

School of Computing and Information Systems  
The University of Melbourne  
COMP90049 Introduction to Machine Learning (Semester 2, 2020)  
Tutorial exercises: Week 5

1. For the following dataset:

<i>apple</i>	<i>ibm</i>	<i>lemon</i>	<i>sun</i>	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) Using the **Euclidean Distance** measure, classify the test instances using the 1-NN method.
  - (b) 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*.
  - (c) Can we do weighted k-NN using **Cosine Similarity**?
2. What is **gradient descent**? Why is it important?
3. [OPTIONAL] (a) What is **Regression**? How is it similar to **Classification**, and how is it different?
- (b) Come up with one typical classification task, and one typical regression task. Specify the range of valid values of  $y$  (results) and possible valid values for  $x$  (attributes)
4. What is **Discretisation**, and where might it be used?
5. Discretise the following dataset according to the (unsupervised) methods of **equal width** and **equal frequency**.

ID	A (°C)	B (mm)	C (hPa)	CLASS
1	22.5	4.6	1021.2	AUT
2	16.7	21.6	1027.0	AUT
3	29.6	0.0	1012.5	SUM
4	33.0	0.0	1010.4	SUM
5	13.2	16.4	1019.5	SPR
6	14.9	8.6	1016.4	SPR
7	18.3	7.8	995.4	WIN
8	16.0	5.6	1012.8	WIN