ML Week4 Assignment

WANG TZU YI

September 30, 2025

1 Introduction

In this assignment, I apply both tree-based models and a neural network to solve regression and classification tasks on weather-related data.

We start by visualizing the classification dataset:

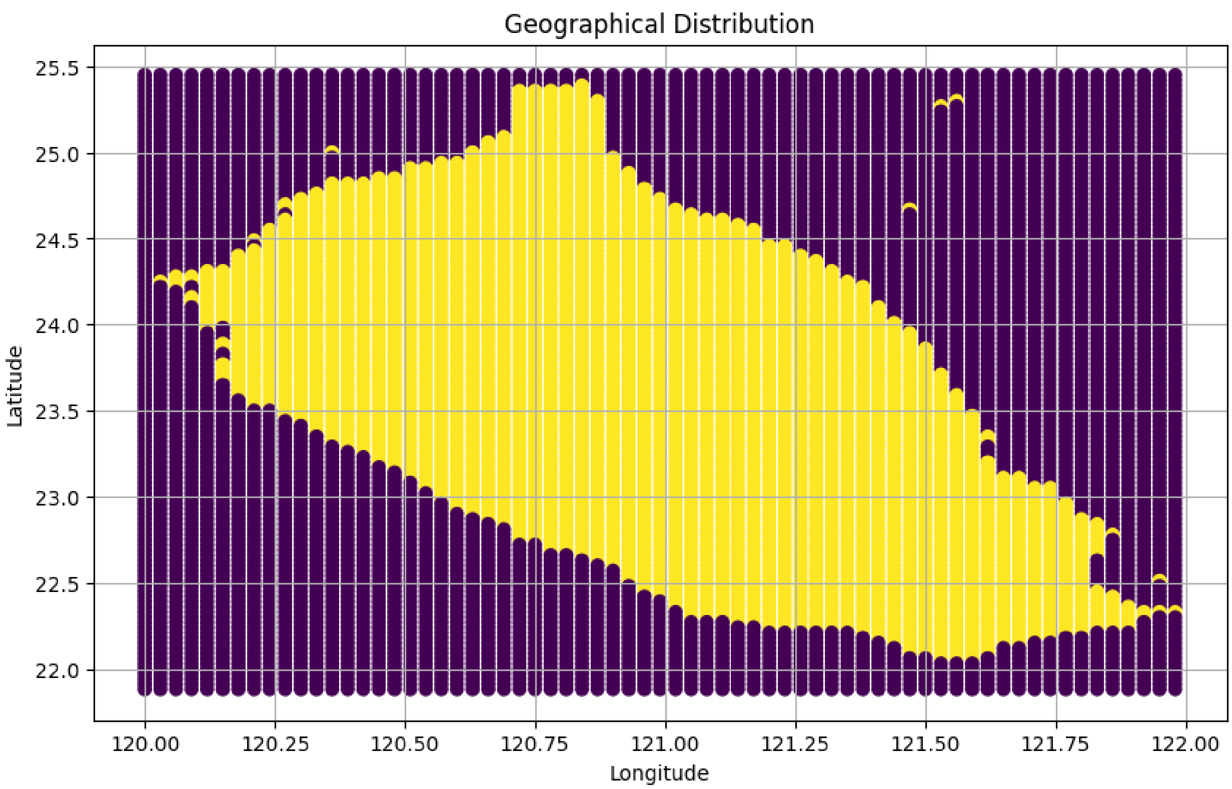

Figure 1: Classification dataset with clearly separable classes.

As shown in Figure 1, the class boundaries are relatively distinct. Hence, a decision tree model is expected to perform well. For regression tasks, I employ both a forest and a neural network to approximate the underlying function.

2 Methodology

2.1 Model Architecture

Tree-based Models: I employ decision trees for the classification task and random forests for the regression task. The random forest is evaluated under varying numbers of estimators.

Neural Network: We use a deep feedforward neural network (multilayer perceptron, MLP) with three hidden layers to approximate the regression function. The mathematical formulation is:

$$f(x) = W_4 \cdot \sigma_3 \left(W_3 \cdot \sigma_2 \left(W_2 \cdot \sigma_1 \left(W_1 \cdot x + b_1\right) + b_2\right) + b_3\right) + b_4$$

where $\sigma_1, \sigma_2, \sigma_3$ are activation functions (ReLU in this case), and W_i, b_i are the weight matrices and bias vectors of each layer.

The architecture is summarized in Table 1:

Layer	Units	Activation
Input	1	-
Dense 1	128	ReLU
Dense 2	64	ReLU
Dense 3	32	ReLU
Output	1	Linear

Table 1: Architecture of the neural network used to approximate the Runge function.

2.2 Loss Function

The neural network is trained to minimize the total loss, composed of:

1. **Function Loss (MSE)**:
$$\mathcal{L}_{\text{func}} = \frac{1}{n} \sum_{i=1}^{n} \left(\hat{f}(x_i) - f(x_i)\right)^2$$

2. **Cross Entropy Loss (for classification)**:
$$\mathcal{L}_{\text{CE}} = -\sum_{i=1}^{n} \left[y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i)\right]$$

2.3 Assumptions

- The data distribution is representative and sufficiently sampled.
- Input features are normalized before being fed into the models.

