ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ & ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
Εθνικόν και Καποδιστριακόν Πανεπιστήμιον Αθηνών
ΙΔΡΥΘΕΝ ΤΟ 1837

ATHENA Research & Innovation Information Technologies

ΙΕΛ ILSP

M902
Βασικές Μαθηματικές Έννοιες στη Γλωσσική Τεχνολογία

Project 3

Κυλάφη Χριστίνα-Θεανώ

LT1200012

November, 2020

TABLE OF CONTENTS

Question 1 ____ 3

Question 2 ____ 4

Question 3 ____ 5

Question 4 ____ 6

Question 5 ____ 7

Question 6 ____ 8

Question 7 ____ 10

Question 8 ____ 11

Question 9 ____ 12

Question 10 ____ 13

Question **1**

Question **2**

Question **3**

Question **4**

Question **5**

Question **6**

Let $A = \begin{bmatrix} a_1^T \\ a_2^T \\ a_3^T \end{bmatrix} = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 0 & -1 \\ 0 & 1 & 2 \end{bmatrix}$, the unknown vector $x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ and the dot products vector

$b = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 5 \\ 8 \\ 9 \end{bmatrix}$.

This system of linear equations in each condensed form, is written as follows:

$$A \cdot x = b$$

The unknown vector x is calculated by:

$$A \cdot x = b \implies A^{-1} \cdot A \cdot x = A^{-1} \cdot b \implies I \cdot x = A^{-1} \cdot b \implies x = A^{-1} \cdot b \qquad (1)$$

$$A^{-1} = \frac{1}{|A|}\, adj(A) \qquad (2)$$

Determinant of matrix A:

$$det(A) = |A| = \begin{vmatrix} 1 & 2 & 0 \\ 2 & 0 & -1 \\ 0 & 1 & 2 \end{vmatrix} = (-1)^{1+1} \cdot 1 \cdot \begin{vmatrix} 0 & -1 \\ 1 & 2 \end{vmatrix} + (-1)^{1+2} \cdot 2 \cdot \begin{vmatrix} 2 & -1 \\ 0 & 2 \end{vmatrix} + (-1)^{1+3} \cdot 0 \cdot \begin{vmatrix} 2 & 0 \\ 0 & 1 \end{vmatrix} =$$

$$1 \cdot 1 \cdot (0 \cdot 2 - (-1) \cdot 1) - 1 \cdot 2 \cdot (2 \cdot 2) + 0 = +1 - 8 = -7 \qquad (3)$$

Adjoint matrix (transpose of cofactor matrix) of A :

$$adj(A) = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix}^T = \begin{bmatrix} A_{11} & A_{21} & A_{31} \\ A_{12} & A_{22} & A_{32} \\ A_{13} & A_{23} & A_{33} \end{bmatrix},$$

where $A_{ij} = (-1)^{i+j} \cdot \det(M_{ij}),\ i,\ j$ the rows and columns of matrix A respectively and M_{ij} the matrix made by removing row i and column j of matrix A. The proper signs of the cofactors in the **adjoint** (adjugate) matrix, are included in the calculation of A_{ij} .

$$A_{11} = (-1)^{1+1} \cdot \begin{vmatrix} 0 & -1 \\ 1 & 2 \end{vmatrix} = 1, \quad A_{12} = (-1)^{1+2} \cdot \begin{vmatrix} 2 & -1 \\ 0 & 2 \end{vmatrix} = -4, \quad A_{13} = (-1)^{1+3} \cdot \begin{vmatrix} 2 & 0 \\ 0 & 1 \end{vmatrix} = 2$$

$$A_{21} = (-1)^{2+1} \cdot \begin{vmatrix} 2 & 0 \\ 1 & 2 \end{vmatrix} = -4, \quad A_{22} = (-1)^{2+2} \cdot \begin{vmatrix} 1 & 0 \\ 0 & 2 \end{vmatrix} = 2, \quad A_{23} = (-1)^{2+3} \cdot \begin{vmatrix} 1 & 2 \\ 0 & 1 \end{vmatrix} = -1$$

