JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

END SEMESTER EXAMINATION-2016

B.Tech (BI) VI Semester

COURSE CODE: 10B11BI612   MAX. MARKS: 35

COURSE NAME: Machine Learning for Bioinformatics

COURSE CREDITS: 04   MAX. TIME: 2 HRS

Note: All questions are compulsory. Carrying of mobile phone during examinations will be treated as case of unfair means.

---

1. A 3000 nt long DNA sequence is used as training set for parameter estimation of the DNA statistical model. The observed counts of dinucleotides are as follows:

	T	C	A	G
T	500	200	300	100
C	200	200	300	300
A	250	350	300	300
G	400	200	300	200

Find the maximum likelihood estimates of transition probabilities P_{AC} and P_{TG} of the first order Markov model for the DNA strand complementary to the training sequence. (3)

2. Distinguish between: (9+2)
(a) Mealy machine and Moore machine with example
(b) Viterbi decoding and posterior decoding with mathematical equations
(c) Context-free and context-sensitive grammar with its respective automatons
(d) Entropy and gain-ratio for getting ID trees

3. Explain the inherent limitations of the traditional backpropagation algorithm in ANN. Discuss the possible solutions. (3)

4. Distinguish between liberal and conservative performance in ROC analysis. Which zone should be used for test case prediction? Why and how do we compare ROC curves? (3)

5. Draw the Parse tree for the following: (5)
(a) PROSITE pattern representing the RNP-1 motif: RGQAFVIF
(b) RNA stem loop structure formed by the sequence: 5'-CAGGAACUG-3'

6. Why are psedocounts used in Forward-Backward algorithm? Why is it called EM algorithm. Elaborate. (5)

7. Derive the various palindromic strings from the following context-free gramarras with lengths between 8 and 16. S-> aSa|bSb|aa|bb. (5)