

**CS480/580: Introduction to Artificial Intelligence**  
**Fall 2009**  
**(Lecture: R 1620-1900)**  
**Homework #6**  
**Points: 20**  
**Due: November 5, 2009**

1. Build a decision tree using the ID3 algorithm (using entropy gain concept) for a CS graduate admission decision system. You are provided with the following training data. Also, determine the accuracy of your decision tree with the given validation data. (i) Draw the decision tree (ii) Show the results of your validation (iii) Using the decision tree, determine the outcome of the test cases given below. **SHOW YOUR WORK**

**Training data**

Instance#	Degree	Major	GPA	GRE	Recommendations	Experience (Years)	Outcome
1	BS	CS	4.0	1450	Strong	0	Admit
2	BS	CS	2.5	900	Weak	0	Reject
3	BS	Non-CS	3.5	1250	Average	0	Admit
4	MS	Non-CS	3.0	1200	Average	2	Admit
5	BS	CS	2.9	1250	Strong	1	Admit
6	BS	Non-CS	2.95	1150	Strong	5	Admit
7	BS	CS	3.5	1050	Weak	0	Reject
8	BS	CS	3.1	1450	Average	0	Admit
9	BS	CS	3.3	1100	Average	0	Reject
10	BS	CS	3.2	1150	Strong	3	Admit
11	MS	CS	3.1	1250	Average	0	Admit
12	BS	Non-CS	3.4	1200	Weak	0	Reject
13	MS	Non-CS	3.3	1125	Weak	1	Reject
14	BS	CS	3.2	1350	Weak	0	Admit
15	BS	CS	3.0	950	Strong	0	Reject
16	BS	CS	3.4	1150	Weak	0	Reject
17	BS	CS	3.5	950	Average	0	Reject

**Validation data**

Instance#	Degree	Major	GPA	GRE	Recommendations	Experience (Years)	Outcome
1	BS	CS	2.9	1150	Weak	0	Reject
2	BS	CS	3.5	1210	Average	0	Accept
3	BS	Non-CS	3.5	1150	Strong	0	Admit
4	BS	Non-CS	3.0	1100	Average	2	Reject
5	MS	CS	3.1	1150	Average	0	Admit
6	MS	Non-CS	3.5	1250	Average	1	Admit
7	BS	CS	3.2	1050	Strong	0	Reject
8	BS	CS	3.3	1450	Average	0	Admit

9	BS	CS	3.3	1100	Average	4	Admit
---	----	----	-----	------	---------	---	-------

Test data

Instance#	Degree	Major	GPA	GRE	Recommendations	Experience (Years)	Outcome
1	BS	CS	3.5	1150	Weak	0	
2	BS	CS	2.9	1110	Strong	6	
3	BS	CS	3.7	950	Average	0	
4	BS	CS	3.2	1150	Weak	5	
5	MS	CS	3.0	1150	Average	0	
6	MS	CS	3.1	1250	Weak	0	
7	MS	Non-CS	3.2	1250	Average	1	
8	BS	Non-CS	3.7	1050	Average	0	
9	BS	Non-CS	3.3	1300	Average	2	
10	BS	Non-CS	3.0	1450	Strong	0	

- Define the cosine similarity between the following two clusters using:  
 (i) Single link (ii) Complete link metrics

Cluster 1: {<3, 5, 7, 3>, <4, 6, 5, 4>}; Cluster 2: {<2,4,4,4>, <6,5,6,4>, <5,6,6,4>}

- Cluster the following data into three clusters. Use the K-means algorithm (use the Euclidean distance): <1,5,6>, <7,7,7>, <-2, 9, 8>, <2,-5, -4>, <0,0,2>, <2,-2,3>, <3,3,3>, <4,1,2>, <6,5,4>, <-2,4,5>, <1,1,1>, <2,-2,-2>, <0, 1, 1>, <-1,-1,-1>