Week 8: Daily Morning Challenge

Day 1: Tuesday 18th February 2020

Question 1: Briefly explain the difference between Supervised and Unsupervised Learning

Supervised learning is where you have input variables and an output variable and you use an algorithm to learn the mapping function from the input to the output. The aim is to approximate the mapping function so that when we have new input data we can predict the output variables for that data.

Unsupervised learning is modelling the underlying or hidden structure or distribution in the data in order to learn more about the data. Unsupervised learning is where you only have input data and no corresponding output variables.

Key differences are:

- In supervised learning, the input and output variable will be given. In unsupervised, only the inputs will be given.
- In supervised learning scientist acts as a guide to teach the algorithm what conclusions or predictions it should come up with. In unsupervised learning there is no correct answer there is no teacher, algorithms are left to their own to discover and present the interesting hidden structure in the data.
- Supervised learning goal is to determine the function so well that when new input data set given, can predict the output.
- The unsupervised learning goal is to model the hidden patterns or underlying structure in the given input data in order to learn about the data.
- Supervised learning is often used for export systems in image recognition, speech recognition, forecasting, financial analysis and training neural networks and decision trees etc
- Unsupervised learning algorithms are used to pre-process the data, during exploratory analysis or to pre-train supervised learning algorithms.

Question 2: Outline the steps involved in a data mining classification process

The two important steps of classification are:

1. Model construction

A predefine class label is assigned to every sample tuple or object. These tuples or subset data

are known as training data set.

The constructed model, which is based on training set is represented as classification rules,

decision trees or mathematical formulae.

2. Model usage

The constructed model is used to perform classification of unknown objects.

A class label of test sample is compared with the resultant class label.

Accuracy of model is compared by calculating the percentage of test set samples, that are

correctly classified by the constructed model.

Test sample data and training data sample are always different.