School of Computing and Information Systems
The University of Melbourne
COMP90049 Introduction to Machine Learning (Semester 2, 2022)

Week 3

1. For the following dataset:

id	*apple*	*ibm*	*lemon*	*sun*	CLASS
TRAINING INSTANCES					
A	4	0	1	1	FRUIT
B	5	0	5	2	FRUIT
C	2	5	0	0	COMPUTER
D	1	2	1	7	COMPUTER
TEST INSTANCES					
T1	2	0	3	1	?
T2	1	2	1	0	?

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

   (ii). 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*.

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

2. Approximately 1% of women aged between 40 and 50 have breast cancer. 80% of mammogram screening tests detect breast cancer when it is there. 90% of mammograms DO NOT show breast cancer when it is **NOT** there[1]. Based on this information, complete the following table.

Cancer	**Probability**
No	99%
Yes	1%

Cancer	**Test**	**Probability**
Yes	Positive	80%
Yes	Negative	?
No	Positive	?
No	Negative	90%

3. Based on the results in question 2, calculate the **marginal probability** of 'positive' results in a Mammogram Screening Test.

4. Based on the results in question 2, calculate *P(Cancer = 'Yes' | Test = 'Positive')*, using the Bayes Rule.

[1] Remember these numbers are not accurate and simplified to ease the calculations in this question.