Dr Jayshree

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST-3 EXAMINATION – MAY 2019

B.Tech VIth Semester (Bioinformatics)

COURSE CODE: 16B11BI612 MAX. MARKS: 35

COURSE NAME: Datawarehousing and mining for bioinformatics

COURSE CREDITS: 4 MAX. TIME: 2HRS

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

---

1. Explain the difference between Inmon and Kimball models of datawarehousing. (5)
2. Explain the limitations of using single identification tree. How are these overcome by using a committee of decision tree approach? (2+6)
3. How do we use the partitioning approaches for clustering data in bioinformatics. Explain the k-means and k-medoids methods in this context. (3+3)
4. Explain briefly with steps how we perform PCA and describe its application briefly. (5)
5. For the following square matrix (5)

$$\begin{matrix} 3 & 0 & 1 \\ -4 & 1 & 2 \\ -6 & 0 & -2 \end{matrix}$$

Decide which, if any of the following vectors are eigen vectors of that matrix and give the corresponding eigen value.

$$\text{(a)}\begin{matrix} 2 \\ 2 \\ -1 \end{matrix} \quad \text{(b)}\begin{matrix} -1 \\ 0 \\ 2 \end{matrix} \quad \text{(c)}\begin{matrix} -1 \\ 1 \\ 3 \end{matrix} \quad (d)\begin{matrix} 0 \\ 1 \\ 0 \end{matrix} \quad (e)\begin{matrix} 3 \\ 2 \\ 1 \end{matrix}$$

6. Explain why gain ratio is a better mesure for identification trees than entropy. Highlight the limitations of identification tree. What is meant by overfitting in ID trees and how do we circumvent it? (6)