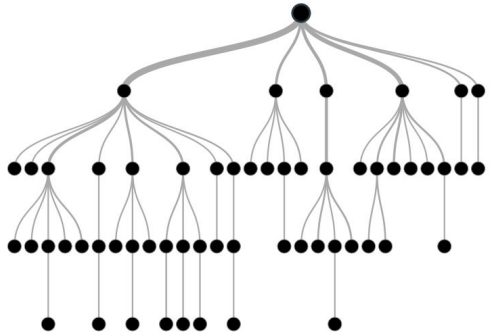
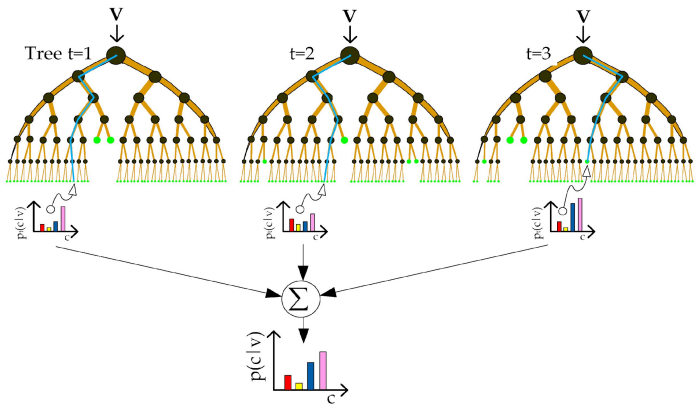
BT2101 Tutorial 1

Purpose of Tutorial

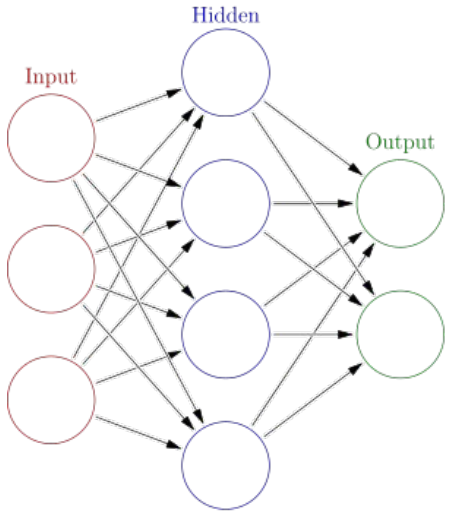
Time Series



Decision Tree



Ensemble Learning



Neural Networks

Agenda

- Form a team of 3-4
- Discussion about Assignment 1
  - Decision Tree Split
- Python Implementation

CART

- Think about
  - Decision Tree Building
  - Entropy = 0 / Entropy =1 ?

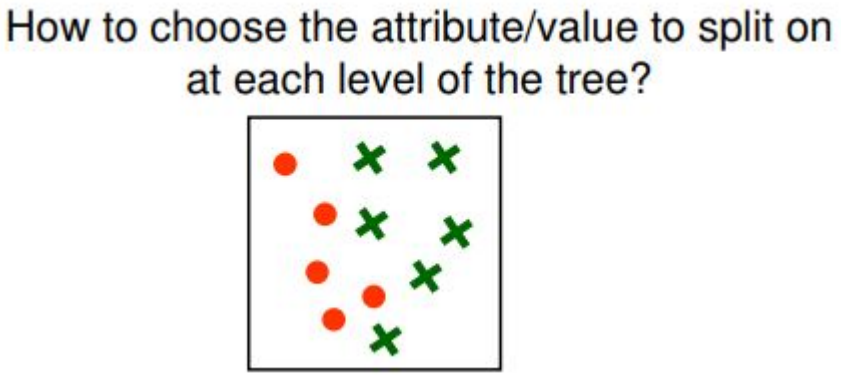
$$E(S) = \sum_{i=1}^{c} -p_i \log_2 p_i$$

