Modelling Decision trees ID3, C45, J48

Tuesday, March 21, 2017 3:02 PM

1. **What is entropy is and how do you calculate it?**

   It is the way to measure the uncertainty of a class in a subset of examples.

   It is interpreted as the number of bits needed to determine if X positive or negative belongs to S.

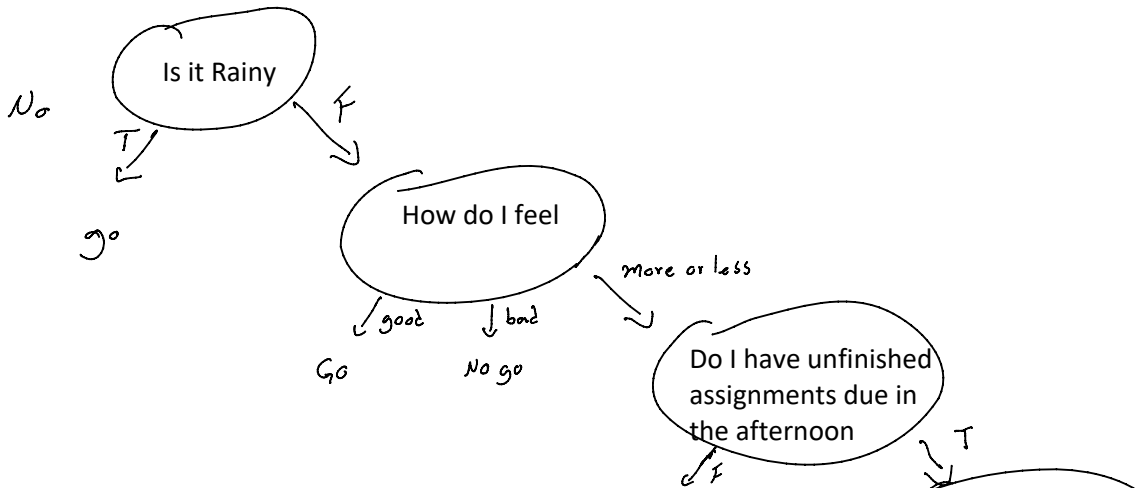
   The entropy value will be a fractional value between 1 and 0.

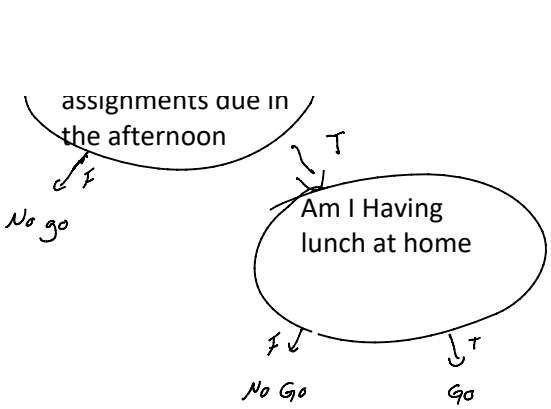
   A pure set will return 0, and a totally uncertain set (50%-50%) will return 1.
2. **What is information gain?**
   It is the number of bits you would gain in each split.
   The weight in each entropy value, the amount of items that fall into a subset.
3. **Why is information entropy relevant to computer science in general?**
   Information entropy is a relevant topic to computer science because it allows us to classify the level of uncertainty of a data set. Knowing this, we can determine if a piece of information is trustworthy or not depending on the evidence that is provided.
4. **What are decision trees such as ID3, C4.5 and J48? What are they used for?**
   Decision trees are data classifiers that arrange data in structures that allow us to operate decision functions among them. A decision tree with enough evidence can determine the expected outcome of a specific scenario based on the data of all previous scenarios.
5. **Explain how ID3 works in your own words.**

   An ID3 is a decision tree that classifies a set of data into different subsets of data in order to assume a decision outcome based on a specific input and the previous evidence. An ID3 will try to divide the set of data into subsets with the same outcome that have any information in common. A subset with the data that leads to the same outcome is considered a pure set, reaching a decision tree full of pure sets in the end is the goal of an ID3 as long as most leaf nodes have a weight (amount of items in the subset) higher than 1.
6. **Give a small example of a decision tree using an example in your daily life**
   A decision tree for my daily life can be used to determine whether I'll go to soccer practice or not. There are couple of factors that determine if I take the decision to go or not. A list of factors may be:

   -How do I feel physically: good | more or less | bad

   -Do I have unfinished assignments due in the afternoon?: yes | no
   -Is it rainy: yes | no
   -Is it sunny: yes | no
   -Am I having lunch at home: yes | no




assignments due in
the afternoon
F
No go
T
Am I Having
lunch at home
F
No Go
T
Go