University of Louisiana at Lafayette

CSCE 561: Information Storage and Retreival

Assignment 3

Md. Enamul Haque
CLID: mxh5576

November 29, 2017

Contents

1.

$$A = \begin{array}{c|cccc} & d_1 & d_2 & d_3 & d_4 \\ \hline t_1 & 3 & 0 & 1 & 3 \\ t_2 & 0 & 1 & 3 & 0 \\ t_3 & 3 & 0 & 0 & 2 \\ t_4 & 2 & 0 & 1 & 3 \end{array}$$

a

$$G_t = I \quad , \text{ so. } \quad t_i . t_i = 1$$

$$q = 2\,\bar{t}_1 + \bar{t}_3 = 2\bar{t}_1 + 0.\bar{t}_2 + \bar{t}_3 + 0.\bar{t}_4$$

$$RSV_q = A^T G_t q^T$$

$$= \begin{bmatrix} 3 & 0 & 3 & 2 \\ 0 & 1 & 0 & 0 \\ 1 & 3 & 0 & 1 \\ 3 & 0 & 2 & 3 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 3 & 0 & 3 & 2 \\ 0 & 1 & 0 & 0 \\ 1 & 3 & 0 & 1 \\ 3 & 0 & 2 & 3 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

$$= \begin{array}{c} \begin{bmatrix} 9 \\ 0 \\ 2 \\ 8 \end{bmatrix} \end{array} \begin{array}{l} d_1 \\ d_2 \\ d_3 \\ d_4 \end{array}$$

$\therefore$ RSVs of d_1 through d_4 with respect to q are 9, 0, 2, 8 respectively

(b)

$$RSV_q = \begin{bmatrix} 3 & 0 & 3 & 2 \\ 0 & 1 & 0 & 0 \\ 1 & 3 & 0 & 1 \\ 3 & 0 & 2 & 3 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0.1 & -0.4 & 0.5 \\ 0.1 & 1 & -0.3 & 0.2 \\ -0.4 & -0.3 & 1 & 0.1 \\ 0.5 & 0.2 & 0.1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 3 & 0 & 3 & 2 \\ 0 & 1 & 0 & 0 \\ 1 & 3 & 0 & 1 \\ 3 & 0 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1.6 \\ -0.1 \\ 0.2 \\ 1.1 \end{bmatrix}$$

$$= \begin{array}{cl} \begin{bmatrix} 7.6 \\ -0.1 \\ 2.4 \\ 8.5 \end{bmatrix} & \begin{array}{l} d_1 \\ d_2 \\ d_3 \\ d_4 \end{array} \end{array}$$

(c) Both documents d_3 and d_4 has greater RSV's in part (b) compared to part (a). The change in RSV's are caused due to the change in weights of Grammian matrix, G_t, which is responsible for representing linear dependence between term-term pairs. In this particular scenario, (t_3, t_1) pair weight in G_t caused change in RSV for d_3. Similarly, (t_4, t_1) and (t_4, t_3) pair caused change in RSV for d_4.

(d) Yes The RSV of a document can be smaller when G_t is incorporated. We can explain the scenario from the components of Grammian matrix, G_t. In the standard G_t, the components other than (t_i, t_i) location (diagonal location) are zeros. Therefore, the overall effect on RSV for document i depends only on the (t_i, t_i) value of G_t. On the contrary, (t_i, t_j) locations can have both +ve and −ve values in the G_t. In that case, we can have the below relation for RSVs with respect to query and G_t.

if $G_{t'} \cdot \bar{q} < G_t \cdot \bar{q}$, then the RSV

for corresponding documents are lower when new $G_{t'}$ is incorporated inplace of G_t. It is to be noted that (t_i, t_j) values become −ve when the angle (θ) between them is between $90°$ and $270°$, for all other cases, the values are +ve.

$$\therefore \; (t_i \cdot t_j) = \begin{cases} -ve, & 90° < \theta < 270° \\ +ve, & 0° < \theta < 90° \text{ or } 270° < \theta < 360° \\ 0, & \text{elsewhere} \end{cases}$$