$$A_{31} = (-1)^{3+1} \cdot \begin{vmatrix} 2 & 0 \\ 0 & -1 \end{vmatrix} = -2, \quad A_{32} = (-1)^{3+2} \cdot \begin{vmatrix} 1 & 0 \\ 2 & -1 \end{vmatrix} = 1, \quad A_{33} = (-1)^{3+3} \cdot \begin{vmatrix} 1 & 2 \\ 2 & 0 \end{vmatrix} = -4$$

$$adj(A) = \begin{bmatrix} 1 & -4 & -2 \\ -4 & 2 & 1 \\ 2 & -1 & -4 \end{bmatrix} \qquad (4)$$

$$\overset{(2),\,(3),\,(4)}{\Longrightarrow} \quad A^{-1} = \frac{1}{-7} \cdot \begin{bmatrix} 1 & -4 & -2 \\ -4 & 2 & 1 \\ 2 & -1 & -4 \end{bmatrix} = \begin{bmatrix} -\frac{1}{7} & \frac{4}{7} & \frac{2}{7} \\ \frac{4}{7} & -\frac{2}{7} & -\frac{1}{7} \\ -\frac{2}{7} & \frac{1}{7} & \frac{4}{7} \end{bmatrix} \qquad (5)$$

$$\overset{(1),\,(5)}{\Longrightarrow} \; x = A^{-1} \cdot b \;\Longrightarrow$$

$$x = \begin{bmatrix} -\frac{1}{7} & \frac{4}{7} & \frac{2}{7} \\ \frac{4}{7} & -\frac{2}{7} & -\frac{1}{7} \\ -\frac{2}{7} & \frac{1}{7} & \frac{4}{7} \end{bmatrix} \cdot \begin{bmatrix} 5 \\ 8 \\ 9 \end{bmatrix} = \begin{bmatrix} -\frac{1}{7} \cdot 5 + \frac{4}{7} \cdot 8 + \frac{2}{7} \cdot 9 \\ \frac{4}{7} \cdot 5 - \frac{2}{7} \cdot 8 - \frac{1}{7} \cdot 9 \\ -\frac{2}{7} \cdot 5 + \frac{1}{7} \cdot 8 + \frac{4}{7} \cdot 9 \end{bmatrix} = \begin{bmatrix} \frac{45}{7} \\ -\frac{5}{7} \\ \frac{34}{7} \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \Longrightarrow$$

$$x_1 = \frac{45}{7}, \; x_2 = -\frac{5}{7}, \; x_3 = \frac{34}{7}$$

Question **7**

Question **8**

Question **9**

Question **10**

In order to find the similar documents based on the 3-dimensional vector representations, $d_n \in \mathbb{R}^3,\ n = 1, \ldots, 5$, a 5x5 matrix was constructed, with each row being each of the 3-D vectors d_1, d_2, d_3, d_4, d_5 , normalised as follows:

$$\sum_{i=1}^{5}\sum_{j=1}^{3}\frac{d_{ij}}{\|d_i\|}$$

where d_{ij} the $j-th$ element of the $i-th$ vector and $\|d_i\| = \sqrt{d_{i1}^2 + d_{i2}^2 + d_{i3}^2}$ the norm of the respective vector i:

