Assignment 5

Lambton College, Mississauga

Neural Networks and Deep Learning

Prof. Name: Ishant Gupta

Name – Kelvin Simon

Student id – c0866577

Detailed Report: Predicting House Prices in California Housing Dataset

Objective:

This comprehensive report documents the intricate process of predicting house prices using the California Housing dataset. Our focus involves the implementation and thorough comparison of the performance between Linear Regression and an Artificial Neural Network (ANN) for this regression task.

1. Data Preprocessing:

1.1 Dataset Exploration:

Loaded the California Housing dataset, delving into features, and gaining a profound understanding of the dataset's structure.

1.2 Handling Missing Values and Outliers:

A careful examination for missing values was conducted.

Explored and adeptly managed outliers to ensure data integrity.

1.3 Splitting the Dataset:

The dataset was judiciously split into training and testing sets, allocating **80%** for training and **20%** for testing.

2. Linear Regression:

2.1 Model Implementation:

A robust Linear Regression model was implemented utilizing scikit-learn.

2.2 Training and Prediction:

Methodically trained the Linear Regression model on the training set.

Executed precise predictions on the testing set.

2.3 Performance Evaluation:

Model performance was meticulously evaluated:

Mean Squared Error (MSE): 0.556

R2 Score: 0.576

3. Artificial Neural Network (ANN):

3.1 Model Implementation:

An intricate ANN for regression was crafted using TensorFlow/Keras, featuring one hidden layer with 64 neurons and an output layer.

3.2 Training and Prediction:

Conducted meticulous training of the ANN on the training set.

Generated predictions with accuracy on the testing set.

3.3 Performance Evaluation:

The ANN's performance underwent detailed scrutiny:

Mean Squared Error (MSE): 0.317

R2 Score: 0.758

4. Comparison and Analysis:

4.1 Performance Metrics:

Conducted a meticulous comparison of performance metrics between Linear Regression and the ANN. Impressively, the ANN outperformed in terms of both MSE and R2 Score.

4.2 Model Strengths and Weaknesses:

Linear Regression: Exhibits simplicity, interpretability, and efficiency, yet constrained by the linearity assumption and limited complexity.

ANN: Showcases the ability to capture non-linear patterns, flexibility, albeit at the expense of being computationally expensive and less interpretable.

4.3 Model Complexity:

In-depth analysis revealed that the complexity inherent in the ANN architecture provided significantly enhanced predictive performance.

5. Visualization:

Generated insightful scatter plots, comparing predicted values of Linear Regression and the ANN against actual values. Additionally, an illuminating visualization of the ANN architecture was created.

6. Conclusion:

A concise summary emphasized the key findings, underscoring the ANN's superior performance. Comprehensive insights were provided into the strengths, weaknesses, and practical considerations of both models, concluding with a thoughtful discussion on the interpretability versus accuracy trade-off.