Data Mining 2
Quiz 1

1) Briefly, describe what is unsupervised learning?

Unsupervised learning is training a machine using information that is neither classified nor labelled and allowing the model to work on its own to discover information/patterns in the dataset. Hidden patterns or data groupings are discovered without any human intervention. Usually classified into 2 categories of algorithms:

- Clustering
- Association

2) Briefly, describe a major challenge in unsupervised learning.

While unsupervised learning has many benefits, it comes with its own challenges. Unlike supervised learning, there is no correct answer here, so algorithms are left to their own to discover interesting patterns in data. Sometimes when it lets the algorithm run without any human intervention challenges can occur like longer training times can be observed. There is also a higher risk of inaccuracy. Not knowing which data was clustered can cause problems too.

3) Give two real world examples of unsupervised learning.

Some of the main applications of unsupervised learning include clustering, dimensionality reduction, visualization, finding association rules and anomaly detection.

1. Recommendation Systems are one of the real world example for unsupervised learning algorithms. Clustering algorithms are used by ecommerce websites like Amazon to produce a user specific recommendation system.

2. Another real world example is credit card fraud detection, unusual credit card transactions are detected to prevent fraud. Anomaly detection techniques are used in this type of situations.

4) Give two examples of a quantitative variable:

Quantitative variable : numerical values that represent some kind of measurement.

1. Height
2. Weight