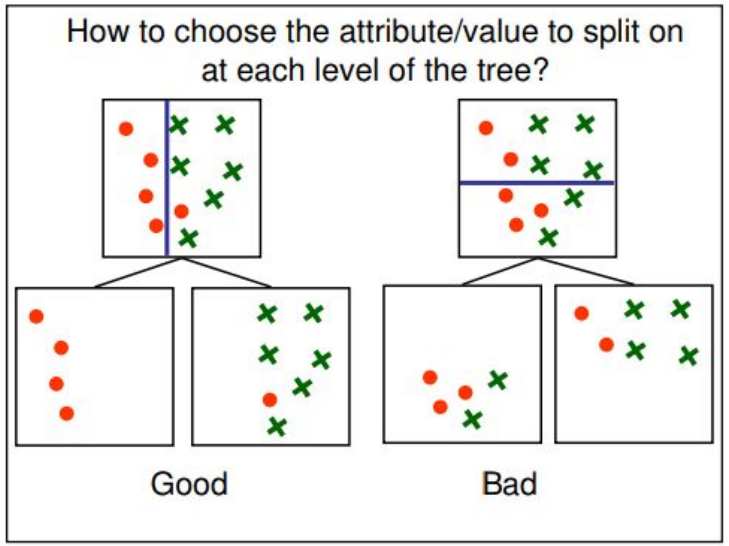
  - Why is information gain a good tool to choose best attribute? What is the intuition behind?

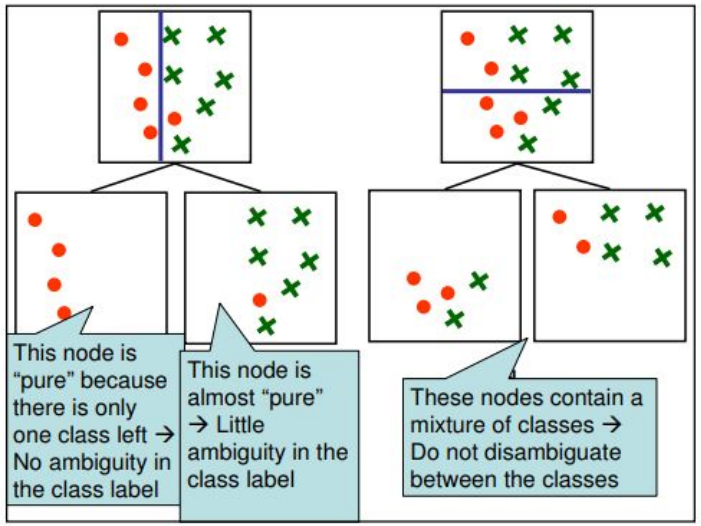
$$Gain(S, A) = E(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} E(S_v)$$

  - What kind of features would you like to collect?

“Bad” root attribute







Implementation in Python

BT2101 Introduction to Decision Tree

1 Goal

In this notebook, we will explore **Decision Tree** including:

