School of Computing and Information Systems
The University of Melbourne
COMP90049 Knowledge Technologies (Semester 2, 2017)
Workshop exercises: Week 10

1. 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.

   For the following dataset:

apple	*ibm*	*lemon*	*sun*	CLASS
TRAINING INSTANCES				
4	0	1	1	FRUIT
5	0	5	2	FRUIT
2	5	0	0	COMPUTER
1	2	1	7	COMPUTER

2. Build a contingency table for each of the four attributes on the data collection above.

   (a) According to "Pointwise Mutual Information", which attribute has the best correlation with the class COMPUTER?

   (b) Use the method of "Mutual Information" to rank the "goodness" of the four features in predicting this two–class problem, according to the following formula:

$$MI(A,C) = \sum_{i\in\{a,\bar{a}\}} \sum_{j\in\{c,\bar{c}\}} P(i,j) \log_2 \frac{P(i,j)}{P(i)P(j)}$$