School of Computing and Information Systems
The University of Melbourne
COMP90049 Knowledge Technologies (Semester 2, 2018)
Workshop exercises: Week 9

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	0	1	0	?

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

   (b) Using the **cosine similarity** measure, classify the test instances using the 3-NN method. Extend this to the **weighted** 3-NN method.

2. For the following dataset:

apple	*ibm*	*lemon*	*sun*	CLASS
TRAINING INSTANCES				
Y	N	Y	Y	FRUIT
Y	N	Y	Y	FRUIT
Y	Y	N	N	COMPUTER
Y	Y	Y	Y	COMPUTER
TEST INSTANCES				
Y	N	Y	Y	?
Y	N	Y	N	?

   Use the method of **Naive Bayes** classification, as shown in lectures, to classify the test instances. Revise some of the assumptions that are built into the model.

   (a) [EXTENSION] Revise the **multinomial distribution**. Naive Bayes can be extended to account for integer frequencies in the data (like in Question 1) using this model. Read up on so-called **multinomial Naive Bayes**.

3. Revise Support Vector Machines, paying particular attention to the terms "linear separability" and "maximum margin".

   (a) What are "soft margins", and when are they desirable?

   (b) Why are we interested in "kernel functions" here?

   (c) Why are SVMs "binary classifiers", and how can we extend them to "multi-class classifiers"?