidden, and output layers optimised for predictive performance

The architecture of the Artificial Neural Network (ANN) model defined in the provided code is equipped with the following layers. Now, We are aiming to balance the predictive performance and the ability to generalize across the dataset. Breakdown of layers are below:

ANN Model Architecture:

1. Input Layer:

- **Input Shape:** (X_train_scaled_df.shape[1],)
 - The input shape corresponds to the number of features in the data(training data). For example, if we have 10 features i.e., columns in our preprocessed dataset then our input shape would be 10.

2. Hidden Layers:

- **First Hidden Layer:**
 - **Number of Neurons:** 64
 - **Activation Function:** ReLU (Rectified Linear Unit)
 - **Details:** This layer contains 64 neurons and a ReLU activation function, which allows the network to learn nonlinearity and enables the network to learn complicated or hidden patterns in the data.
- **Second Hidden Layer:**
 - **Number of Neurons:** 32
 - **Activation Function:** ReLU
 - **Details:** This layer contains 32 neurons as each layer learns different features to create better representation of data, also using the ReLU activation function to further process output of the first layer.
- **Third Hidden Layer:**
 - **Number of Neurons:** 128
 - **Activation Function:** ReLU
 - **Details:** This layer has 128 neurons with ReLU activation which makes it the largest hidden layer. It helps the model to learn more complex representations.
- **Dropout Layer:**
 - **Dropout Rate:** 0.3 (30%)

- **Details:** A dropout layer is added next to hidden layer to reduce the overfitting by randomly putting a fraction of input units to 0 while training.
 - **Fourth Hidden Layer:**
 - **Number of Neurons:** 64
 - **Activation Function:** ReLU
 - **Details:** Another layer with 64 neurons and ReLU activation function, following the dropout layer to continue the learning process while maintaining the regularization.
 - **Dropout Layer:**
 - **Dropout Rate:** 0.3 (30%)
 - **Details:** We add another dropout layer to prevent overfitting.
3. **Output Layer:**
- **Number of Neurons:** 1
 - **Activation Function:** Sigmoid
 - **Details:** The output layer is with a single neuron with a sigmoid activation function. This is typical for tasks which involve binary classification (like predicting churn), where the output is a probability score between 0 and 1, representing the likelihood of a customer churning.

Summary of Model Layers:

- **Input Layer:** Dimension matching the number of input features.
- **Hidden Layers:**
 - Layer 1: 64 neurons, ReLU activation
 - Layer 2: 32 neurons, ReLU activation
 - Layer 3: 128 neurons, ReLU activation
 - Dropout Layer: 30% dropout rate
 - Layer 4: 64 neurons, ReLU activation
 - Dropout Layer: 30% dropout rate
- **Output Layer:** 1 neuron, Sigmoid activation

Optimizations and Considerations:

- **Activation Functions:** ReLU is used in hidden layers to handle the vanishing gradient problem and efficiently learn complex patterns.
- **Dropout:** Dropout layers are strategically placed to prevent overfitting, especially since neural networks with many layers and neurons can easily overfit on smaller datasets.

- **Output Layer:** A sigmoid activation function is employed in the output layer to produce a probability that can be used for binary classification decisions.