University
of
St Andrews

CS5010 Artificial Intelligence Principals

Notes For Lecture 4

Zhongliang Guo

November 29, 2021

1 Introduction

I checked the previous examination, and I found those text are not useful for exam. The only thing useful is the information theory and decision tree. So I will divide this notes into miscellaneous, Information Theory, and Decision Tree.

2 Information Theory

Those part Oggie did not illustrate, but it is essential to understand it, so I need to explain it.

2.1 Information entropy

When there are multiple cases of a event, the uncertainty of which case the event is for the observer is called information entropy.

2.2 Information

The thing that can remove or reduce the observer's uncertainty about the matter is called information.

2.3 Relationship of above notions

The two are equal in quantity and opposite in meaning. Access to information means eliminating uncertainty.

2.4 Function of Information

1. Adjust probability
2. Exclude interference
3. Determine the situation

2.5 How to quantify information

Compared to the mass (physical quantity), we use a certain reference as the standard, called unit 1, i.e., 1 kg. So, when we define the amount of information, we should also pick a reference.

Take the coin flip event as a reference.

Mass is linearly related, but information is index Related. i.e., The possible outcomes that can be produced by flipping 3 coins are $2^3 = 8$, not $2 * 3 = 6$. (**Unit:** bit)

So when finding the amount of information, you should not use division, you should use log.

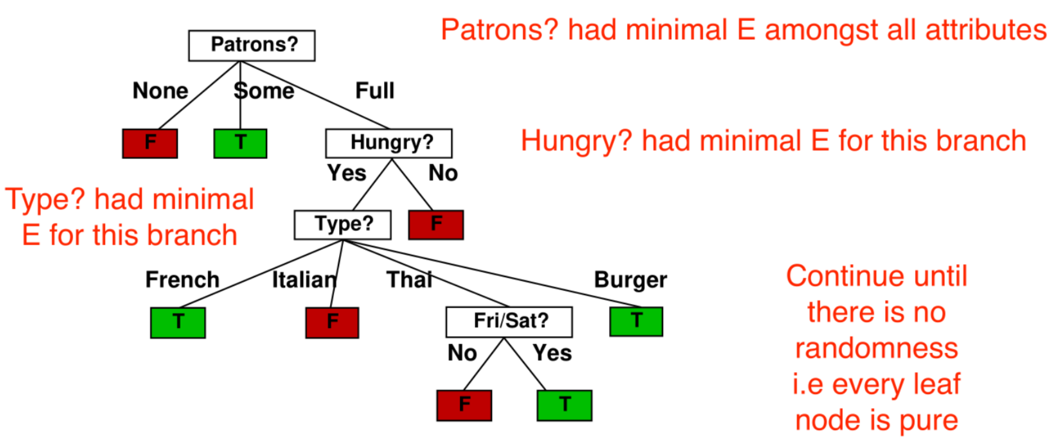
Information entropy when flip a coin with equal probability 3 times: $\log_2 8 = 3$ bits. (but everything happened when this coin is equal probability)

How about the situation when it is unequal probability? – Weighting using probabilities.

When $P(A) = \frac{1}{6}$, $P(B) = \frac{1}{6}$, $P(C) = \frac{1}{2}$, $P(D) = \frac{1}{6}$, The quantity of information is $\frac{1}{6}\log_2(\frac{1}{6})^{-1} + \frac{1}{6}\log_2(\frac{1}{6})^{-1} + \frac{1}{2}\log_2(\frac{1}{2})^{-1} + \frac{1}{6}\log_2(\frac{1}{6})^{-1}$

