AI Course Module 4 Notes

**4.1**

Some of the most popular decision tree algorithms in machine learning are:

* Classification and Regression Tree (CART)
* Iterative Dichotomiser3 (ID3)
* C4.5 (successor of ID3))
* Chi-squared Automatic Interaction Detection (CHAID)
* Decision Stump
* Conditional Decision Trees

A decision tree algorithm is used both to find an answer to a question and to analyze data. Ex. Titanic survival rate based on demographics.

Machine learning is most valuable when the data sets provided for the machine to learn from are expansive. However, large data sets also mean that good decision tree algorithms must be constructed for information to be extracted from them.

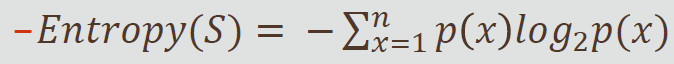
**4.2**

Information entropy refers to uncertainty in data. Specifically, it refers to randomness. Data with a high degree of randomness will consequently result in more information that can be used.

Given new data = new information = high entropy (learning took place), while known data = old information = no entropy ( no learning took place)

Variance measures how far a data set is spread out. This is technically defined as “he average of the squared differences from the mean.”

If there’s a high variance value, little is known of the data. However, we can be confident in low-variance data.

Entropy is measured as:

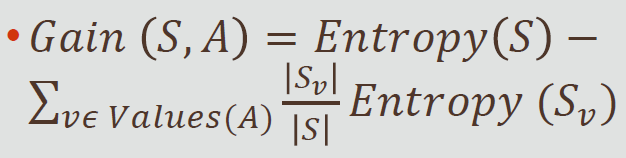
This equation will return how much information to expect from some action. High number = more information obtained.

**4.3**

The ID3 algorithm will both calculate entropy and calculate gain.

Information gain measures how much entropy is reduced when partitioning on an attribute (A). Higher number = better data classification, showing the best choice of the dependent data for this entropy.

Gain Formula:



**4.4**

In this section, we use the HashMap class. This class is simply another collection class. This class is functionally the same as the Dictionary class in Python.