Lab-6, Neural Networks

Objective: To familiarize students with the application of neural networks in RapidMiner for basic data analysis and predictive modeling.

Task 1:

1. Data Loading and Exploration:
    a. Load the Sonar (or Ripley) dataset into RapidMiner.
    b. Explore basic statistics: number of instances, attributes, class distribution, etc.
2. Data Preprocessing:
    a. Normalize the data (e.g., z-score normalization) so that all features are on a similar scale.
    b. Split the data into training (70%) and testing (30%) sets.
3. Neural Network Configuration:
    a. Initiate the Neural Net operator from the 'Modeling' section.
    b. Configure the neural network:
        i. Set the number of hidden layers and neurons.
        ii. Choose a suitable activation function (e.g., Sigmoid).
        iii. Set other parameters: epochs, learning rate, etc.
4. Training the Neural Network:
    a. Train the neural network on the training dataset.
    b. Observe the error convergence through the epochs.
5. Evaluation:
    a. Use the Performance operator to evaluate the trained model on the testing dataset.
    b. Compare results: accuracy, precision, recall, F1-score, etc.
6. Hyperparameter Tuning:
    a. Adjust various hyperparameters of the neural network (e.g., number of neurons, learning rate, epochs).
    b. Train and evaluate the model again to observe any improvements.
7. Visualization:
    a. Use the Plot View operator to visualize the neural network architecture.
    b. Use scatter plots to visualize actual vs. predicted values.
8. Comparison with Other Models:
    a. Use the Decision Tree or any other classifier available in RapidMiner.
    b. Train and evaluate the alternative model on the same dataset.
    c. Compare the performance of the neural network against this model.
9. Understanding Overfitting:
    a. Train the neural network with an excessive number of neurons and/or layers.
    b. Observe the performance on training vs. testing datasets to understand the concept of overfitting.
10. Final Discussion and Reflection:
    a. Discuss the benefits and drawbacks of using neural networks.
    b. Reflect on scenarios where neural networks might be the preferred choice over other algorithms.
    c. Explore potential use cases beyond the Iris dataset.

Task 2:

Dataset: Wine Quality dataset (Contains attributes like fixed acidity, volatile acidity, citric acid, residual sugar, etc., and a quality score between 0 to 10).

1. Data Loading and Exploration:

   a. Load the Wine Quality dataset into RapidMiner.
   b. Explore basic statistics: number of instances, attributes, quality distribution, etc.
   c. Visualize relationships between some key attributes and wine quality using scatter plots.
2. Data Preprocessing:

   a. Handle any missing values, if present.
   b. Normalize or standardize the data to bring all features to a similar scale.
   c. Split the data into training (70%) and testing (30%) sets.
3. Neural Network Configuration:

   a. Initiate the Neural Net operator from the 'Modeling' section. You can also try the deep learning operator.
   b. Configure the neural network with one or more hidden layers.
   c. Set appropriate activation functions, epochs, learning rate, etc.
4. Training the Neural Network:

   a. Train the neural network on the training dataset.
5. Evaluation:

   a. Evaluate the trained model on the testing dataset.
   b. Focus on metrics like Mean Absolute Error (MAE) or Root Mean Square Error (RMSE) given the regression nature of the task.
6. Hyperparameter Tuning:

   a. Experiment with different configurations (number of neurons, layers, learning rate) to optimize performance.
   b. Retrain and re-evaluate the model to observe improvements.
7. Comparison with Other Models:

   a. Use the Decision Tree or Linear Regression operator in RapidMiner.
   b. Train and evaluate the alternative model on the same dataset.
   c. Compare the performance of the neural network against these simpler models.
8. Feature Importance Analysis (Advanced and optional):

   a. Investigate which features are most influential in predicting wine quality.
   b. Compare the importance of features as identified by the neural network with those identified by simpler models like Decision Trees.
9. Final Discussion and Reflection:

   a. Discuss the advantages and challenges of using neural networks for this task.
   b. Reflect on how preprocessing or feature