**Homework 1: Decision Tree**

**Report ([student\_id] [student\_name])**

**Part I. Implementation**

PASTE YOUR CODE WITH COMMENT HERE

**Part II. Results & Analysis**

|  |  |  |
| --- | --- | --- |
|  | train data accuracy | test data accuracy |
| d=5 | % | % |
| d=10 | % | % |
| d=15 | % | % |
| d=20 | % | % |

**Part III. Answer the questions**

1. Please describe a problem you encountered and how you solved it.

A:

2. Why do we need to change the string inside the dataset into numbers?

A:

3. What are underfit and overfit? How to observe them in decision tree?

A:

4. How to prevent overfit in decision tree? List three methods.

A:

5. Is it possible that the train data accuracy become 100%?

A: