Web information retrieval - 2017/2018

Exam - February 11th, 2019

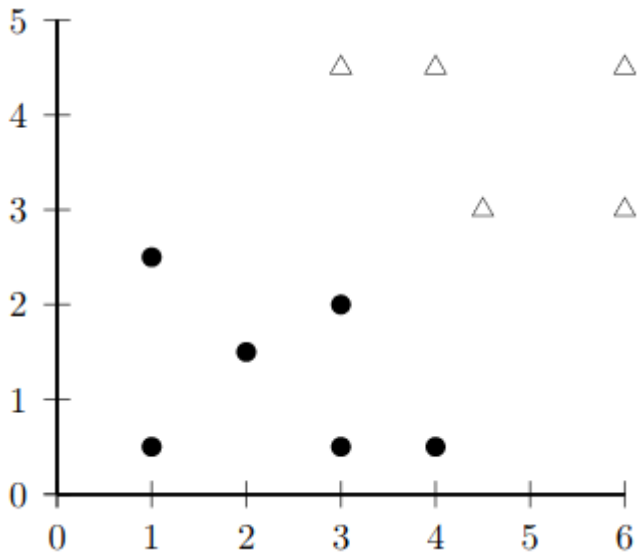
Time: 60 minutes

Assignment 1

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

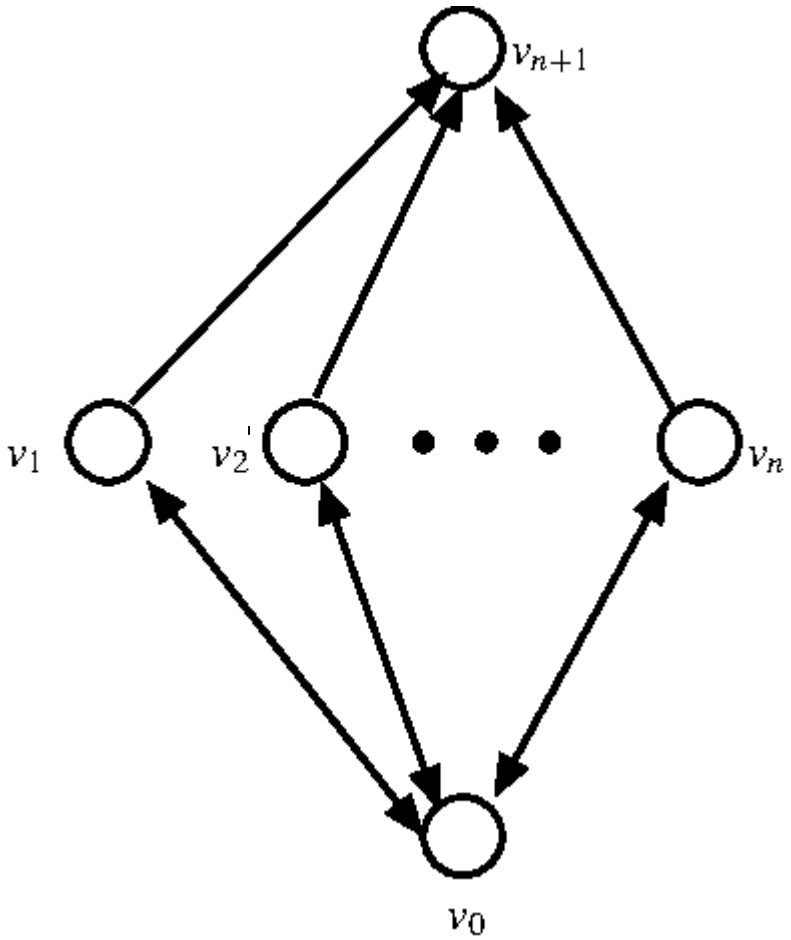
Assignment 2

In the picture below, identify the linear maximum margin (SVM) classifier for the binary problem triangle vs dot. *Draw three lines*: the two boundaries of the maximum margin and the maximum margin hyperplane. Which of the vectors are support vectors? You can solve this problem "visually" by drawing your solution into the figure.



Assignment 3

3.1. Work out the Pagerank equations for the network below with teleporting probability α, under the standard personalization vector $\frac{1}{n}\mathbf{1}$. **Hint:** help yourself with symmetries whenever possible.



3.2. Consider personalized Pagerank for the same network with teleporting probability α, this time with personalization vector $\mathbf{p} = (0, \ldots, 1)^T$ (i.e., all components are 0 except for the $(n+1)$-th). Compute the corresponding Pagerank vector π. *Motivate your answer*.

$$\pi = ((1-\alpha)M + \alpha p)\pi$$

$$\begin{bmatrix} \pi_1 \\ \vdots \\ \pi_i \\ \vdots \end{bmatrix} = (1-\alpha) \sum_{j \to i} \frac{\pi_j}{d_j} + \alpha \begin{bmatrix} p_1 \\ \vdots \\ p_i \\ \vdots \end{bmatrix}$$

$$\pi = (1-\alpha)M\pi + \alpha p$$

$$p_i = \begin{cases} \frac{1}{|S|} & i \in S \\ 0 \end{cases}$$