Homework 1

Zhong Yun 2016K8009915009

September 12, 2018

Problem 1 *Construct a data, each conditional data has its own value. The constructed data should prove that the decision tree calculated with greedy algorithm is not an optimal decision tree.*

Answer 1 Example:Suppose there are three key factors that determine if a basketball game wins or loses. Data Set is as follows:

No.	***Win or Lose***	***Time***	***Be home or away***	***Weather***
1	*win*	*morning*	*home*	*sunny*
2	*win*	*morning*	*home*	*rain*
3	*win*	*morning*	*away*	*rain*
4	*win*	*morning*	*home*	*rain*
5	*lose*	*morning*	*away*	*sunny*
6	*lose*	*night*	*home*	*sunny*

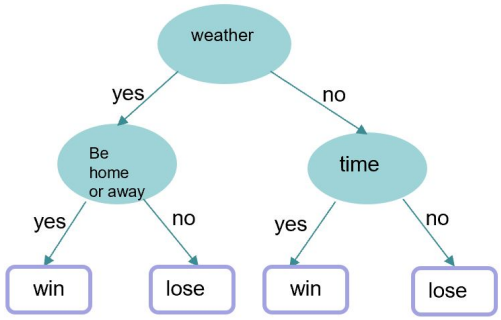
First, let's create a decision tree using Greedy Algorithm:
Conditional Entropy:(Here we use Shannon Information Entropy to compute Information Entropy)

$$\begin{aligned} H(Win\ or\ Lose|Time) &= P(morning)H(Win\ or\ Lose|morning) + P(night)H(Win\ or\ Lose|night) \\ &= P(morning)\left(P(win|morning)log_2\frac{1}{P(win|morning)} + P(lose|morning)log_2\frac{1}{P(lose|morning)}\right. \\ &= 0.918 \end{aligned}$$

Similarly, we have that

$$\begin{aligned} H(Win\ or\ Lose|Be\ home\ or\ away) &= 0.874 \\ H(Win\ or\ Lose|Weather) &= 0.459 \\ &\dots \end{aligned}$$

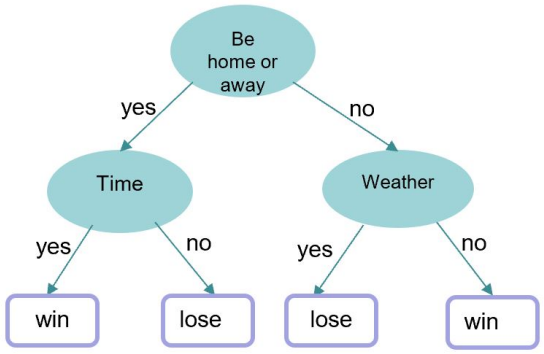
Finally, we have a decision tree(see tree [a]):



Tree [a]

Test the data set given,the accuracy of decision tree [a] is $\frac{5}{6}$.(The sixth data do not comform to decisicion tree [a])

However, we can create a better decision tree b for the data set(see tree [b]):



Tree [b]

The accuracy of tree [b] is $\frac{6}{6} = 1 \geq \frac{5}{6}$.
It is proved that decision tree calculated with greedy algorithm may not be an optimal decision tree.