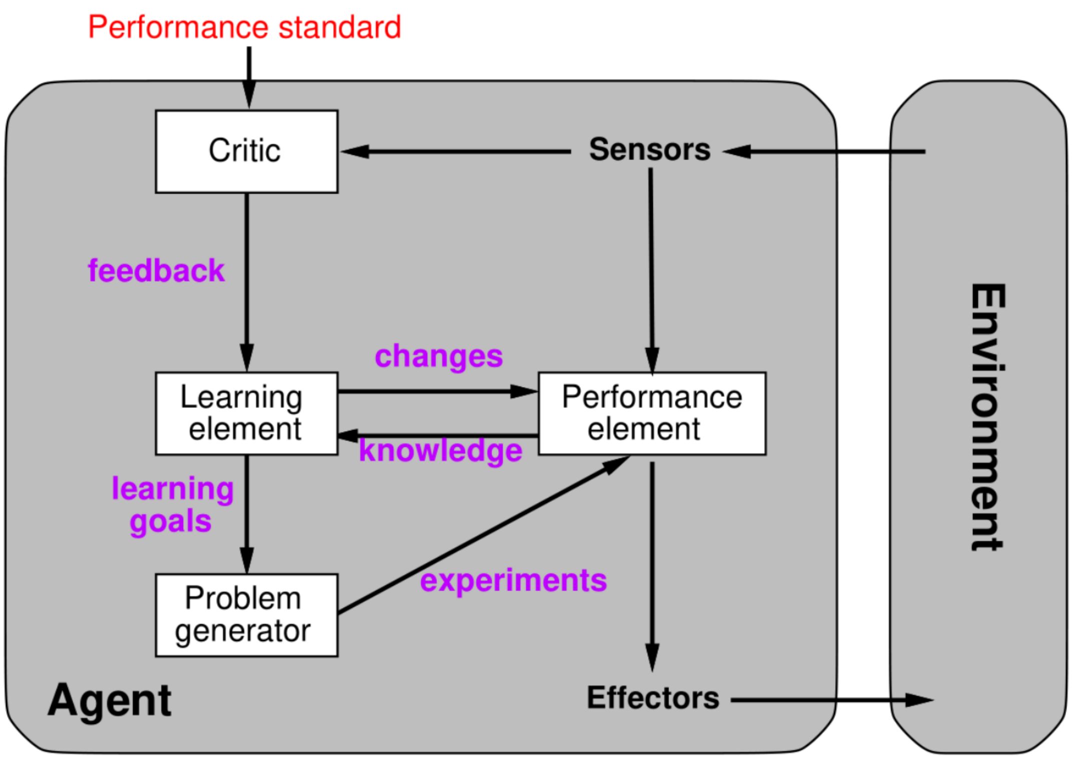
3 Decision Tree

The process is that, calculate all the information entropy of each attributes, order nodes (attributes) from smallest to largest.



4 Miscellaneous

4.1 Learning agents (was never tested in exam, info from previous students)



4.1.1 Learning

Learning is essential for unknown environments,
i.e., when designer lacks omniscience

Learning is useful as a system construction method,
i.e., expose the agent to reality rather than trying to write it down

Learning modifies the agent's decision mechanisms to improve performance

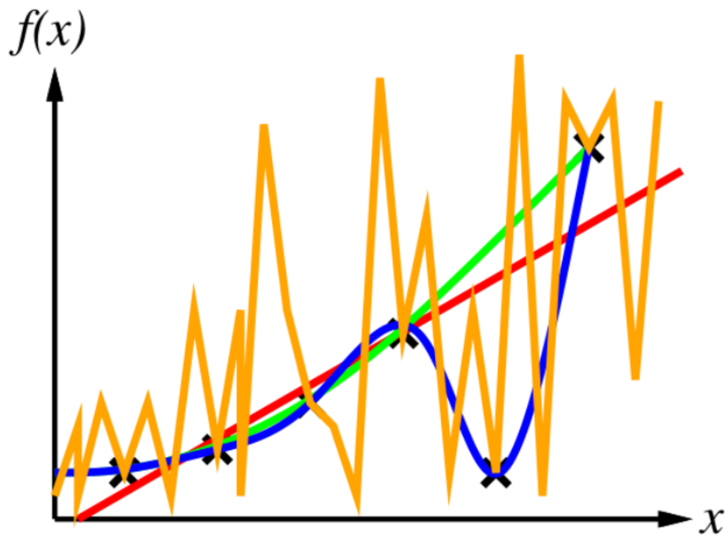
4.1.2 Learning element

Design of learning element is dictated by

- what type of performance element is used
- which functional component is to be learned
- how that functional component is represented
- what kind of feedback is available

4.2 Inductive learning

Different functions are used to fit the data. None of the lines seem to make much sense except for the green and red lines.



4.2.1 Ockham's razor principal

To solve above problem, this principal is needed.

Ockham's razor: maximize a combination of consistency and simplicity

My understanding: The simplest is often the best