(e) In GVSM,

$$G_t = AA^T = \begin{bmatrix} 19 & 3 & 15 & 16 \\ 3 & 10 & 0 & 3 \\ 15 & 0 & 13 & 12 \\ 16 & 3 & 12 & 14 \end{bmatrix}$$

$$\therefore RSV_q = [\bar{d}_\alpha]_{A^T} \cdot G_t \cdot [q^T]_{A^T}$$

$$= \begin{bmatrix} 3 & 0 & 3 & 2 \\ 0 & 1 & 0 & 0 \\ 1 & 3 & 0 & 1 \\ 3 & 0 & 2 & 3 \end{bmatrix} \cdot \begin{bmatrix} 19 & 3 & 15 & 16 \\ 3 & 10 & 0 & 3 \\ 15 & 0 & 13 & 12 \\ 16 & 3 & 12 & 14 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 3 & 0 & 3 & 2 \\ 0 & 1 & 0 & 0 \\ 1 & 3 & 0 & 1 \\ 3 & 0 & 2 & 3 \end{bmatrix} \cdot \begin{bmatrix} 53 & 6 & 43 & 44 \end{bmatrix}$$

$$= \begin{array}{cccc} \begin{bmatrix} 376 & 6 & 115 & 377 \end{bmatrix} \\ d_1 \quad d_2 \quad d_3 \quad d_4 \end{array}$$

RSV.s are 376, 6, 115, 377 respectively.

```
#!/usr/bin/env python3
# -*- coding: utf-8 -*-
"""
Created on Sun Nov 26 19:06:46 2017

@author:
    Md Enamul Haque

Desc:
    The order is: d1, d3, d5, d6, d7 in the D matrix.
    d6 and d7 are negated due to the non-relevancy.
    Mapping in the Yset: 0-->d1, 1-->d3, 2-->d5, 3-->d6, 4-->d7
*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*
"""

import numpy as np

def get_yset(DQ):
    yset = list()
    for i in range(len(DQ)):
        if DQ[i] <= 0:
            yset.append(i)

    return yset

def update_q(q, yset):
    l = len(yset)
    Dvec = np.array(D.shape[1] * [0])
    if l>0:
        for i in yset:
            Dvec = Dvec + D[i]
    q = q + Dvec

    return q

if __name__ == "__main__":
    print(__doc__)

    iteration_count = 500
    empty = "\u03a6"
    global D

    D = np.array([
                  [0,2,0,2],
                  [1,3,0,0],
                  [1,3,1,0],
                  [0,-2,-1,-1],
                  [0,-3,-1,-2]
                  ])

#    D = np.array([
#              [0,2,0,2],
#              [1,3,1,0],
#              [0,-3,-1,-2],
#              [1,3,0,0],
#              [0,-2,-1,-1]
#              ])


#    D = np.array([
#              [0,-1,-1,0],
#              [0,0,-1,1],
```

Q2

(a)
$$\hat{d}_1 = [0 \quad 2 \quad 0 \quad 2]$$
$$\hat{d}_3 = [1 \quad 3 \quad 0 \quad 0]$$
$$\hat{d}_5 = [1 \quad 3 \quad 1 \quad 0]$$
$$\hat{d}_6 = [0 \quad -2 \quad -1 \quad -1]$$
$$\hat{d}_7 = [0 \quad -3 \quad -1 \quad -2]$$

Let's assume, $\bar{q}^0 = [0 \quad 0 \quad 0 \quad 0]$

$$y(q^0) = \{d_1, d_3, d_5, d_6, d_7\}$$

$[y(q^0) = \{d_\alpha \mid d_\alpha q^T < 0]$

$$\bar{q}_1 = \bar{q}_0 + \sum_i \hat{d}_i$$

$$= [2 \quad 3 \quad -1 \quad -1]$$

$$d_\alpha \cdot \bar{q}_1^T = \begin{bmatrix} 0 & 2 & 0 & 2 \\ 1 & 3 & 0 & 0 \\ 1 & 3 & 1 & 0 \\ 0 & -2 & -1 & -1 \\ 0 & -3 & -1 & -2 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 3 \\ -1 \\ -1 \end{bmatrix} = \begin{bmatrix} 4 \\ 11 \\ 10 \\ -4 \\ -6 \end{bmatrix}$$