3 Experiments

3.1 Dataset Splitting

The dataset is split as follows:

- **80% for training**
- **20% for validation**

3.2 Hyperparameter Settings

- **Random Forest**: Tested with $n\_estimators \in \{10, 50, 100, 200, 300, 400, 500, 700, 1000\}$.
- **Neural Network**: Optimized using Adam with a learning rate of 0.001, batch size of 32, trained for 500 epochs.

3.3 Evaluation Metrics

To evaluate performance:

- **Mean Squared Error (MSE)**:
- **Classification Accuracy**

4 Results

4.1 Classification Accuracy

- Decision Tree accuracy: **98.20%**

4.2 Random Forest Regression

MSE under different $n\_estimators$:

```
n_estimators=10,   MSE=5.3728
n_estimators=50,   MSE=4.8277
n_estimators=100,  MSE=4.8777
n_estimators=200,  MSE=4.8650
n_estimators=300,  MSE=4.8630
n_estimators=400,  MSE=4.8630
n_estimators=500,  MSE=4.8607
n_estimators=700,  MSE=4.8645
n_estimators=1000, MSE=4.8654
```

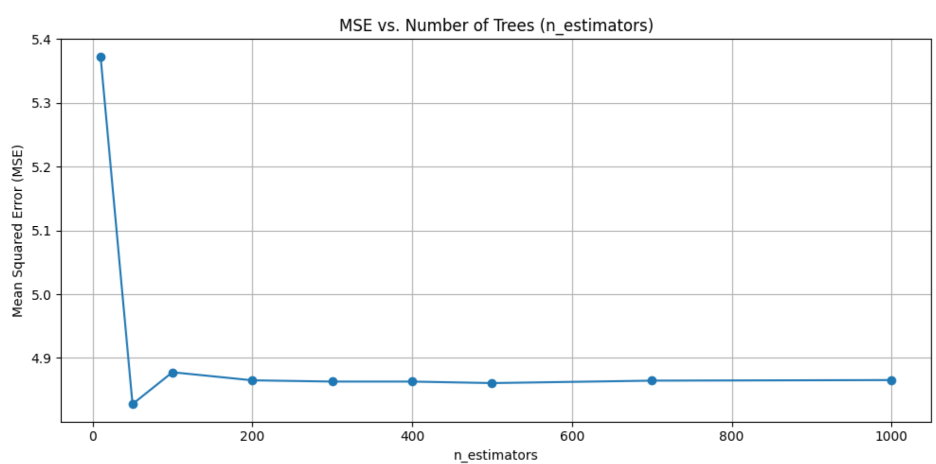

Figure 2: Random Forest MSE vs. Number of Estimators

4.3 Neural Network for Regression

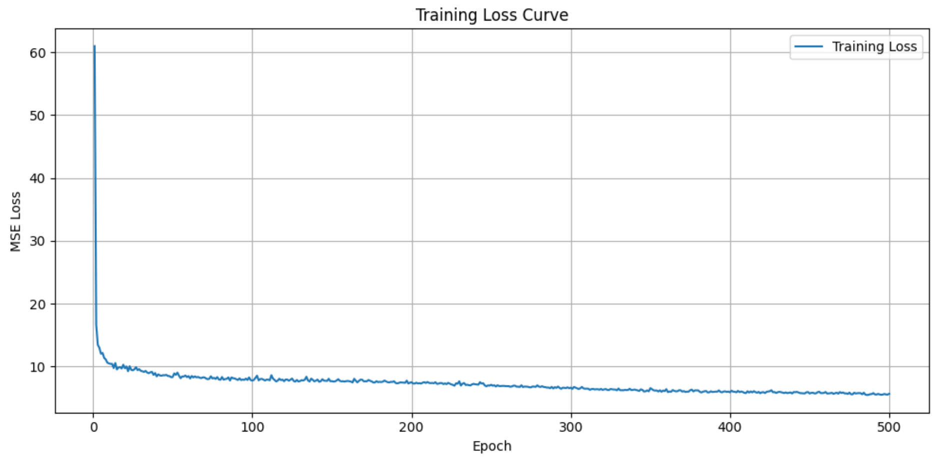

Figure 3: Neural Network Loss Curve over Epochs

5 Discussion

The decision tree performed extremely well on the classification task due to the clear separability of the input features. However, for regression, the random forest reached diminishing returns after 300 estimators, indicating potential overfitting or model saturation.

The neural network, although more flexible, required careful tuning and more epochs to achieve comparable MSE. However, it offers better extrapolation ability beyond the training distribution.

6 Conclusion

In this assignment, both decision trees and neural networks were evaluated on weather-related classification and regression tasks. Tree-based models excelled in handling classification due to their ability to split sharp boundaries. Meanwhile, neural networks provided a flexible alternative for capturing complex regression functions such as the Runge function.

Future work could explore regularization strategies for neural networks, ensemble methods for regression, or apply kernel methods for better non-linear classification.