School of Computing and Information Systems
The University of Melbourne
COMP90049 Introduction to Machine Learning (Semester 2, 2020)
Tutorial exercises: Week 8

1. For the following dataset; Classify the test instances using the ID3 *Decision Tree* method:

ID	Outl	Temp	Humi	Wind	PLAY
		TRAINING INSTANCES			
A	s	h	h	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	c	n	T	?
H	s	m	h	F	?

   (i). Using the **Information Gain** as a splitting criterion

   (ii). Using the **Gain Ratio** as a splitting criterion

2. Given the same dataset, we wished to perform feature selection on this dataset, where the class is PLAY:

ID	Outl	Temp	Humi	Wind	PLAY
A	s	h	h	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

   (i). Which of Humi and Wind has the greatest *Pointwise Mutual Information* for the class Y? What about N?

   (ii). Which of the attributes has the greatest *Mutual Information* for the class, as a whole? (Note that we need to extend the lecture definition to handle non–binary attributes.)