$$\therefore \quad y(q_1) = \{ d_6 , d_7 \}$$

$$\bar{q}_2 = \bar{q}_1 + \sum_{d_i \in y(q_1)} \hat{d}_i$$

$$= \begin{bmatrix} 2 & 3 & -1 & -1 \end{bmatrix} + \begin{bmatrix} 0 & -5 & -2 & -3 \end{bmatrix}$$

$$= \begin{bmatrix} 2 & -2 & -3 & -4 \end{bmatrix}$$

- - - - - - - - - - - - - - - - - -

$$d_\alpha \cdot \bar{q}_2^T = \begin{bmatrix} 0 & 2 & 0 & 2 \\ 1 & 3 & 0 & 0 \\ 1 & 3 & 1 & 0 \\ 0 & -2 & -1 & -1 \\ 0 & -3 & -1 & -2 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ -2 \\ -3 \\ -4 \end{bmatrix}$$

$$= \begin{bmatrix} -12 \\ -4 \\ -7 \\ 11 \\ 17 \end{bmatrix}$$

$$\therefore \quad y(q_2) = \{ d_1 , d_3 , d_5 \}$$

$$\bar{q}_3 = \begin{bmatrix} 2 & -2 & -3 & -4 \end{bmatrix} + \begin{bmatrix} 2 & 8 & 1 & 2 \end{bmatrix}$$

$$= \begin{bmatrix} 4 & 6 & -2 & -2 \end{bmatrix}$$

$$d_\alpha \cdot q_3^T = \begin{array}{cc} \begin{bmatrix} 8 \\ 22 \\ 20 \\ -8 \\ -12 \end{bmatrix} & \begin{array}{c} d_1 \\ d_3 \\ d_5 \\ d_6 \\ d_7 \end{array} \end{array}$$

$$\therefore \quad y(q_3) = \{ d_6, d_7 \}$$

$$\bar{q}_4 = [4 \quad 6 \quad -2 \quad -2] + [0 \quad -5 \quad -2 \quad -3]$$

$$= [4 \quad 1 \quad -4 \quad -5]$$

$$d_\alpha \cdot \bar{q}_4^T = \begin{bmatrix} -8 \\ 7 \\ 3 \\ 7 \\ 11 \end{bmatrix}$$

$$\therefore \quad y(q_4) = \{ d_1 \}$$

$$\bar{q}_5 = [4 \quad 1 \quad -4 \quad -5] + [0 \quad 2 \quad 0 \quad 2]$$

$$= [4 \quad 3 \quad -4 \quad -3]$$

$$d_\alpha q_5^T = \begin{bmatrix} 0 \\ 13 \\ 9 \\ 1 \\ 1 \end{bmatrix}$$

$$\therefore \quad y(q_5) = \{d_1\}$$

$$\bar{q}_6 = [4 \quad 3 \quad -4 \quad -3] + [0 \quad 2 \quad 0 \quad 2]$$
$$= [4 \quad 5 \quad -4 \quad -1]$$

$$d_\alpha \bar{q}_6^T = \begin{bmatrix} 8 \\ 19 \\ 15 \\ -5 \\ -9 \end{bmatrix}$$

$$\therefore \quad y(q_6) = \{d_6, d_7\}$$

$$\bar{q}_7 = [4 \quad 5 \quad -4 \quad -1] + [0 \quad -5 \quad -2 \quad -3]$$
$$= [4 \quad 0 \quad -6 \quad -4]$$

$$d_\alpha \bar{q}_7^T = \begin{bmatrix} -8 \\ 4 \\ -2 \\ 10 \\ 14 \end{bmatrix}$$

$$\therefore \quad y(q_7) = \{d_1, d_5\}$$

$$\bar{q}_8 = [4 \quad 0 \quad -6 \quad -4] + [1 \quad 5 \quad 1 \quad 2]$$
$$= [5 \quad 5 \quad -5 \quad -2]$$

