Decision Tree Visualization Lab

Objective

The goal of this lab is to build and visualize a Decision Tree using Python's scikit-learn library and generate a graphical representation of the tree using graphviz.

Requirements

* Python 3.x
* Required Python Libraries:
  + scikit-learn (for building the decision tree)
  + graphviz (for visualizing the decision tree)
  + pandas (optional, for handling data)
  + matplotlib (optional, for plotting)

Setup Instructions

1. Install Required Libraries

First, ensure that you have the necessary libraries installed. You can install them using pip:

bash

Copy code

pip install scikit-learn graphviz matplotlib pandas

2. Install Graphviz

Graphviz is used for rendering the visualization of the decision tree. Follow these steps to install Graphviz:

* Download the appropriate version for Windows from the Graphviz download page.
* Install the software by running the setup installer.
* After installation, add the Graphviz bin directory to your system's PATH environment variable:
  + Go to C:\Users\Bhavy\Graphviz\bin (or your Graphviz installation directory).
  + Add this path to the Path environment variable in System Properties > Environment Variables.

3. Test Installation

Verify that Graphviz is correctly installed by running the following command in your terminal:

bash

Copy code

dot -version

This should output the version of Graphviz installed on your system. If you see an error, make sure the Graphviz bin directory is added to your system's PATH.

Code

Below is the code to build and visualize the Decision Tree:

python

Copy code

# Import required libraries

import numpy as np

import pandas as pd

from sklearn.datasets import load\_iris

from sklearn.tree import DecisionTreeClassifier, export\_graphviz

import graphviz

import matplotlib.pyplot as plt

# Load dataset

data = load\_iris()

X = data.data

y = data.target

# Train a decision tree classifier

clf = DecisionTreeClassifier(criterion="entropy")

clf = clf.fit(X, y)

# Export decision tree as a dot file

dot\_data = export\_graphviz(clf, out\_file=None,

feature\_names=data.feature\_names,

class\_names=data.target\_names,

filled=True, rounded=True,

special\_characters=True)

# Visualize the tree using Graphviz

graph = graphviz.Source(dot\_data)

graph.render("entropy\_tree") # Saves as entropy\_tree.pdf

graph.view()

4. Explanation of Code

* Data Loading: The load\_iris() function from sklearn.datasets loads the Iris dataset, which is used to train the decision tree model.
* Model Training: A decision tree classifier is created using DecisionTreeClassifier from sklearn.tree. The criterion for splitting is set to entropy.
* Visualization: The export\_graphviz function generates a .dot file which is then rendered and visualized using graphviz.

Troubleshooting

Graphviz Installation Issues

* If you get an error like ExecutableNotFound: failed to execute WindowsPath('dot'), this means that the Graphviz executables are not properly added to the system’s PATH.
* Ensure that Graphviz is installed and the path to the bin directory is included in your PATH environment variable.

Python Library Errors

* If you encounter errors related to missing libraries (ModuleNotFoundError), ensure that all dependencies are installed:
  + scikit-learn
  + graphviz
  + matplotlib
  + pandas

You can install these libraries using:

bash

Copy code

pip install scikit-learn graphviz matplotlib pandas