



Fundamentals of Data Analytics **HW2**



Decision Tree Practice

- **Goal**

Understand what decision trees do and the difference between real behavior of classification model and observed data

- **Description**

Construct a decision tree to observe the difference between real **'right'** data and modeled data

- Due: **4/18 Thu. 9am !!!**



Flow

- **Step 1 -**

Design features/attributes for your problem and a set of rules to classify data

- **Step 2 -**

Use the data generated in step 1 to build a decision tree

- **Step 3 -**

Compare the rules in the decision tree from step 2 and the rules you used to generate your 'right' data

- **Step 4 -**

Discuss anything you can



Simple Example

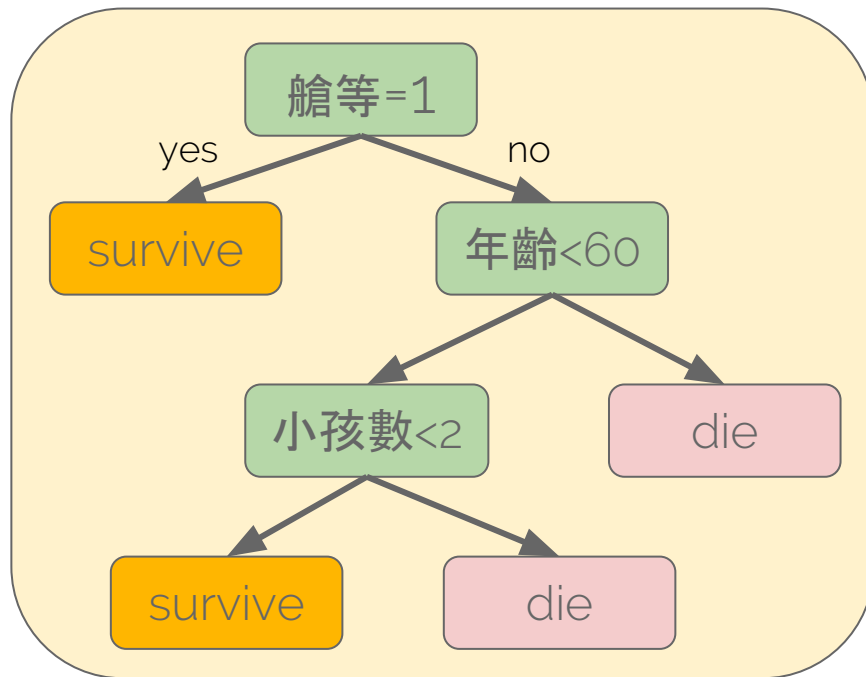
1. Question:

Titanic- Who will survive?

2. Attributes/Features:

艙等(1~3)、性別、年齡、兄弟姊妹數、父母數、小孩數、票價...

3. Designed Rules:





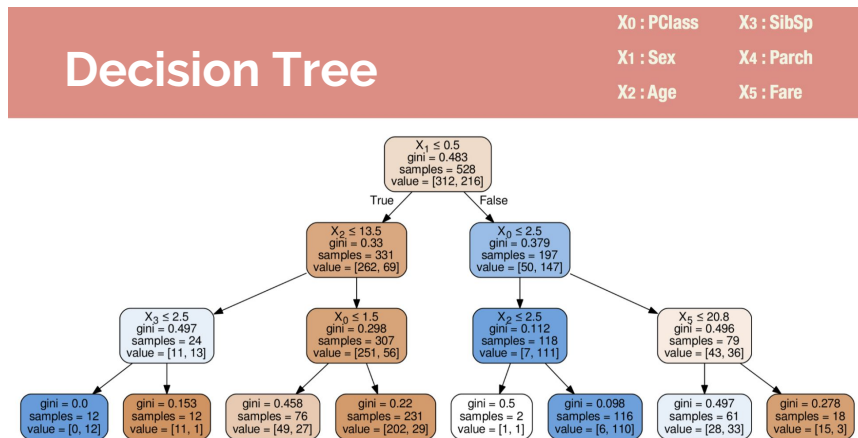
Simple Example -conti.

4. generate data

index	Pclass	Sex	...	Survival
1	1	1	...	1
2	1	0	...	1
3	2	1	...	0
4	3	0	...	0
...

5. construct decision tree

6. compare the rules





Requirement

- You should use **jupyter / ipython notebook** to implement this homework.
- Please upload your **code, generated data** and **report** to your github.