End Semester Examination

Course Name: Artificial Intelligence Code: CS 561

Full Marks-100 Time: 3 hours

Answer ALL the questions

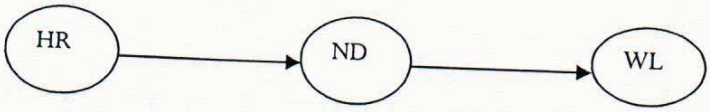
Make reasonable assumptions as and whenever necessary. You can answer the questions in any sequence. However, the answers to all the components of any particular question should appear together.

1. (a). Consider the weighted term vectors of three documents as:

$D_1 = 3T_1 + 2T_2 + 6T_3$ $D_2 = 5T_1 + 8T_2 + 4T_3$ $D_3 = 2T_1 + 4T_2 + 6T_3$

For a query, $Q = 15T_1 + 5T_2 + 15T_3$ compute the similarities using *inner product* and *cosine similarity* metrics. With respect to this problem, which one is the better measurement?

(b). For the following Bayesian network, the symbols HR, ND and WL are three random variables representing the events *Heavy rain, No drainage* and *Water logging*, respectively.



P(+hr)=0.1, P(+nd|+hr)=0.8, P(+nd| -hr)=0.1, P(+wl|+nd)=0.3, P(+wl|-nd)=0.1, where hr, nd and wl (both positive and negative) denote the evidences of the random variables. Compute P (WL| +hr) [***show each step clearly***]

10+ 10

2. (a). Distinguish between single objective optimization and multiobjective optimization. Define the terms "*Pareto optimality*", "*Non-dominated front*" with respect to multiobjective optimization. Write down the various steps of differential evolution. How is it different from genetic algorithm? What is time complexity of multinomial naïve Bayes model? **4+4+10+2**

3. Based on the data given in the following table, (i). estimate a multinomial Naïve Bayes classifier; (ii). Apply classifier to the test document (***Show each step with proper explanation***)

20

	docID	Words in document	In c= Sports?
Training set	1	Cricket Football Computers	yes
	2	Kabadi Hockey Chess	yes
	3	Laptop Desktop	no
	4	Sapporo Osaka Taiwan	no
Test set	5	Cricket Cricket Desktop	?

4. The following table summarizes a dataset with three attributes A,B,C and two class labels +, -. Build a two-level decision tree.

A	B	C	Number of instances	
			+	-
T	T	T	5	0
F	T	T	0	20
T	F	T	20	0
F	F	T	0	5
T	T	F	0	0
F	T	F	25	0
T	F	F	0	0
F	F	F	0	25

a. According the classification error rate, which attribute would be chosen as the first splitting attribute? For each attribute, show the gains in classification error rate.

b. How many instances are misclassified by the resulting decision tree?

20

5. Write short notes on the following topics (Any **FOUR**) **4*5=20**

(a). Syntactic ambiguities in NLP

(b). Forward search algorithms for HMM evaluation problem

(c). Viterbi decoding in HMM

(d). Bagging vs. boosting

(e). Multivariate naive Bayes classifier