Machine Learning - Assignment 1 Report

Hardik Soni (20CS30023) Saurabh Jaiswal (20CS30047)

September 18, 2024

Contents

Abstract

Question 1: Decision Tree

1. Split Data-set A into 80%-20% to form training and testing sets, respectively. Build a Decision Tree Classifier using ID3 algorithm. Train the classifier using Information Gain (IG) measure (no packages to be used for Decision Tree Classifier).

2. Repeat (1) for 10 random splits. Print the best test accuracy and the depth of that tree.

3. Perform reduced error pruning operation over the tree obtained in (2). Plot a graph showing the variation in test accuracy with varying depths. Print the pruned tree obtained in hierarchical fashion with the attributes clearly shown at each level.

4. Prepare a report including all your results.

1 DataSet

Customer segmentation is the practice of dividing a customer base into groups of individuals that are similar in specific ways relevant to marketing, such as age, gender, interests and spending habits.

- **Gender** : The Gender of the Person. Categorical Data (Male / Female)
- **Ever_Married** : Has the person ever married. Categorical Data (Yes/No)
- **Age** : The Age of the Person. Numerical Data [0-100]
- **Graduated** : Has the person completed his graduation. Categorical Data (Yes/No)
- **Profession** : The Profession of the Person. Categorical Data
- **Work_Experience** : The Work Experience of the Person in

- **Spending_Score**
- **Family_Size**
- **Var_1**
- **Segmentation**

This data set contains a total of 8068 customer details which are categorized into 4 segments(A, B, C, D).

2 The Decision Tree Algorithm Used

This is a great section to explain the choices your group made to measure the resonance frequencies. Use your answers to questions 1–4 to discuss these choices.

3 Important Terms and Expressions

Please don't include any data tables in your LaTeX write up. Making tables in LaTeX is very boring, although there are programs to convert your Excel file to LaTeX form! We'll worry about that later. Simply submit your Excel or Google Sheets workbook via Sapling, sharing on Google drive, or email.

4 Results and Procedures

Here you discuss your observations and results.

4.1 One Part

The subsection command let's you further divide your sections up. Comment on your observations and results here.

You should write a few sentences about your answers to questions 6–8.

4.2 Second Part

Here's a good spot to discuss your plot (again, don't include the plot, we'll learn how to do this later). Use your answer to question 7.

4.3 When to Make New Sections

As a general rule of thumb, don't make smaller sections, subsections, etc, unless there are at least two of them at that level (just like with lists).