Roll No: 17K-3795
Sec. 'Gr-2.
Muhammad Mustafa Mangi.

Question 1: (a)(1)

- No partial ~~Hard~~ matching in boolean model.
- Difficult to perform relevance feedback
- No Ranking Concept.
- Boolean queries are hard to formulate.

a(2):- VSM use linear Algebra for ranked retrieval. Hence, position of word does not matter

- It assumes that each term of query ~~is~~ is independent

- Often ~~retest~~ retrieved false positive result.

a(3) Doc with similar content but different vocab are not associated in VSM. Hence retrived false -ve.
↓
(Synonym).

a(4) : All term in VSM treated as independent, but human languages have a context eg the word use by human are based on same context.

Ques 1b:

$d_1 : \langle (0.34 * 0.14); (0.12 \times 0.38); 0; (0.26 \times 0.24); 0; 0; 0 \rangle$

$d_2 : \langle 0, (0.29 \times 0.38), (0.33 \times 0.51), 0, 0, (0.22 \times 0.32), 0 \rangle$

$d_3 : \langle (0.12 \times 0.14), (0.19 \times 0.38), 0, 0, 0, 0, (0.21 \times 0.15) \rangle$

~~query : < 0, 0, 1, 1, 0,~~

query : $\langle 0, 0, 1, 1, 0, 0, 1 \rangle$ ~~not~~ only if

⇓

~~query : < 0, 0, 0.81, 0.54, 0, 0, 0.45 >~~

$\cos(d_1, q) : (0.26 \times 0.24)(0.54) / (0.09 \times \sqrt{3}) = 0.40$

$\cos(d_2, q) : (0.33 \times 0.51)(0.81) / (0.213 \times \sqrt{3}) = 0.45$

$\cos(d_3, q) : (0.21 \times 0.15)(0.45) / (0.08 \times \sqrt{3}) = 0.22$

Ranking : d_2, d_1, d_3. Ans!

Question No 2.

(a)

Precision and Recall together not a very good evaluation because if a system have high recall but low precision as many extra document. if a system have low recall and high prec.. so it would be miss many document.

(b) Break even point is used to evaluate IR System when precision and recall are equal.

Question 24.

Total doc : 100

1) prec : 5/12. = 0.416

2) recall : 5/8 = 0.625

F1 :

$$(\beta^2 + 1) * P * R / \beta^2 * P + R.$$

$$F1 : \frac{2.P*R}{P+R} = \frac{2(0.416)(0.625)}{0.416 + 0.625}$$

F1 = 0.499

(3) MAP : $\frac{1}{5}\left(\frac{1}{1} + \frac{2}{2} + \frac{3}{5} + \frac{4}{9} + \frac{5}{12}\right)$

= ~~0.632.~~ 0.692

(4) MAP : $\frac{1}{8}\left(\frac{1}{1} + \frac{2}{2} + \frac{3}{5} + \frac{4}{9} + \frac{5}{12.} + \frac{6}{13} \right.$

$\left. \frac{7}{14} + \frac{8}{15}\right)$

MAP = 0.619 (Largest)

(5) MAP : $\frac{1}{8}\left(\frac{1}{1}+\frac{2}{2}+\frac{3}{5}+\frac{4}{9}+\frac{5}{12}+\frac{6}{998}+\frac{7}{999}+\frac{8}{1001}\right)$

MAP = 0.435 (smallest)

Question 3

(i) $d_1 = \langle 0, 0, 0, 0, 0, 1, 1 \rangle$

$d_2 = \langle 0, 1, 0, 0, 1, 0, 1 \rangle$

$d_3 = \langle 1, 0, 0, 1, 0, 0, 0 \rangle$

$d_4 = \langle 0, 0, 1, 1, 0, 0, 1 \rangle$

$d_5 = \langle 0, 0, 0, 1, 0, 0, 2 \rangle$

$\cos(d_1, d_5) = 2/(\sqrt{2} * \sqrt{5}) = 0.632$

$\cos(d_2, d_5) = 2/(\sqrt{3} \times \sqrt{5}) = 0.516$

$\cos(d_3, d_5) = 1/(\sqrt{2} * \sqrt{5}) = 0.316$

$\cos(d_4, d_5) = 3/(\sqrt{3} \times \sqrt{5}) = 0.77$

d_5 belong to Chine.