$$D = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \\ d_5 \end{bmatrix} \implies D_{5x3} = \begin{bmatrix} 8 & 6 & 0 \\ 0 & 6 & 8 \\ 6 & 0 & 8 \\ 2 & 3 & 0 \\ 9 & 6 & 0 \end{bmatrix} \implies D_{normal} = \begin{bmatrix} \frac{8}{10} & \frac{6}{10} & \frac{0}{10} \\ \frac{0}{10} & \frac{6}{10} & \frac{8}{10} \\ \frac{6}{10} & \frac{0}{10} & \frac{8}{10} \\ \frac{2}{3.6055} & \frac{3}{3.6055} & \frac{0}{3.6055} \\ \frac{9}{10.8166} & \frac{6}{10.8166} & \frac{0}{10.8166} \end{bmatrix} = \begin{bmatrix} 0.8 & 0.6 & 0 \\ 0 & 0.6 & 0.8 \\ 0.6 & 0 & 0.8 \\ 0.5547 & 0.8320 & 0 \\ 0.8320 & 0.5547 & 0 \end{bmatrix}$$

Finally, we multiply **normalized** matrix **D** with its **transpose**, to get the inner product of every pair of vectors and perform the comparisons between the document representations:

$$D_{normal} \cdot D_{normal}^T = \begin{bmatrix} 0.8 & 0.6 & 0 \\ 0 & 0.6 & 0.8 \\ 0.6 & 0 & 0.8 \\ 0.5547 & 0.8320 & 0 \\ 0.8320 & 0.5547 & 0 \end{bmatrix} \cdot \begin{bmatrix} 0.8 & 0 & 0.6 & 0.5547 & 0.8320 \\ 0.6 & 0.6 & 0 & 0.8320 & 0.5547 \\ 0 & 0.8 & 0.8 & 0 & 0 \end{bmatrix} =$$

$$\begin{bmatrix} 1 & 0.36 & 0.48 & 0.9429 & 0.9984 \\ \mathbf{0.36} & 1 & 0.64 & 0.4992 & 0.3328 \\ \mathbf{0.48} & \mathbf{0.64} & 1 & 0.3328 & 0.4992 \\ \mathbf{0.9429} & \mathbf{0.4992} & \mathbf{0.3328} & 1 & 0.9230 \\ \mathbf{0.9984} & \mathbf{0.3328} & \mathbf{0.4992} & \mathbf{0.9230} & 1 \end{bmatrix} = D_{normal\_dot\_product}$$

We keep only the **lower triangular matrix**, where each value represents the dot product of the $j-th$ vector (d_j) with the $i-th$ vector (d_j), where $i,\ j$ the rows and the columns of the matrix $D_{dot\_product}$ respectively.

The formula of the dot product of two vectors, using the included angle θ , is:

$$d_i \cdot d_j = \|d_i\| * \|d_j\| * cos\theta \implies cos\theta = \frac{d_i \cdot d_j}{\|d_i\| * \|d_j\|} = cos\theta_{d_i,d_j}$$

The process we followed above, calculates exactly the value of the cosine of the included angle θ, giving information on the relation between those vectors. If the angle's value is close to 0, then the vectors are orthogonal, if it is close to 1 they are

Therefore, the results are the follwing:

Inner Product (of normalized vectors)	Value	Similarity
$cos\theta_{d_1,d_2} = 0.36$		Low
$cos\theta_{d_1,d_3} = 0.48$		
$cos\theta_{d_1,d_4} = 0.9429$	$> 0.94 \implies \theta_{d_1,d_4} < 19^{\circ}$	**High**
$cos\theta_{d_1,d_5} = 0.9984$	$\sim 1 \implies \theta_{d_1,d_5} \sim 0^{\circ}$	**High**
$cos\theta_{d_2,d_3} = 0.64$		
$cos\theta_{d_2,d_4} = 0.4492$		
$cos\theta_{d_2,d_5} = 0.3328$		Low
$cos\theta_{d_3,d_4} = 0.3328$		Low
$cos\theta_{d_3,d_5} = 0.4992$		
$cos\theta_{d_4,d_5} = 0.9230$	$> 0.92 \implies \theta_{d_4,d_5} < 23^{\circ}$	**High**

Table 3.10

The **Documents 1**, **4** and **5** are **very similar**.

Also, for this exercise, some code was developed to validate the results (included in the uploaded .zip file).