**DATA MINING -** **Project 2**

**P76061425 林聖軒**

**Environment**

DISTRIB\_ID=Ubuntu  
DISTRIB\_RELEASE=18.04  
DISTRIB\_CODENAME=bionic  
DISTRIB\_DESCRIPTION="Ubuntu 18.04.1 LTS"

**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

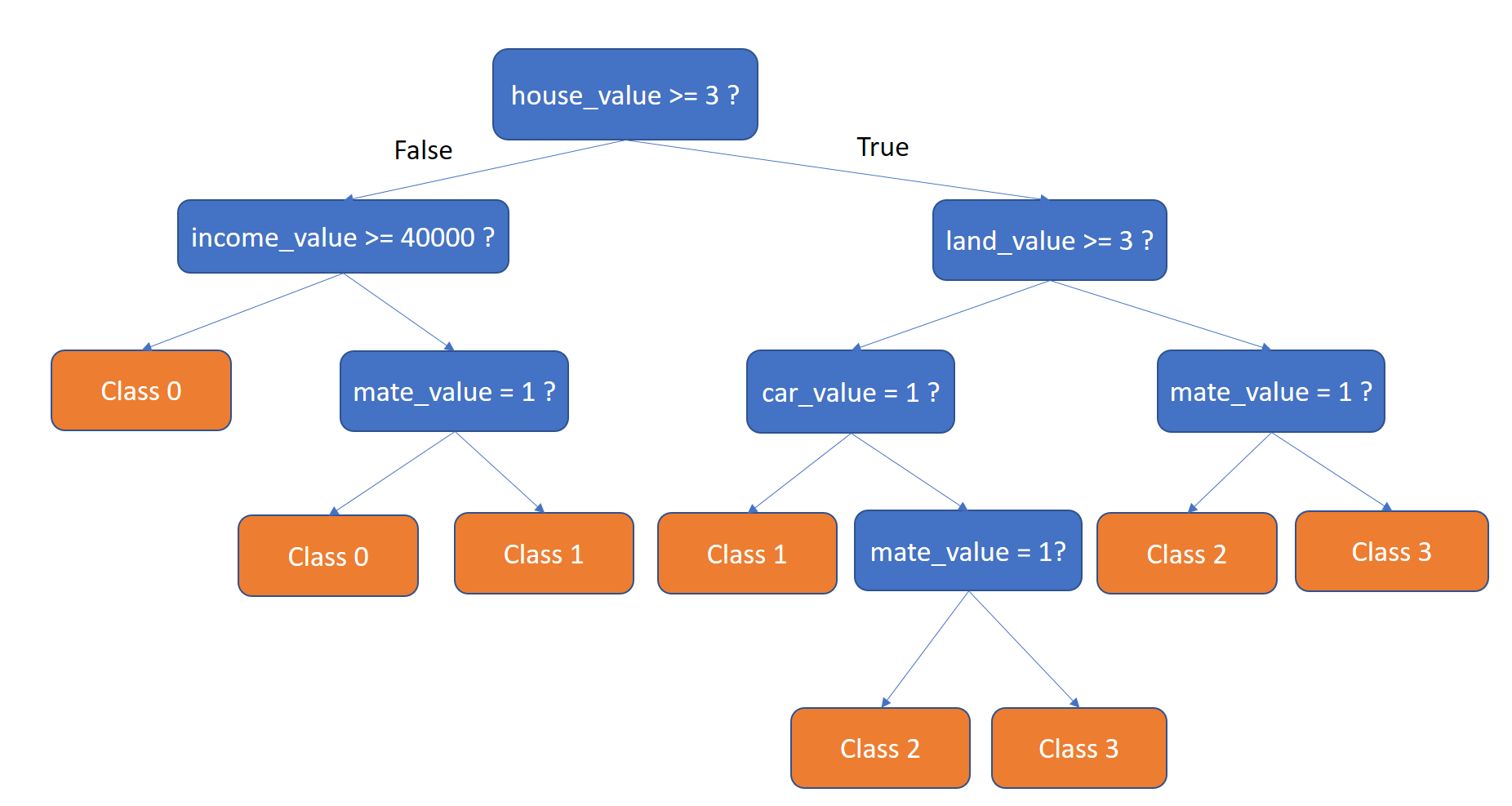
* **Data Generator**

$ python3 data\_generator.py [-h]

