**Quiz 3 – Spring 2017**

**CS583: Data Mining and Text Mining**

Name:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ UID\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

This is a closed-book quiz.

|  |  |
| --- | --- |
|  | **Marks** |
| Q1 |  |
| Q2 |  |
| **Total** |  |

1. (10 marks) There is **only one best answer** for the multiple choice questions.

(a). In PU learning, what does the theory say about how the problem may be solved?

1. It can be solved as an optimization problem with labeled and unlabeled examples
2. It can be solved as a machine learning problem
3. It can be solved as a constrained optimization problem that tries to minimize the number of unlabeled examples labeled as positive with some constraints
4. It can be solved as a constrained optimization problem that tries to minimize the number of labeled examples labeled as negative with some constraints
5. The EM algorithm for learning from labeled and unlabeled examples uses the following strategy.
6. Runs the Naïve Bayesian classifier iteratively until it gets the best F score.
7. Runs the Naïve Bayesian classifier iteratively until it gets the best accuracy.
8. Runs the Naïve Bayesian classifier iteratively until it converges.
9. Runs the Naïve Bayesian classifier iteratively until it converges to the global optimum.
10. PageRank is based on
11. Degree prestige
12. Rank prestige
13. Proximity prestige
14. Betweenness prestige
15. The HITS algorithm produces
16. A hub score for each page
17. An authority score for each page
18. A hub score and an authority score for each page
19. A rank score for each page
20. What is the degree centrality of node *i*?

*i*

2. (10 marks) Given the hyperlink graph below:

3

4

6

1

2

5

We want to use the Markov chain model to compute the PageRank value of each node.

1. (3 marks) Give the initial transition probability matrix.
2. (3 marks) Is it a stochastic matrix? Why?
3. (2 marks) Is the matrix irreducible? Why?
4. (2 marks) Is the matrix aperiodic? Why?