Title: Neural Networks and Exploratory Data Analysis on the Iris Dataset

Total Marks: 20

Objective:

In this assignment, you'll explore the Iris dataset and build a simple neural network to classify different types of Iris flowers. This will help you get hands-on experience with data analysis, data preprocessing, and building neural networks.

Data:

The Iris dataset includes 150 samples of Iris flowers from three different species: Iris setosa, Iris versicolor, and Iris virginica. Each flower has four measurements: sepal length, sepal width, petal length, and petal width.

Tasks:

1. **Data Exploration (6 marks):**

- Load the Iris dataset using `pandas` or `sklearn`. *(1 mark)*

- Perform some basic data analysis to understand the dataset:

  - Create histograms for each of the four features. *(2 marks)*

  - Make scatter plots to show how the features relate to each other. *(2 marks)*

  - Calculate key statistics like the average, median, and standard deviation for each feature. *(1 mark)*

2. **Data Preprocessing (3 marks):**

- Adjust the data so that all features have a similar scale (normalize them). *(1.5 marks)*

- Convert the species names into numbers so they can be used by the neural network (one-hot encoding). *(1.5 marks)*

3. **Building the Neural Network (4 marks):**

- Use `TensorFlow` or `Keras` to create a simple neural network with one hidden layer.

- The neural network should include:

- An input layer that matches the number of features (4). *(1 mark)*

- A hidden layer with 8 to 16 neurons. *(1.5 marks)*

- An output layer with three neurons, one for each Iris species. *(1.5 marks)*

4. **Training the Model (3 marks):**

- Split the data into training and testing sets (e.g., 80% for training, 20% for testing). *(1 mark)*

- Train the neural network on the training data for 20-50 rounds (epochs). *(1 mark)*

- Plot how the accuracy and loss change during training and validation. *(1 mark)*

5. **Model Evaluation (2 marks):**

- Test the trained model on the test data and report its accuracy. *(1 mark)*

- Create a confusion matrix to see how well the model is performing on each species. *(1 mark)*

6. **Discussion (2 marks):**

- Talk about how well your model worked and any patterns or problems you noticed during training. *(1 mark)*

- Suggest ways to improve the model or what changes could be made to get better results. *(1 mark)*

Bonus (Optional):

- Try using different numbers of neurons in the hidden layer or add another hidden layer to see how it affects the model's performance.

- Visualize the decision boundaries of the neural network by projecting the features into 2D space.