$$d_\alpha \bar{q}_8^T = \begin{bmatrix} 14 \\ 20 \\ 15 \\ -7 \\ -14 \end{bmatrix}$$

$$\therefore \; y(q_8) = \{d_6, d_7\}$$

$$\bar{q}_9 = [5 \quad 5 \quad -5 \quad -2] + [0 \quad -5 \quad -2 \quad -3]$$
$$= [5 \quad 0 \quad -7 \quad -5]$$

$$d_\alpha \bar{q}_9^T = \begin{bmatrix} -10 \\ 5 \\ -2 \\ 12 \\ 17 \end{bmatrix}$$

$$\therefore \; y(q_9) = \{d_1, d_5\}$$

$$\bar{q}_{10} = [5 \quad 0 \quad -7, \quad -5] + [1 \quad 5 \quad 1 \quad 2]$$
$$= [6 \quad 5 \quad -6 \quad -3]$$

$$d_\alpha \bar{q}_{10}^T = \begin{bmatrix} 4 \\ 21 \\ 15 \\ -1 \\ -3 \end{bmatrix}$$

$$\therefore \; y(q_{10}) = \{d_6, d_7\}$$

$$\bar{q}_{11} = [6 \;\; 5 \;\; -6 \;\; -3] + [0 \;\; -5 \;\; -2 \;\; -3]$$
$$= [6 \;\; 0 \;\; -8 \;\; -6]$$

$$d_\alpha \bar{q}_{11}^T = \begin{bmatrix} -12 \\ 6 \\ -2 \\ 14 \\ 20 \end{bmatrix}$$

$$\therefore \; y(q_{11}) = \{d_1, d_5\}$$

$$\bar{q}_{12} = [6 \;\; 0 \;\; -8 \;\; -6] + [1 \;\; 5 \;\; 1 \;\; 2]$$
$$= [7 \;\; 5 \;\; -7 \;\; -4]$$

$$d_\alpha \bar{q}_{12}^T = \begin{bmatrix} 2 \\ 22 \\ 15 \\ 1 \\ 0 \end{bmatrix}$$

$$\therefore \; y(q_{12}) = \{d_7\}$$

$$\bar{q}_{13} = \begin{bmatrix} 7 & 2 & -8 & -6 \end{bmatrix}$$

$$d_\alpha \bar{q}_{13}^T = \begin{bmatrix} -8 \\ 13 \\ 5 \\ 10 \\ 14 \end{bmatrix}$$

$$\therefore \; y(q_{13}) = \{ d_1 \}$$

$$\bar{q}_{14} = \begin{bmatrix} 7 & 4 & -8 & -4 \end{bmatrix}$$

$$d_\alpha \bar{q}_{14}^T = \begin{bmatrix} 0 \\ 19 \\ 11 \\ 4 \\ 4 \end{bmatrix} \qquad \therefore \; y(q_{14}) = \{ d_1 \}$$

$$\bar{q}_{15} = \begin{bmatrix} 7 & 6 & -8 & -2 \end{bmatrix}$$

$$d_\alpha \bar{q}_{15}^T = \begin{bmatrix} 8 \\ 25 \\ 17 \\ -2 \\ -6 \end{bmatrix} \qquad \therefore \; y(q_{15}) = \{ d_6, d_7 \}$$

$$\bar{q}_{16} = \begin{bmatrix} 7 & 1 & -10 & -5 \end{bmatrix}$$

$$d_\alpha \bar{q}_{16}^T = \begin{bmatrix} -8 \\ 10 \\ 0 \\ 13 \\ 17 \end{bmatrix} \qquad \therefore \; y(q_{16}) = \{ d_1 \}$$

$$\bar{q}_{17} = [8 \quad 6 \quad -9 \quad -3]$$

$$d_\alpha \bar{q}_{17}^T = \begin{bmatrix} 6 \\ 26 \\ 17 \\ 0 \\ -3 \end{bmatrix} \quad \therefore \quad y(q_{17}) = \{d_6, d_7\}$$

