DATA MINING - Project 2

P76061425 林聖軒

Usage

- **Decision Tree**

```
$ python3 decisionTree.py [-h]
```

optional Options	Description
-h, --help	show this help message and exit
-train, TRAIN_PATH	Input training data file, default = ./data/train_data.txt
-test TEST_PATH	Input testing data file, default = ./data/test_data.txt

訓練 **decisionTree**，並將訓練出的 **decisionTree** 結果 **output** 至當前目錄的 **tree.pdf** 中。

需要安裝 **graphviz**

```
$ apt-get install graphviz
```

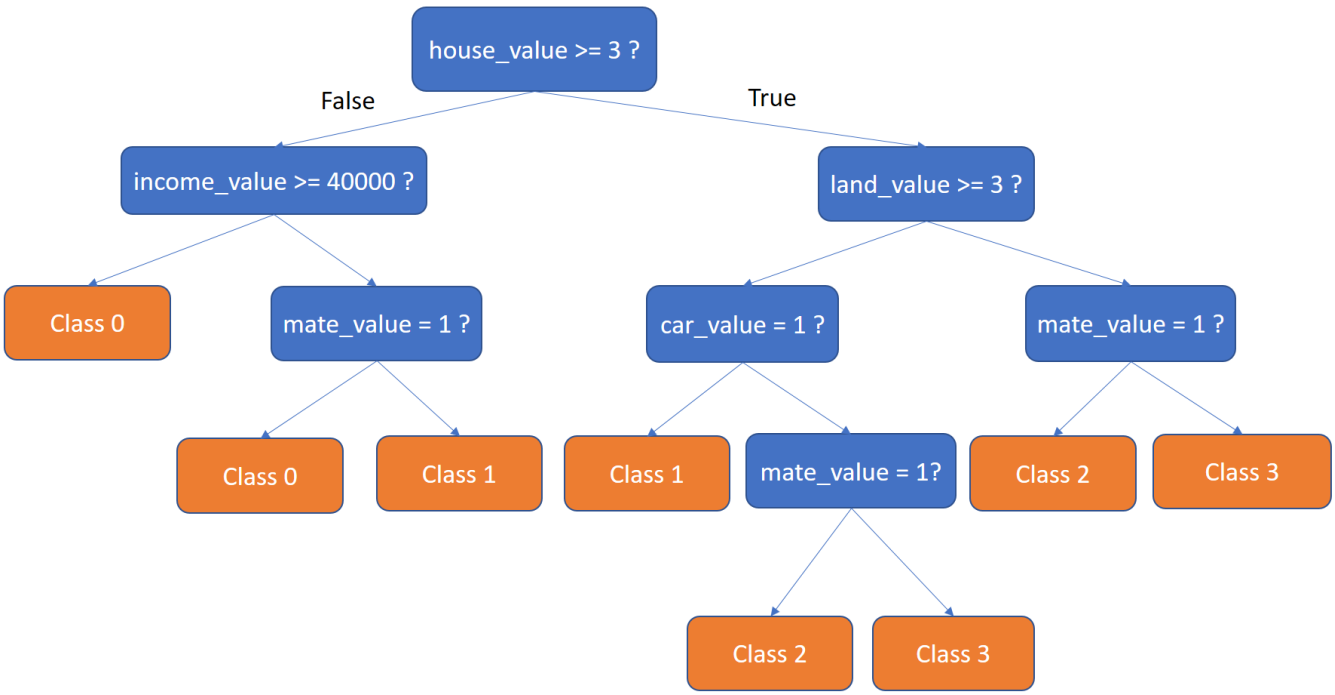
- **Data Generator**

```
$ python3 data_generator.py [-h]
```