Q3(ii)

$d_5 = \langle 0, 0, 0, 1, 0, 0, 2 \rangle$

$\mu_{\text{claim=no}} = \frac{1}{2}(\vec{d_3} + \vec{d_4})$

$= \frac{1}{2}\langle 1, 0, 1, 2, 0, 0, 1 \rangle$

$\mu_{\text{claim=no}} = \langle 0.5, 0, 0.5, 1, 0, 0, 0.5 \rangle$

$\mu_{\text{claim=yes}} = \frac{1}{2}(\vec{d_1} + \vec{d_2})$

$= \frac{1}{2}\langle 0, 0, 0, 0, 0, 1, 1 \rangle \langle 0, 1, 0, 0, 1, 0, 1 \rangle$

$= \frac{1}{2}\langle 0, 1, 0, 0, 1, 1, 2 \rangle$

$\mu_{\text{claim=yes}} = \langle 0, 0.5, 0, 0, 0.5, 0.5, 1 \rangle$

$|\mu_{\text{claim=yes}}| = |\langle 0, 0.5, 0, 0, 0.5, 0.5, 1 \rangle - \langle 0, 0, 0, 1, 0, 0, 2 \rangle|$

$= \sqrt{(0.5)^2 + (-1) + (0.5)^2 + (0.5)^2 + (-1)^2}$

$|\mu_{\text{yes}} - d_5| = 1.658$

$$|\mu_{\text{win-No}} - d_5| = |\langle 0.5, 0, 0.5, 1, 0, 0, 0.5\rangle - \langle 0, 0, 0, 1, 0, 0, 2\rangle|$$

$$= \sqrt{(0.5)^2 + (0.5)^2 + (-1.5)^2}$$

$$= \boxed{1.658}$$

Ans!

Chance of d_5 in both class is same.

Q3 (iii) :

Priors Prob:

$P(Yes) = 2/4 = 0.5$

$P(No) = 2/4 = 0.5$

$P(Sapporo|Yes) = \frac{0+1}{5+7} = \frac{1}{12} = 0.0833$

$P(Taiwan|Yes) = \frac{3}{12} = 0.25$

$P(Sapporo|No) = \frac{3}{12} = 0.25$

$P(Taiwan|No) = \frac{2}{5+7} = 0.167$

$P(Yes|d_5) = 0.5 \times 0.25 \times 0.25 \times 0.933$

$= 0.0026$

$P(No|d_5) = 0.5 \times (0.167)^2 \times 0.25$

$= 0.0034.$

Hence D_5 belong to China as per Multi. Naive Bayes.

Question No 4.

Let $C_1 = d_2$ & $C_2 = d_5$

$(C_1, d_1) = \sqrt{1^2+0} = 1$
$(C_1, d_2) = \sqrt{0^2+0^2} = 0$
$(C_1, d_3) = \sqrt{2^2+0^2} = 2$
$(C_1, d_4) = \sqrt{1^2+3^2} = \sqrt{10}$
$(C_1, d_5) = \sqrt{0^2+3^2} = 3$
$(C_1, d_6) = \sqrt{2^2+3^2} = \sqrt{13}$

$(C_2, d_1) = \sqrt{10}$
$(C_2, d_2) = 3$
$(C_2, d_3) = \sqrt{13}$
$(C_2, d_4) = 1$
$(C_2, d_5) = 0$
$(C_2, d_6) = 2.$

$C_1 = \{d_1, d_2, d_3\}$

$C_2 = \{d_4, d_5, d_6\}.$

→ New Centroid

$C_1 = (2.33, 4)$

→ New Centroid

$C_2 (2.33, 1)$

$(C_1, d_1) = \sqrt{(1-2.33)^2+(4-4)^2} = 1.33.$
$(C_1, d_2) = 0.333$
$(C_1, d_3) = 1.67$
$(C_1, d_4) = 3.28$
$(C_1, d_5) = 3.01$
$(C_1, d_6) = 3.43.$

$(C_2, d_1) = \sqrt{(1-2.33)^2+(4-1)^2} = 3.2$
$(C_2, d_2) = 3.018$
$(C_2, d_3) = 3.433$
$(C_2, d_4) = 1.333$
$(C_2, d_5) = 0.333$
$(C_2, d_6) = 1.667$

Final Cluster:

