

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

T-3 EXAMINATION, May 2018

B.Tech (BI) VI Semester

COURSE CODE: 10B11BI612

MAX. MARKS: 35

COURSE NAME: Machine Learning in Bioinformatics

COURSE CREDITS: 04

MAX. TIME: 2 Hr

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

1. Why are pseudocounts used in Forward-Backward algorithm? Why is it called Expectation-Maximization algorithm? Elaborate by describing the algorithm. (4)
2. Distinguish between: (10)
 - (a) Sensitivity and PPV
 - (b) Viterbi decoding and posterior decoding
 - (c) Evaluation and decoding in HMM
 - (d) Deterministic and non-deterministic automaton
 - (e) Mealy machine and Moore machine
3. You are given 100 sequences from a cohort of individuals with a certain disease. You design an HMM to model these sequences. How will you calculate the log-likelihood of this model? How can you use this to optimize the model? (4)
4. Take any hypothetical example of an HMM (any kind of biological example), assuming any probabilities randomly. Describe how and why you would apply forward algorithm to this data? (4)
5. Derive the various palindromic strings from the following context-free grammar with lengths ranging between 8 and 16: $S \rightarrow aSa \mid bSb \mid aa \mid bb$. (4)
6. Write any two hypothetical RNA sequences that can form the hairpin structures. Construct a context-free grammar to describe these sequences. (4)
7. Describe the applications of kNN to the diagnosis of colorectal cancer using an electronic nose. (3)
8. Explain why we use cross-validation in machine learning. (2)