Web information retrieval - 2017/2018

Exam - April 12th, 2019

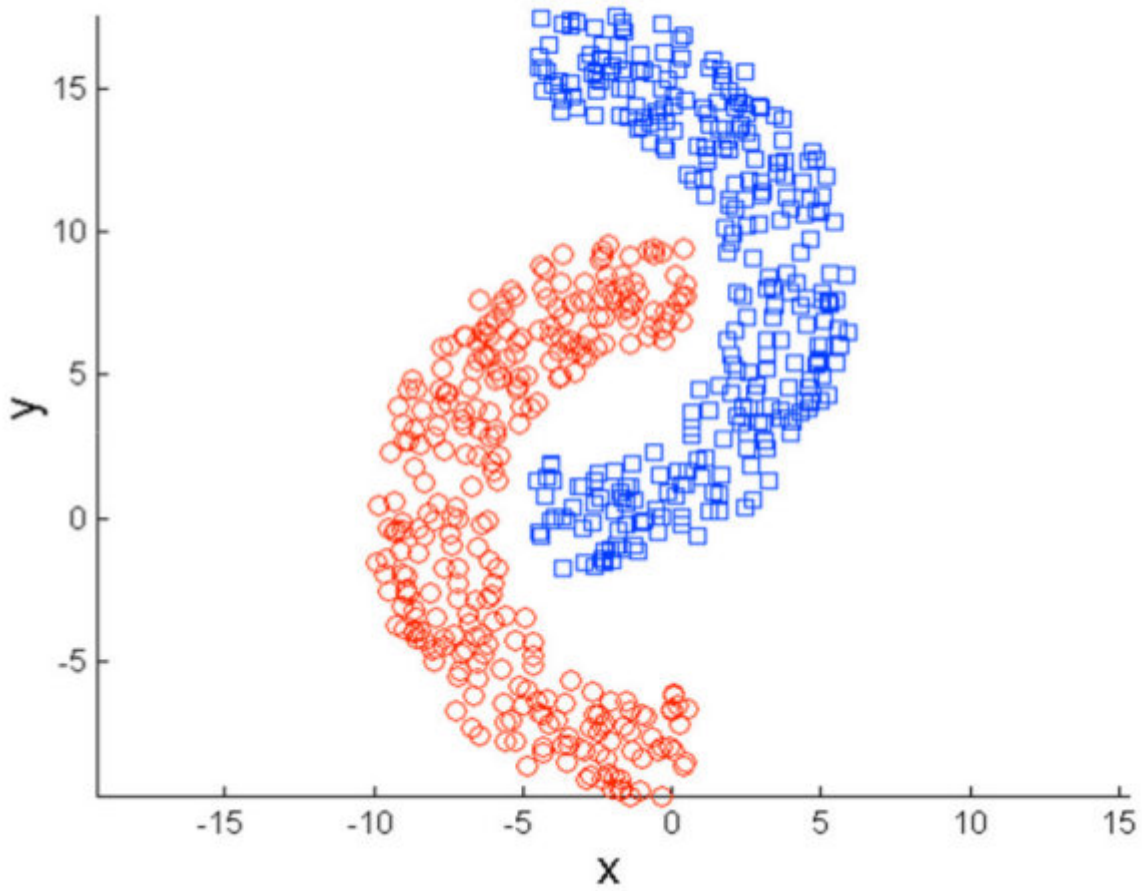
Time: 60 minutes

Assignment 1

Assume you have a boolean retrieval system. Write down the algorithm to answer the query T_1 `or` T_2 (i.e., to perform the union of the corresponding postings lists), where T_1 and T_2 are two terms of the vocabulary.

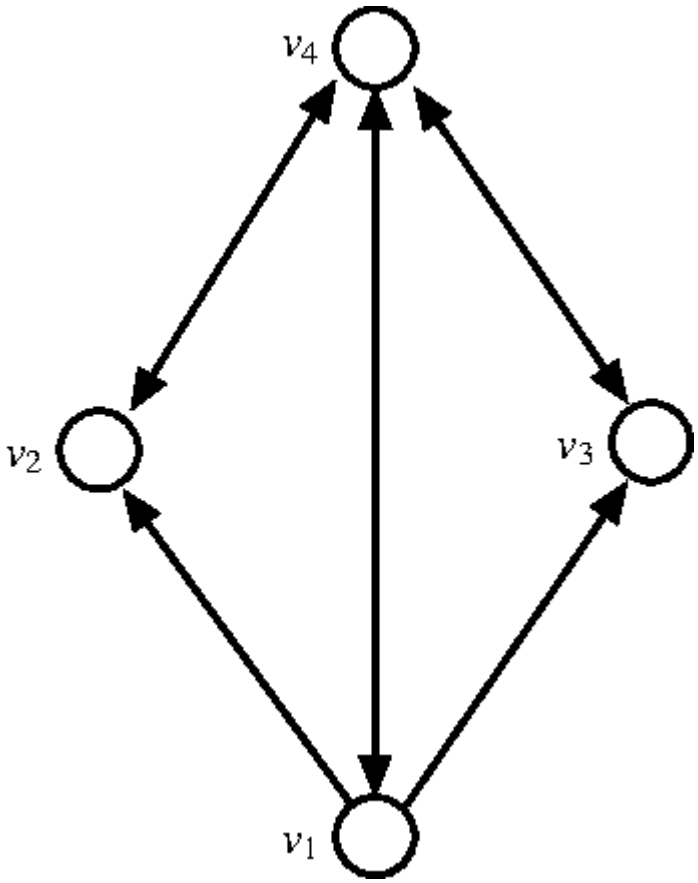
Assignment 2

Assume points in the picture below form a training set for a classification task, with two labels corresponding to the two clusters that are clearly present in the picture. Discuss which, between SVM (Support Vector Machines) and kNN you would use to perform classification. *You should clearly motivate your answer, referring to the properties of SVM and kNN*.



Assignment 3

We are given the following graph. Assume a general teleporting probability α.



3.1. Write down all the necessary equations needed to calculate the personalized pagerank with respect to the personalization vector $\{1, 0, 0, 0\}$.

3.2. Assume you have computed personalized pagerank vectors π_1 and π_2 corresponding to personalization vectors $\mathbf{p}_1 = \{1, 0, 0, 0\}^T$ and $\mathbf{p}_2 = \{0, 1, 0, 0\}^T$. Explain in detail how to calculate the personalized pagerank with respect to the personalization vector $\{0.5, 0.5, 0, 0\}$, *without* solving the corresponding system of linear equations from scratch.

Introduce whatever notation you think necessary.