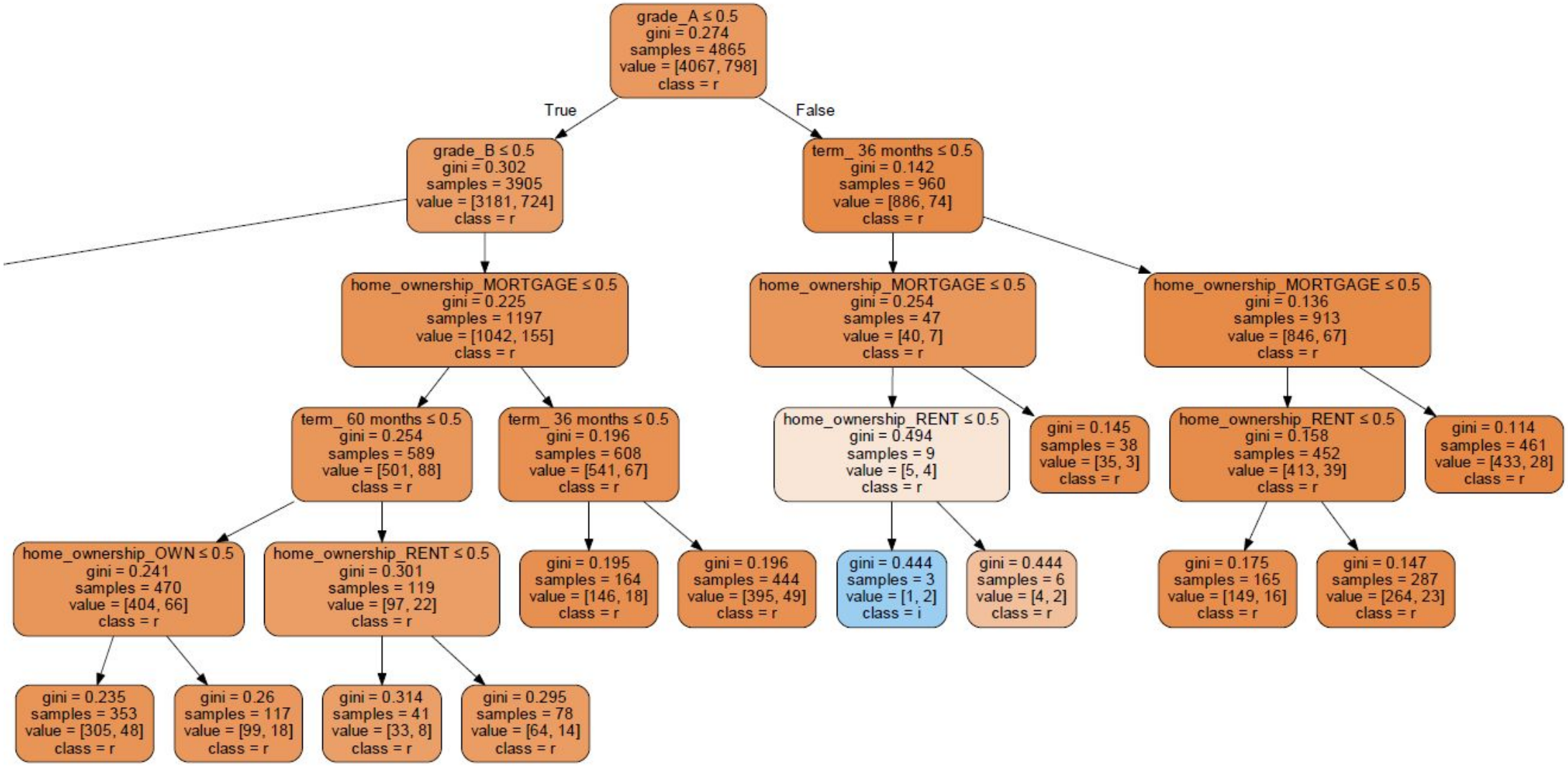
- User-defined functions
- Open-source package: `scikit-learn`

For the **Decision Tree** method, you will:

- Use numpy to write functions
- Write binary recursive splitting functions
- Write decision functions
- Write pruning functions
- Use open-source package to do classification

In [ ]:
```
# -*- coding:utf-8 -*-
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from math import sqrt, log
from __future__ import division
from collections import defaultdict
%matplotlib inline
```

Implementation in Python



Programming Assignment 1

Using the BT2101 Tutorial 1 Programming code (Decision Tree.ipynb), please answer the questions in the jupyter notebook

1. Set your own stopping conditions for pre-pruning;
2. Try to do post-pruning;
3. Try to use different criteria for binary split, such as misclassification and Gini index;
4. Write a function to calculate misclassification error rate

5.1 What if features are continuous? Explain in words what would happen

5.2 What if output is continuous? Explain in words what happen

Answer all in the jupyter notebook.

Instructions

Submit Python Notebook to IVLE HW1 Named:
AXXXX_T1_program.ipynb

Include your answers in the jupyter notebook

Submit a DRAFT by Aug 28 before Tutorial 1 (by midnight)

Submit a FINAL program by Sept 1 (by midnight)

Thank you!