$C_1 = \{d_1, d_2, d_3\}$

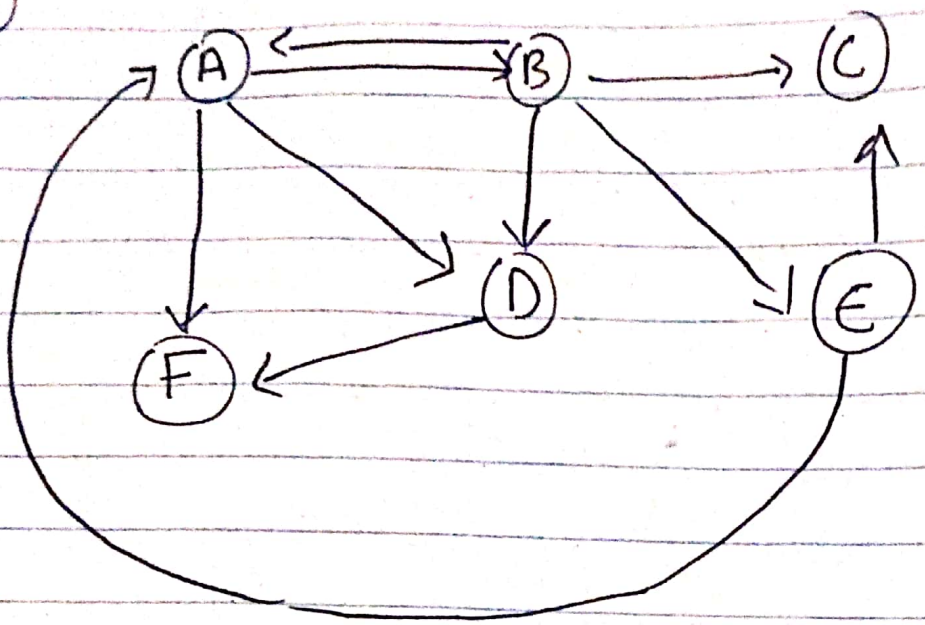
$C_2 = \{d_4, d_5, d_6\}$.

Time Complexity: $O(\bar{I}KNM)$

To find optimal clustering one need to produce all possible clustering arrangement. Hence k-Mean Coverage local minimum.

Question 6

(a)



: Pictorial Representation:

A	0	1	0	1	0	1
B	1	0	1	1	1	0
C	0	0	0	0	0	0
D	0	0	0	0	0	1
E	1	0	1	0	0	0
F	0	0	0	0	0	0

Adj. Matrix.

No, Webpages with no outlinks are those nodes which make the given statement false.

(b) initially: $H = [1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1]$

$A = [1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1]$

$$h^0 = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$h^0 = \begin{bmatrix} 3 \\ 4 \\ 0 \\ 1 \\ 2 \\ 0 \end{bmatrix} \Rightarrow \text{Normalizing.} \quad h^0 = \begin{bmatrix} 0.547 \\ 0.730 \\ 0 \\ 0.182 \\ 0.365 \\ 0 \end{bmatrix}$$

$$h' = \begin{bmatrix} 0.912 \\ 1.094 \\ 0 \\ 1 \\ 0.547 \\ 0 \end{bmatrix} \Rightarrow \text{Norm..} \quad \begin{bmatrix} 0.499 \\ 0.599 \\ 0 \\ 0.548 \\ 0.299 \\ 0 \end{bmatrix}$$

$$a^0 = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$a^0 = \begin{bmatrix} 2 \\ 11 \\ 2 \\ 2 \\ 21 \\ 2 \end{bmatrix} \Rightarrow \text{Norm} \Rightarrow a^0 \begin{bmatrix} 0.471 \\ 0.235 \\ 0.471 \\ 0.471 \\ 0.235 \\ 0.471 \end{bmatrix}$$

$$a' = \begin{bmatrix} 0.47 \\ 0.471 \\ 0.47 \\ 0.706 \\ 0.235 \\ 0.942 \end{bmatrix} \Rightarrow \text{Norm} \Rightarrow \begin{bmatrix} 0.306 \\ 0.307 \\ 0.306 \\ 0.460 \\ 0.153 \\ 0.614 \end{bmatrix}$$

Ans

Question 6 (c)

$d = 0.85$

$$\begin{bmatrix} 0 & 1/4 & 0 & 1 & 0 & 0 \\ 1/3 & 0 & 0 & 1 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0.85 \\ 0.85 \\ 0.85 \\ 0.85 \\ 0.85 \\ 0.85 \end{bmatrix}$$

$$\begin{bmatrix} 1.06 \\ 1.56 \\ 0 \\ 0 \\ 0.283 \\ 0 \end{bmatrix}$$ ← iteration # 1

$\Rightarrow$ iteration #2.

$$\begin{bmatrix} 0.39 \\ 0.49 \\ 0 \\ 0 \\ 0.353 \\ 0 \end{bmatrix}$$