Workshop: Week 5

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**.

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

2) 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
		TEST INSTANCES		
2	0	3	1	?
1	2	1	0	?

a) Using the **Euclidean distance** measure, classify the test instances using the 1-NN method.

b) Using the **Manhattan distance** measure, classify the test instances using the 3-NN method, for the three weightings we discussed in the lectures: *majority class*, *inverse distance ($\varepsilon = 1$)*, *inverse linear distance*.

c) Can we do weighted k-NN using **cosine similarity**?