$$\bar{q}_{18} = [8 \quad 1 \quad -11, -6]$$

$$d_\alpha \bar{q}_{18}^T = \begin{pmatrix} -10 \\ 11 \\ 0 \\ 15 \\ 20 \end{pmatrix} \quad \therefore \quad y(q_{18}) = \{d_1, d_5\}$$

$$\bar{q}_{19} = [9 \quad 6 \quad -10 \quad -4]$$

$$d_\alpha \bar{q}_{19}^T = \begin{pmatrix} 4 \\ 27 \\ 17 \\ 2 \\ 0 \end{pmatrix} \quad \therefore \quad y(q_{19}) = \{d_7\}$$

$$\bar{q}_{20} = [9 \quad 3 \quad -11 \quad -6]$$

$$d_\alpha \bar{q}_{20}^T = \begin{pmatrix} -6 \\ 18 \\ 7 \\ 11 \\ 14 \end{pmatrix} \quad \therefore \quad y(q_{20}) = \{d_1\}$$

$$\bar{q}_{21} = \begin{bmatrix} 9 & 5 & -11 & -4 \end{bmatrix}$$

$$d_\alpha \bar{q}_{21}^{T} = \begin{pmatrix} 2 \\ 24 \\ 13 \\ 5 \\ 4 \end{pmatrix}$$

Optimal Query $= \begin{bmatrix} 9 & 5 & -11 & -4 \end{bmatrix}$

(b)

$$b_1 = [0 \quad -1 \quad -1 \quad 0]$$
$$b_2 = [0 \quad 0 \quad -1 \quad 1]$$
$$b_3 = [1 \quad 0 \quad 0 \quad -2]$$
$$b_4 = [1 \quad 1 \quad 0 \quad -1]$$
$$b_5 = [1 \quad 0 \quad -1 \quad -2]$$
$$b_6 = [1 \quad 1 \quad -1 \quad -1]$$

Initialize, $\bar{q}_0 = [0 \quad 0 \quad 0 \quad 0]$

$$y(q_0) = \{b_1, b_2, b_3, b_4, b_5, b_6\}$$

$\left[y(q_0) = \{ d_\alpha \mid d_\alpha \bar{q}_0^T \leq 0 \} \right]$

$$\bar{q}_1 = \bar{q}_0 + \sum_{b_i \in y(q_0)} b_i$$

$$= [4 \quad 1 \quad -4 \quad -5]$$

$$b_\alpha \bar{q}_1^T = \begin{bmatrix} 3 \\ -1 \\ 14 \\ 10 \\ 18 \\ 14 \end{bmatrix} \qquad \therefore\ y(q_1) = \{b_2\}$$

$$\bar{q}_2 = [4 \quad 1 \quad -5 \quad -4]$$

$$b_\alpha \bar{q}_2^T = \begin{bmatrix} 4 \\ 1 \\ 12 \\ 9 \\ 17 \\ 14 \end{bmatrix} \qquad \therefore\ \text{Optimal Query} = [4, 1, -5, -4]$$

(c) $\bar{q}_0 = [0 \ 0 \ 0 \ 0]$

For $b_1 \Rightarrow$

$b_1 \cdot q_0^T = 0 \longrightarrow b_1$ is misclassified

$\therefore \bar{q}_1 = \bar{q}_0 + \bar{b}_1$

$= [0 \ -1 \ -1 \ 0]$

For $b_2 \Rightarrow$

$b_2 \cdot q_1^T = 1 \longrightarrow b_2$ is correctly classified

$\therefore \bar{q}_2 = \bar{q}_1$

For $b_3 \Rightarrow$

$b_3 \cdot \bar{q}_2^T = 0 \longrightarrow b_3$ is misclassified

$\therefore \bar{q}_3 = \bar{q}_2 + \bar{b}_3$

$= [1 \ -1 \ -1 \ -2]$

For $b_4 \Rightarrow$

$b_4 \cdot \bar{q}_3^T = 2 \longrightarrow b_4$ is correctly classified

$\therefore \bar{q}_4 = \bar{q}_3$

For $b_5 \Rightarrow$

$b_5 \cdot \bar{q}_4^T = 6 \longrightarrow b_5$ is correctly classified

$\therefore \bar{q}_5 = \bar{q}_4$

For $b_6 \Rightarrow$

$b_6 \cdot \bar{q}_5^T = 3 \longrightarrow b_6$ is correctly classified

$\therefore \bar{q}_6 = \bar{q}_5 = [1 \quad -1 \quad -1 \quad -2]$

(d) Rocchio's method based optimal query representation:

