School of Computing and Information Systems
The University of Melbourne
COMP90049 Knowledge Technologies (Semester 2, 2017)
Workshop exercises: Week 7

1. What is data mining/machine learning? What makes this a knowledge task?
2. What is the difference between supervised and unsupervised machine learning? Give examples of some supervised and unsupervised techniques.
3. In the context of (supervised) machine learning:
   (a) What is an instance?
   (b) What is an attribute? What different kinds of attribute are there?
   (c) What is a class?

   Consider the following dataset:

id	*apple*	*ibm*	*lemon*	*sun*	LABEL
A	4	0	1	1	FRUIT
B	5	0	5	2	FRUIT
C	2	5	0	0	COMP
D	1	2	1	7	COMP
E	2	0	3	1	?
F	1	0	1	0	?

4. Treat the problem as an unsupervised machine learning problem (excluding the *id* and LABEL attributes), and calculate the clusters according to k-**means** with $k = 2$, using the Manhattan distance:
   (a) Starting with seeds A and D.
   (b) Starting with seeds A and F.
5. Perform **agglomerative clustering** of the above dataset (excluding the *id* and LABEL attributes), using the Euclidean distance and calculating the **group average** as the cluster centroid. Do you expect to observe a different dendrogram if we were instead using the cosine similarity?