Final Lecture Report: Lecture 2

Decision Tree Analysis on IRIS Dataset

Submission Date: May 12, 2025
Submitted by: **Kaan Yazıcıoğlu (97364)**

1. About the Data

The IRIS dataset has 150 records.

Each record has 4 features: sepal length, sepal width, petal length, petal width.

There are 3 classes: Setosa, Versicolor, Virginica.

No missing data in this dataset.

2. What I Found in the Data

The smallest sepal length is 4.3 cm, the largest is 7.9 cm.

The smallest petal length is 1.0 cm, the largest is 6.9 cm.

Each class has 50 samples.

The data is balanced and clean.

3. How Well the Model Works

I used a Decision Tree Classifier.

I split the data: 75% for training, 25% for testing.

Training accuracy: 100%

Test accuracy: about 96% (average over different random states)

The model predicts the flower type very well.

4. Results with Different Random States

I tested the model with 11 different random states.

Training accuracy was always 100%.

Test accuracy changed a little, between 89% and 100%.

Mean test accuracy: 96%

The model is stable and gives similar results every time.

5. Results with Different Test Sizes

I tried different test set ratios: 10%, 25%, 50%, 75%, 90%.

Training accuracy stayed high.

Test accuracy was highest when the test set was 10% or 25%.

When the test set was very big (90%), test accuracy dropped a little.

6. What I Did

Loaded the IRIS data from sklearn.

Checked the data and features.

Trained a Decision Tree model.

Tested the model with different random states and test sizes.

Calculated and compared accuracy scores.

Made this report.

7. Plots and Charts

Accuracy vs Test Set Ratio Plot: Shows how accuracy changes with different test set sizes.

Decision Tree Visualization: Shows the tree structure and decision rules.