$$q^* = \kappa \left[\frac{1}{n_0} \sum_{d' \in rel} \frac{\underline{d}'}{\|\underline{d}'\|} - \frac{1}{n_1} \sum_{d \in nrel} \frac{\underline{d}}{\|\underline{d}\|} \right]$$

considering $\kappa = 1$,

$$q^* = \frac{1}{3} \left(\frac{\underline{d}_1}{\|\underline{d}_1\|} + \frac{\underline{d}_3}{\|\underline{d}_3\|} + \frac{\underline{d}_5}{\|\underline{d}_5\|} \right) - \frac{1}{2} \left(\frac{\underline{d}_6}{\|\underline{d}_6\|} + \frac{\underline{d}_7}{\|\underline{d}_7\|} \right)$$

$$= \frac{1}{3} \left[\frac{(0,2,0,2)}{2\sqrt{2}} + \frac{(1,3,0,0)}{\sqrt{10}} + \frac{(1,3,1,0)}{\sqrt{11}} \right]$$

$$- \frac{1}{2} \left[\frac{(0,2,1,1)}{\sqrt{6}} + \frac{(0,3,1,2)}{\sqrt{14}} \right]$$

$$= (0{\cdot}21,\ 1{\cdot}09,\ 0{\cdot}1,\ 0{\cdot}47) - (0,\ 0{\cdot}9,\ 0{\cdot}33,\ 0{\cdot}47)$$

$$= (0{\cdot}21,\ 0{\cdot}19,\ -0{\cdot}23,\ 0)$$

Q3. i) standard perceptron criteria:

Optimal Query = $[9, 5, -11, -4] = q$

$$RSV = W'\underline{q}^T = \begin{bmatrix} 2 \\ 22 \\ -6 \\ 19 \\ -2 \end{bmatrix} \begin{array}{l} d2 \\ d4 \\ d8 \\ d9 \\ d10 \end{array}$$

document ranking = d4, d9, d2, d10, d8

Relevance ranking = REL, REL, REL, NREL, NREL

$$\therefore R_{norm} = \frac{1}{2}\left(1 + \frac{I^+ - I^-}{I^+_{max}}\right) = 1 \quad \left[I^+ = 6,\ I^- = 0,\ I^+_{max} = 6 \right]$$

ii) Generalized perceptron:

