School of Computing and Information Systems The University of Melbourne COMP90049

Knowledge Technologies (Semester 1, 2018) Workshop exercises: Week 10

- 1. What is the **Naive Bayes** classifier? How does it classify data? What assumptions do we need to make about the data?
- 2. For the following dataset:

apple	ibm	lemon	sun	CLASS					
Training Instances									
Y	N	Y Y F		FRUIT					
Y	N	Y	Y	FRUIT					
Y	Y	N	N	COMPUTER					
Y	Y	Y	Y	COMPUTER					
Test Instances									
Y	N	Y	Y	?					
Y	N	Y	N	?					

Classify the test instances according to the method of Naive Bayes, as discussed in this subject.

3. 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	
	a	10	2	3	1	
Classified	b	2	5	3	1	
Classified	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?)
- 4. How is **holdout** evaluation (like in the Project 2 data) different to **cross-validation** evaluation?