Decision Tree Notes

JEFF LABONTE

1 Information Entropy

On this page I will note what I have learned throughout all the research I made towards better understanding of decision trees

Information Entropy : It is a concept from information theory. It tells how much Information there is in an event. In general, the more uncertain or random the event is, the more Information it will contain. More clearly stated, information is a decrease of incertainty or entropy

from `https: // simple. wikipedia. org/ wiki/ Information_ entropy`

Basicly, it is the probality of the information that will comes out of a system, which will produce more and who is going to produce less informations.
i.e letters generated by a machine that produces a sequence of letters.

Example and formulas took from `https://www.khanacademy.org/computing/ computer-science/informationtheory/moderninfotheory/v/information-entropy`

Example from Khah Academy : machine 1 has
- **50%** chance to produce the letter **A**
- **12.5%** chance to produce the letter **B**
- **12.5%** chance to produce the letter **C**
- **25%** chance to produce the letter **D**

Machine 2 has
- **25%** chance to produce the letter **A**
- **25%** chance to produce the letter **B**
- **25%** chance to produce the letter **C**
- **25%** chance to produce the letter **D**

The multiplier to the letter is found by finding the level at which the letter is in the tree of propablity.

For machine one :
$$nbr\_of\_bounces = p_a * 1 + p_b * 3 + p_c * 3 + p_d * 2$$

$nbr\_of\_bounces = 0.50 * 1 + 0.125 * 3 + 0.125 * 3 + 0.25 * 2$
$nbr\_of\_bounces = 1.75$ *bounces to reach an answer*

For machine two :

$nbr\_of\_bounces = p_a * 1 + p_b * 3 + p_c * 3 + p_d * 2$
$nbr\_of\_bounces = 0.25 * 2 + 0.25 * 2 + 0.25 * 2 + 0.25 * 2$
$nbr\_of\_bounces = 2$ *bounces to reach an answer*

So if we have to guess 200 values from the machine.

- We would have to ask 175 times machine 1
- We would have to ask 200 times machine 2

Which means machine 1 will produce less information. It also means that there less **incertainty** with *machine 1*. This is called **entropy**