

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT
 TEST-2, EXAMINATION – 2016
 B. TECH. BIOINFORMATICS (VI SEM)

Course code: 10B11BI612

Time: 1.5 hours

Course name: Machine Learning for Bioinformatics

MM-25 marks

Q. 1. Cluster the genes for the following data: (5)

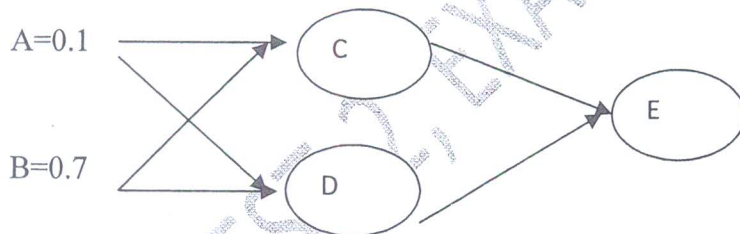
	Gene 1	Gene 2	Gene 3	Gene 4	Gene 5
Patient 1	3	2	3	1	0
Patient 2	2	1	1	0	0
Patient 3	1	0	0	1	3
Patient 4	0	1	2	0	3

Q.2 MCC and ROC analysis are better measures than other heuristics for evaluating the performance of a classifier. Comment. (2)

Q.3. Explain what is over-fitting and how cross-validation of different kinds is used to overcome it. (3)

Q.4. Explain Li's approach for obtaining the committee of decision trees. How is it better than using single identification trees? (5)

Q.5. Given that the target=1 and learning rate=1, perform a forward and a reverse pass on the neural network shown below: (5)



The values of the weights are $w_{AC}=0.1$, $w_{AD}=0.3$ while $w_{BC}=0.7$, $w_{BD}=0.2$. $w_{CE}=0.2$, $w_{DE}=0.1$

Q.6. Explain the following terms: (5)

- Confusion matrix
- Bias input
- Supervised machine learning
- Positive predictive value
- Error function