

School of Computing and Information Systems
The University of Melbourne
COMP30027 Machine Learning (Semester 1, 2021)

Week 3

- 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's **NOT** there¹. 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%

- Based on the results in question 1, calculate the **marginal probability** of 'positive' results in a Mammogram Screening Test.
- Based on the results in question 1, calculate $P(\text{Cancer} = \text{'yes'} \mid \text{Test} = \text{'positive'})$, using the Bayes Rule.
- Given the following dataset:

ID	Outl	Temp	Humi	Wind	PLAY
TRAINING INSTANCES					
A	s	h	n	F	N
B	s	h	h	T	N
C	o	h	h	F	Y
D	r	m	h	F	Y
E	r	c	n	F	Y
F	r	c	n	T	N
TEST INSTANCES					
G	o	m	n	T	?
H	?	h	?	F	?

- Explain which of the given instances are 'test' instances and which ones are the 'train' instances.
 - Build a Naïve Bayes model for the given training instances.
- Using the Naïve Bayes model that you developed in question 4, classify the given test instances.
 - No smoothing.
 - Using the "epsilon" smoothing method.
 - Using "Laplace" smoothing ($\alpha = 1$)

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