



[25] Question 1. [25].

- 1.1 Symmetric [1]
- 1.2 Determinant = -5 [1]
- 1.3 Inverse = $\begin{bmatrix} -0.4 & 0.6 \\ 0.6 & -0.4 \end{bmatrix}$ (2 pts for correct method) [4]
- 1.4 $\lambda = -1, 5$ [4]
- 1.5 $\begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}, \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$ if normalized +1; if principal identified +1 [4]
 [2]
- 1.6 0 opt: always zero [2]
- 1.7 $A \times v = \begin{bmatrix} 5 \\ 5 \end{bmatrix} = \lambda v$ [3]
- 1.8 showing, either by coefficients or plot. [4]
 ↳ computed, but didn't show → (-2)

[7] Q2 → graph.

- all points correct → 7
- only +ve points plotted → -3
- ~~two~~ points co-ord 0.1 insted of 0.16 → -1

[25] Q3

- 3.1. Similarity $(Q, d_1) = 0.7852$ [2]
 $(Q, d_2) = 0.7688$ [2]

- 3.2. $v_1 = [2.53 \quad 14.56 \quad 4.6 \quad 0 \quad 0 \quad 2.09]$ } from slides [3]
 $v_2 = [3.37 \quad 6.93 \quad 2.55 \quad 0 \quad 1.07 \quad 0]$ } [3]
 (-1 if no zero) $(-1/2$ if one # b wrong) but calc is ok.

- 3.3 $\text{Sim}(Q, d_1) = 0.43$ [3]
 $\text{Sim}(Q, d_2) = 0.512$ [3]

3.4 → Not Required.

- 3.5.1 doubles [2]
- 3.5.2. doesn't change [2]
- 3.5.3 doubles [2]
- 3.5.4. doesn't chge. [3]

zero points if steps are missing entirely.

7
13
7.

X ~~~~~

Word	Ndocs	Ptr
a	1	
bonfire	2	
of	1	
rage	1	
raging	1	
stop	1	
the	1	
vanities	1	

1. Doc 1: "Bonfire of vanities"
2. Doc 2: "a raging bonfire"
3. Doc 3: "stop the rage"

DOCID	OCCUR	POS
2	1	1
1	1	1
2	1	3
1	1	2
3	1	3
2	1	2
3	1	1
3	1	2
1	1	3

~~~~~

## Question 6

**Consider the strings "Kambhampati" and "Kambhapatti".**

**(a) What is the Levenshtein distance between the strings? Give the edit operations that you used in computing this distance.**

1> Levenshtein distance between the strings is 2.

String s1 = **Kambhapatti**

Op1: s2 = insert(s1, 7, m) = **Kambhampatti** (insert m in s1 at 7<sup>th</sup> position)

Op2: s3 = delete(s2, 11, t) = Kambhampati (delete t at 11<sup>th</sup> position from s2)

**(b) What is the Jaccard similarity between these strings based on 3-gram distance (assume that you generate 3-grams by taking three consecutive letters at a time, and shifting by one letter each time. also assume that ^ is the special character for beginning space and \$ is the special character for ending space).**

1> Using 3-grams:

Kambhampati: A = {^ka, kam, amb, mbh, bha, ham, amp, mpa, pat, ati, ti\$}

Kambhapatti: B = {^ka, kam, amb, mbh, bha, hap, apa, pat, att, tti, ti\$}

2> A intersection B = {kam, amb, mbh, bha, pat} = 7

A union B = { kam, amb, mbh, bha, ham, amp, mpa, pat, ati, hap, apa, att, tti} = 15

2> Jaccard's similarity: sim = (A intersection B) / (A union B)

sim = 7/15 = 0.4667

