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

Tutorial exercises: Week 4

1. Consider the following 10 instances, given so-called "gold standard" labels (assuming a 3-class problem), and the output of four supervised machine learning models:

Instance	Gold	①	②	③	④
1	A	A	A or B	A	A
2	B	A	B or C	A	?
3	A	A	A	A	A
4	C	C	B or C	A	?
5	B	B	A or B or C	A	?
6	C	A	A or C	A	?
7	C	A	A or B or C	A	?
8	A	C	A or B	A	A
9	A	A	A	A	?
10	A	A	A or C	A	A

   (a) Where possible, calculate the **accuracy** and **error rate** of the four models.

   (b) Where possible, calculate the **precision** and **recall**, treating class A as the "positive" class. Do the same for the B and C classes, in turn, and then calculate the **macro-averaged precision and recall**.

2. What is the difference between evaluating using a **holdout** strategy and evaluating using a **cross–validation strategy**?

   (a) What are some reasons we would prefer one strategy over the other?

3. For the following dataset:

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	?

   (a) Classify the test instances using the method of 0-R.

   (b) Classify the test instances using the method of 1-R.

   (c) Classify the test instances using the ID3 **Decision Tree** method:

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

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