

CSCE 561 Assignment #3, Fall 2019

Dr. Vijay V Raghavan

Assigned: November 14, 2019

Due: November 27, 2019

(TOPICS: VSM, GVSM, PCA, Generalized PCA, Rocchio's Method)

Note:

1. You must show all details of work for each question
 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 Gramian matrix. The term-document relationship is shown in Table 1.

	<u>d</u> 1	<u>d</u> 2	<u>d</u> 3	<u>d</u> 4
<u>t</u> 1	3	0	1	3
<u>t</u> 2	0	1	3	0
<u>t</u> 3	3	0	0	2
<u>t</u> 4	2	0	1	3

Table 1

(a). If $G_t = I$, and that a query is given by $q = 2t_1 + t_3$, calculate the RSVs for d_1 through d_4 with respect to q .

(b). Repeat part (a) if

$$G_t = \begin{bmatrix} t_1 \cdot t_1 & t_1 \cdot t_2 & t_1 \cdot t_3 & t_1 \cdot t_4 \\ t_2 \cdot t_1 & t_2 \cdot t_2 & t_2 \cdot t_3 & t_2 \cdot t_4 \\ t_3 \cdot t_1 & t_3 \cdot t_2 & t_3 \cdot t_3 & t_3 \cdot t_4 \\ t_4 \cdot t_1 & t_4 \cdot t_2 & t_4 \cdot t_3 & t_4 \cdot t_4 \end{bmatrix} = \begin{bmatrix} 1 & 0.1 & -0.4 & 0.5 \\ 0.1 & 1 & -0.3 & 0.2 \\ -0.4 & -0.3 & 1 & 0.1 \\ 0.5 & 0.2 & 0.1 & 1 \end{bmatrix}$$

Note: Typo for t_1, t_2, t_3, t_4 in G_t . Consider them as vectors represented as $\underline{t}_1, \underline{t}_2, \underline{t}_3, \underline{t}_4$.

(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 \underline{d}_1 through \underline{d}_4 with respect to \underline{q} , assuming the GVSM model is employed.

In this case, the document & queries are viewed as representations, assuming that the matrix A^T is the basis.

Q2. (40 Points) Answer the questions based on the following table shown in the figure

	\underline{t}_1	\underline{t}_2	\underline{t}_3	\underline{t}_4	Relevance
\underline{d}_1	2	1	1	0	REL
\underline{d}_5	0	1	3	1	REL
\underline{d}_7	2	0	3	1	NREL
\underline{d}_3	0	1	3	0	REL
\underline{d}_6	1	0	2	1	NREL

Fig 1: W matrix for training.

- (a). Use the gradient descent approach based on (standard) Perception Criterion discussed in Class and determine the optimum query.
- (b). Uses the gradient descent approaches based on Generalized Perception Criterion discussed in class and determine the optimum query. We have:
 $\underline{b}_1 = \underline{d}_1 - \underline{d}_7$, $\underline{b}_2 = \underline{d}_1 - \underline{d}_6$, $\underline{b}_3 = \underline{d}_5 - \underline{d}_7$, $\underline{b}_4 = \underline{d}_5 - \underline{d}_6$, $\underline{b}_5 = \underline{d}_3 - \underline{d}_7$, $\underline{b}_6 = \underline{d}_3 - \underline{d}_6$
- (c). Use Generalized Perception—learning by sample discussed in Class and determine the optimum query. We have:
 $\underline{b}_1 = \underline{d}_1 - \underline{d}_7$, $\underline{b}_2 = \underline{d}_1 - \underline{d}_6$, $\underline{b}_3 = \underline{d}_5 - \underline{d}_7$, $\underline{b}_4 = \underline{d}_5 - \underline{d}_6$, $\underline{b}_5 = \underline{d}_3 - \underline{d}_7$, $\underline{b}_6 = \underline{d}_3 - \underline{d}_6$
 You need learn only forwardly.
 For forward learning, you need to follow the order \underline{b}_1 , \underline{b}_2 , \underline{b}_3 , \underline{b}_4 , \underline{b}_5 , \underline{b}_6 ;
- (d). Obtain the optimal query based on Rocchio's method.

Q3. (10 Points) Use the following W' as test data set to compare the above four methods (use the four optimal queries derived by them). Use R_{norm} measure for evaluation.

	\underline{t}_1	\underline{t}_2	\underline{t}_3	\underline{t}_4	Relevance
\underline{d}_2	2	0	2	0	REL
\underline{d}_4	0	2	3	1	REL
\underline{d}_8	0	0	1	1	NREL
\underline{d}_9	0	1	2	0	REL
\underline{d}_{10}	1	1	3	2	NREL

Fig 2: W' matrix for testing