$$\underline{q}_{opt} = [4, 1, -5, -4]$$

$$RSV = W'\underline{q}_{opt} = [-6 \quad 6 \quad -4 \quad 6 \quad -7]$$

document ranking = d9, d4, d8, d2, d10

Relevance ranking = REL, REL, NREL, REL, NREL

$\therefore I^+ = 5,\ I^- = 1,\ I^+_{max} = 6$

$\therefore R_{norm} = 0'83$

iii) Generalized perceptron – learning by sample

$$\underline{q}_{opt} = [1 \ -1 \ -1 \ 2]$$

$$RSV = W'\underline{q}_{opt} = [2 \ -2 \ -2 \ -1 \ -2]$$

document ranking = d2, d9, d10, d8, d4

Relevance ranking = REL, REL, NREL, NREL, NREL

$$\therefore I^+ = 4, \ I^- = 2, \ I^+_{max} = 6$$

$$\therefore R_{norm} = 0{\cdot}67$$

iv) Rocchio's method

$$\underline{q}_{opt} = [0{\cdot}21, \ 0{\cdot}19, -0{\cdot}23, 0]$$

$$RSV = [0{\cdot}38, \ 0{\cdot}76, \ -0{\cdot}04, 0{\cdot}59, 0{\cdot}32]$$

document ranking = d4, d9, d2, d10, d8

Relevance ranking = REL, REL, REL, NREL, NREL

$$I^+ = 6, \ I^- = 0, \ I^+_{max} = 6$$

$$\therefore R_{norm} = 1.$$

Method	R_{norm}
standard perceptron	1·0
Generalized Perceptron	0·83
GP – learning by sample	0·67
Rocchio's	1.0

```
#!/usr/bin/env python3
# -*- coding: utf-8 -*-
"""
Created on Sun Nov 26 19:06:46 2017

@author:
    Md Enamul Haque
"""
import numpy as np
import sys


def getIplusMax(relevance):
    #unique_elements, counts_elements = np.unique(relevance, return_counts=True)
    #iplusmax = counts_elements[0] * counts_elements[1]
    iplusmax = relevance.count('REL') * relevance.count('NREL')

    return iplusmax

def getIplus(relevance):

    iplus = 0
    for i in range(len(relevance)):
        if relevance[i] == 'REL':
            c = relevance[i+1:].count('NREL')
            iplus = iplus + c

    return iplus

def getIminus(relevance):

    iminus = 0
    for i in range(len(relevance)):
        if relevance[i] == 'NREL':
            c = relevance[i+1:].count('REL')
            iminus = iminus + c

    return iminus


def get_Rnorm(relevance):

    iplusmax = getIplusMax(relevance)
    iplus = getIplus(relevance)
    iminus = getIminus(relevance)
    print("I+: ", iplus)
    print("I-: ", iminus)
    print("I+ max: ", iplusmax)
    rnorm = 0.5 * (1 + ((iplus-iminus)/iplusmax))

    return rnorm


if __name__ == "__main__":
    global D

    doclist = [2,4,8,9,10]
    rel_list = ['REL','REL','NREL','REL','NREL']

    D = np.array([
                 [0,2,0,2],
                 [2,3,1,0],
```

CMPS 561 Assignment #3, Fall 201~~0~~7

Vijay V Raghavan

Assigned: November 16, 2017
Due: November 29, 2017

Note:

1. All details of work for each question must be submitted.
2. Staple the question and answer sheet together
3. Make a cover with Name, CLID
4. Number all pages and give an index to each question.
*5. Most importantly, any sort of cheating will **NOT** be tolerated. More information can be found on class Web page on cheating policy.*

Q1. **30 Points**

In the Vector Space Model, the relationships among different terms can be expressed as a term-term matrix G_t, which is called the Grammarian matrix. The term-document relationship is shown in Table 1.

	d_1	d_2	d_3	d_4
t_1	3	0	1	3
t_2	0	1	3	0
t_3	3	0	0	2
t_4	2	0	1	3

Table 1

(a). If $G_t = I$, and that a query is given by $q = 2\vec{t}_1 + \vec{t}_3$, calculate the RSVs for d_1 through d_4 with respect to q.

(b). Repeat part a) if $$\boldsymbol{G}_t = \begin{bmatrix} t_1.t_1 & t_1.t_2 & t_1.t_3 & t_1.t_4 \\ t_2.t_1 & t_2.t_2 & t_2.t_3 & t_2.t_4 \\ t_3.t_1 & t_3.t_2 & t_3.t_3 & t_3.t_4 \\ t_4.t_1 & t_4.t_2 & t_4.t_3 & t_4.t_4 \end{bmatrix} = \begin{bmatrix} 1 & 0.1 & -0.4 & 0.5 \\ 0.1 & 1 & -0.3 & 0.2 \\ -0.4 & -0.3 & 1 & 0.1 \\ 0.5 & 0.2 & 0.1 & 1 \end{bmatrix}$$

(c). Pick a document for which RSV in part (b) is greater than that in part (a) and explain which element(s) (term relationships) from G_t cause this change.

(d). Can the RSV of a document become smaller when G_t is incorporated into the RSV computation? If yes, explain what the characteristics of term-term relationships matrix are that will cause this effect.

(e). Compute the RSVs for d_1 through d_4 with respect to q, assuming the GVSM model is employed. In this case, no assumption about G_t is needed.