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 parts of any particular question should appear together.

1. (a) A feedforward neural network where connections go from the current layer to the next layer ONLY is called a "pure feedforward network". Give an example to show that if this purity is NOT insisted on, that is, connections can jump layers, then one can save on the number of neurons.

5

(b) The derivation of backpropagation was done in the class. It was shown that the weight change rule based on gradient descent ALWAYS reduces error. Thus the algorithm is greedy and can get stuck in local minimum.

Introduce probability. At every iteration, gradient descent rule will give the weight change expression, but we decide to change the weight with probability 'p'. Thus there is a probability of not changing the weight amounting to '1-p'. Assume 'p' to be equal to sigmoid of the net input at the destination neuron for the connection, i.e., same as the transfer function of the neuron (input-output).

Show that this can solve the local minimum problem, probabilistically speaking. Show mathematically.

15

2. The following table summarizes a dataset with three attributes A, B, C and two class labels '+' and '—'. 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 contingency table and the gains in classification error rate.

b) Define the terms "over-fitting and under-fitting" with proper explanations. When and under what conditions sampling is required? How decision tree behaves under the following conditions:
   i.   All the records in a node belong to only one class,
   ii. Records at any particular node are equally distributed among all the classes, and
   iii. No record satisfies the particular condition.

12+8

3. Based on the data given in the following table,

    i.   Estimate a multivariate naïve Bayes classifier;

    ii.   Apply classifier to the test document (**Show each step with proper explanation**)

                 20

Dataset	docID	Words in document			In c= India?
Training set	1	Kolkata	Delhi	Patna	Yes
	2	Cricket	Mumbai	Kolkata	Yes
	3	Delhi	IITP	Patna	Yes
	4	Cricket	Japan	London	No
Test set	5	Kolkata	Patna	Mumbai	?

4. (a) Formulate a HMM by deriving the evaluation (Forward and Backward Recursions) and decoding (Viterbi) problems.

   (b) A knowledge base contains the following statements:
    Everyone who loves all animals is loved by someone.
    Anyone who kills an animal is loved by no one.
    John loves all animals.
    Either John or Jack killed Tuna, the cat.

    (i) Convert these statements into FOL.
    (ii) Convert each FOL statement in (i) to CNF.
    (iii) Using resolution, prove that **Jack killed the cat**.

                 10+10