School of Computing and Information Systems
The University of Melbourne
COMP30027 MACHINE LEARNING (Semester 1, 2019)

Tutorial exercises: Week 5

1. For the following dataset:

apple	*ibm*	*lemon*	*sun*	CLASS
TRAINING INSTANCES				
4	0	1	1	FRUIT
5	0	5	2	FRUIT
2	5	0	0	COMPUTER
1	2	1	7	COMPUTER
TEST INSTANCES				
2	0	3	1	?
1	2	1	0	?

   (a) Classify the test instances according to the method of **Nearest Prototype**.

   (b) Using the **Euclidean distance** measure, classify the test instances using the 1-NN method.

   (c) Using the **Manhattan distance** measure, classify the test instances using the 3-NN method, for the three weightings we discussed in the lectures: majority class, inverse distance, inverse linear distance.

   (d) Can we do weighted k-NN using **cosine similarity**?

2. Revise SVMs, particularly the notion of "linear separability".

   (a) If a dataset isn't linearly separable, an SVM learner has two major options. What are they, and why might we prefer one to the other?

   (b) Contrary to many geometric methods, SVMs work better (albeit slower) with large attribute sets. Why might this be true?

3. We have now seen a decent selection of (supervised) learners:

   - Naive Bayes
   - 0-R
   - 1-R
   - Decision Trees
   - k-Nearest Neighbour
   - Nearest Prototype
   - Support Vector Machines

   (a) For each, identify the model built during training.

   (b) Rank the learners (approximately) by how fast they can classify a large set of test instances. (Note that this is largely independent of how fast they can build a model, and how well they work in general!)