School of Computing and Information Systems
The University of Melbourne
COMP90049 Knowledge Technologies (Semester 2, 2018)
Workshop sample solutions: Week 7

1. Finish any remaining questions from last week, if necessary.

&

---

2. What is data mining/machine learning? What makes this a knowledge task?

   - Data mining: extracting implicit, previously unknown, potentially useful information from data
   - Machine learning: algorithms for acquiring structual descriptions from examples (special case of above?)
   - Knowledge task: the information/descriptions we produce are unknown and useful to humans

3. What is the difference between supervised and unsupervised machine learning?

   - Generally speaking, supervised techniques in machine learning start from exemplars — labelled with classes — in a set of training data, and use these to classify unknown instances in a set of test data.
   - Unsupervised methods are not based on a set of labelled training data: they are often broken down into **weakly** unsupervised methods, where the class set is known, but the system does not have access to labelled training data; and **strongly** unsupervised methods, where even the class set is unknown.
   - For example, Naive Bayes, Support Vector Machines, Decision Trees, and k-Nearest Neighbour are all examples of supervised systems.
   - Clustering (e.g. k-means, Expectation Maximisation) is an example of an unsupervised methodology.

4. In the context of (supervised) machine learning:

   (a) What is an instance?

   - An instance is a single exemplar from the data, consisting of a bundle of (possibly unknown) attribute values (feature values) and a class value, mapping on to the concept that we wish to predict.

   (b) What is an attribute? What different kinds of attribute are there?

   - An attribute is a single measurement of some aspect of an instance, for example, the frequency of some event related to this instance, or the label of some meaningful category.
   - Attributes are usually classified as either nominal (labels with no ordering), ordinal (labels with an ordering), or continuous (numbers, even if they perhaps aren't continuous in the mathematical sense).

   (c) What is a class?

   - A class is the thing (usually attribute) we want to learn. It may be nominal ("classification") or continuous ("regression").