School of Computing and Information Systems
The University of Melbourne
COMP90049 Knowledge Technologies (Semester 1, 2017)
Workshop exercises: Week 10

1. A **confusion matrix** is an indication of the performance of a classifier over a set of test data, by counting the various output instances:

		Actual			
		a	*b*	*c*	*d*
Classified	*a*	10	2	3	1
	b	2	5	3	1
	c	1	3	7	1
	d	3	0	3	5

   (a) Calculate the classification **accuracy** of the system.

   (b) Calculate the **precision**, **recall**, **F-score** (where $\beta = 1$), **sensitivity**, and **specificity** for class *d*. (Why can't we do this for the whole system? How can we consider the whole system?)

2. How is **holdout** evaluation (like in the Project 2 data) different to **cross-validation** evaluation?

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	?

   Classify the test instances using a Decision Tree:

   (a) Using the Gini Index as a splitting criterion

   (b) Using the Information Gain as a splitting criterion