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

Tutorial exercises: Week 3

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	?

1. Build a probabilistic **model** based around the given training instances:

    (a) Calculate the **prior** probability $P(Outl = \text{s})$. Calculate the prior probabilities of the other attribute values in this data.

    (b) Find the **entropy** of (the distribution of the attribute values) for each of the six attributes, given this probabilistic model.

    (c) Calculate the **joint** probability $P(Outl = \text{s} \cap Temp = \text{h})$. Calculate some other joint probabilities, for pairs of attribute values from different attributes.

    (d) Calculate the **conditional** probability $P(Outl = \text{s}|Temp = \text{h})$. Calculate some other conditional probabilities.

2. Ensure that you can derive the **Naive Bayes** formulation.

3. Using the probabilistic model that you developed above, classify the test instances according to the method of **Naive Bayes**.

    (a) Using the "epsilon" smoothing method.

    (b) Using "Laplace" smoothing.