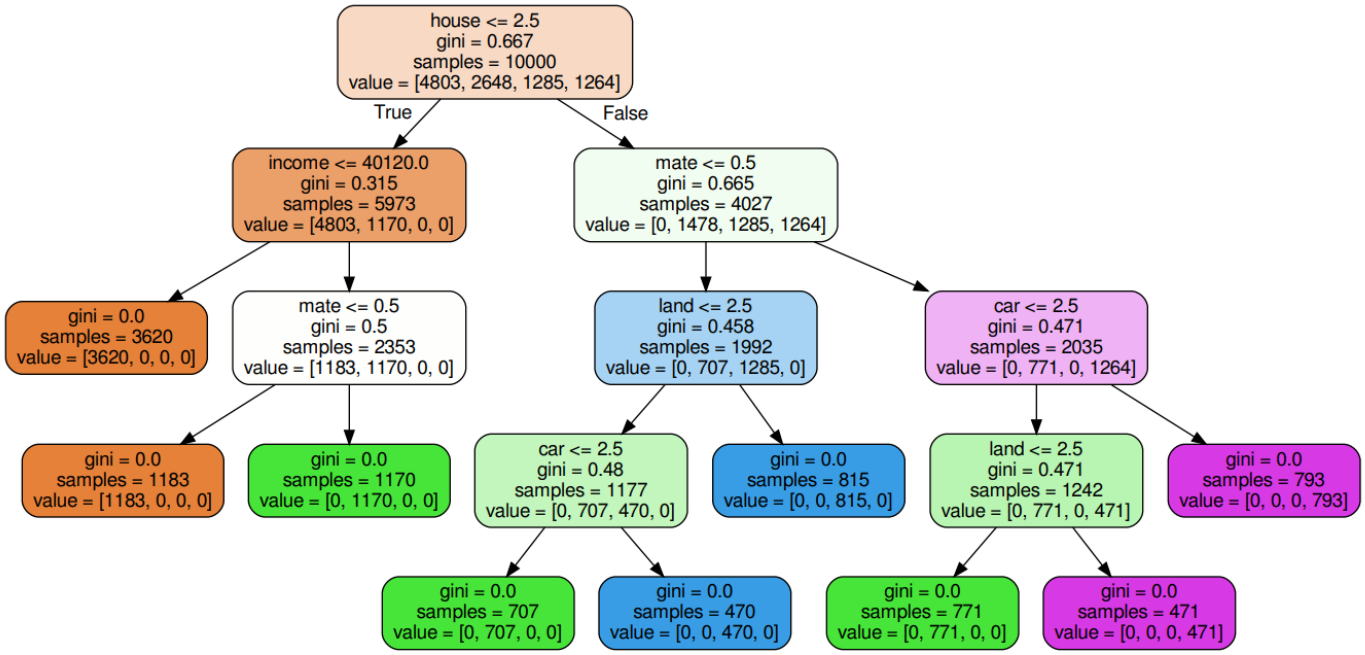
optional Options	Description
-h, --help	show this help message and exit
-n, DATA_AMOUNT	The number of data you want to generate, default = 10000

執行後會在 **data** 資料夾內生成 **10000** 筆 **training data(train_data.txt)**及 **testing data(test_data.txt)**。

- **Absolutely Right Rules**
  - **Attributes_list = house, car, land, income, mate, class**
  - **house_value = [0, 5]**
  - **car_value = [0, 5]**
  - **land_value = [0, 5]**
  - **income_value = [-50000, 100000]**
  - **mate_value = [0, 1]**
  - **class_vlaue = {0, 1, 2, 3}**



- **Decision tree**
  - **Training size = 10000**
  - **Testing size = 10000**
  - **Criterion = Gini**



  - **Accuracy = 0.9997**
  - **Precision = 0.9999826689774697**
  - **Recall = 0.9999677377726158**

- 從結果圖可以得知，與 **Absolutely Right Rules** 相比，**Decision tree** 所建立出的 **model** 和實際的 **rules** 並非完全相同，但有很高的相似度，由於 **Absolutely Right Rules** 的規則很簡單，因此結果的 **Accuracy, Precision, Recall** 均有相當好的表現。