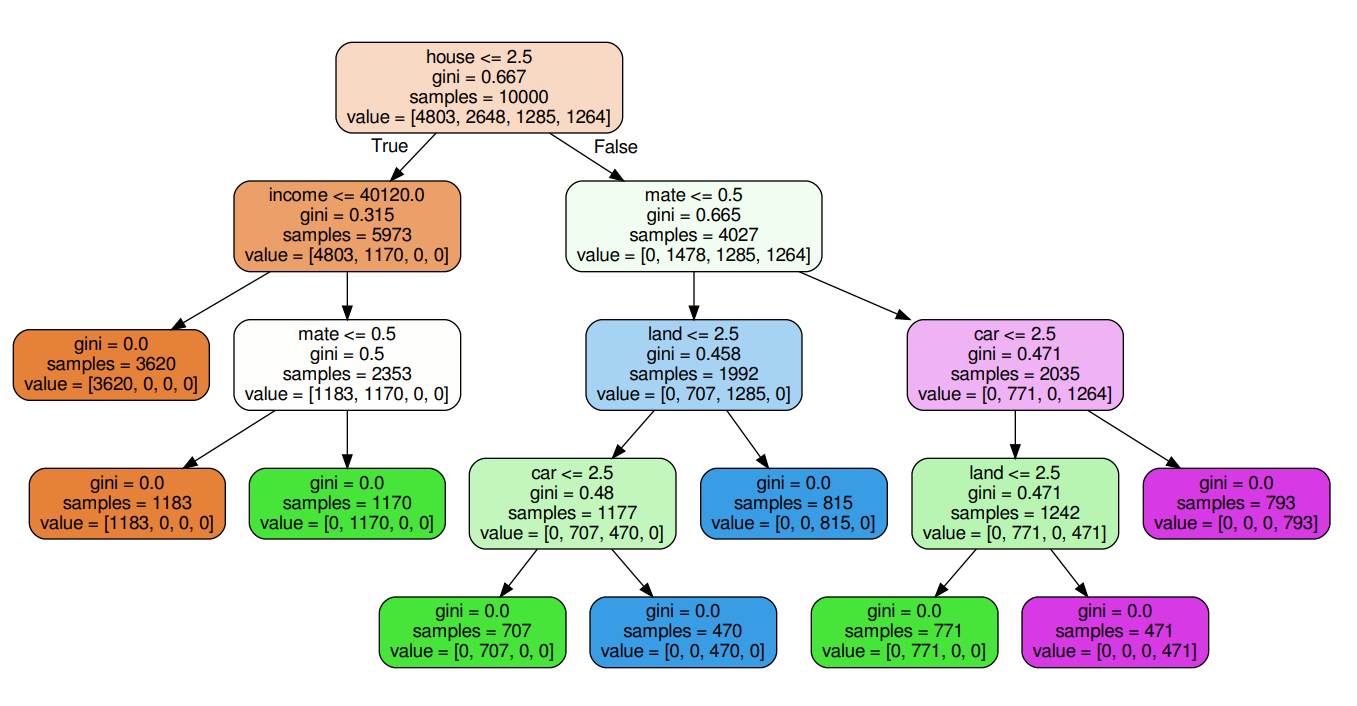
| **optional Options** | **Description** |
| --- | --- |
| -h, --help | show this help message and exit |
| -n, DATA\_AMOUNT | The number of data you want to generate, default = 10000 |

**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。**