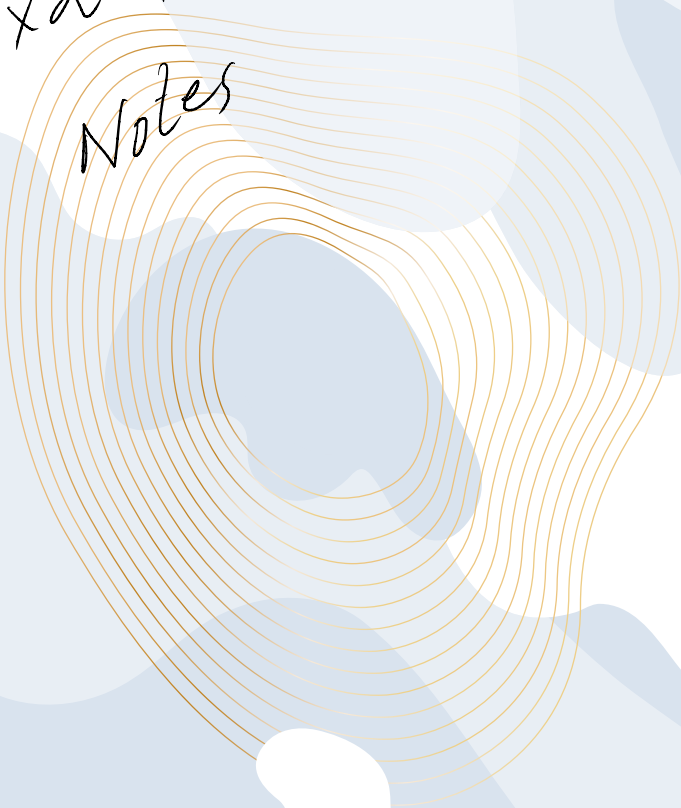
Exam Notes



Laplace smoothing :-

$$\therefore P(w_i / w_{i-1}) = \frac{C(w_{i-1}, w_i) + 1}{C(w_{i-1}) + |V|}$$

Here $|V|$ = number of vocabulary

Minimum edit distance:-

S_1	I	N	T	E	*	N	T	I	O	N
	d	s	s		i	s				
S_2	*	E	X	E	C	U	T	I	O	N

$S_1 = S_2$

- Insertion = 1 3
- Deletion = 1 3
- Substitute = 1 3

Total cost = 5

If cost of sub = 2.

Total cost = 8

- Initialization:-

$D(i, 0) = i$

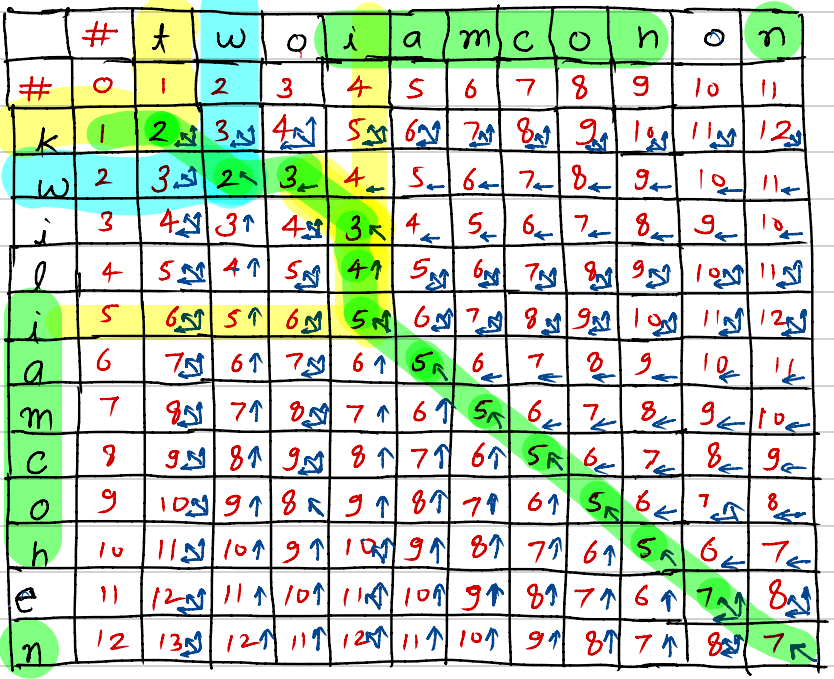
$D(0, j) = j$

go upto m & n.

$$D(i,j) = \min \begin{cases} D(i-1, j) + 1 & \text{delete} \\ D(i, j-1) + 1 & \text{insert} \\ D(i-1, j-1) + 2 ; & \text{if } x(i) \neq y(j) \\ \text{substitute} + 0 ; & \text{if } x(i) = y(j). \end{cases}$$

S_1 = kwiliamcohen

S_2 = twoiamcohon.



Ans 2:- $|V| = 8$

(a) <s> You are Rio </s>
7 1 2 3 8

(b) <s> Rio You are </s>

(c) <s> You are Rio </s>
4 5

(d) <s> You like Yellow orange Rio </s>
6

(1) $P(\text{Rio} \mid \text{are}) = \dfrac{\text{count}(\text{are Rio}) + 1}{\text{count}(\text{are}) + |V|}$

$= \dfrac{2+1}{3+8} = \dfrac{3}{11}$

(2) $P(\text{<s> You like yellow Rio </s>}) = ?$

$= P(\text{You} / \text{<s>}) \cdot P(\text{like} / \text{You}) \cdot P(\text{yellow} / \text{like}) \cdot$
$P(\text{Rio} / \text{yellow}) \cdot P(\text{/s} / \text{Rio})$

$= \dfrac{3+1}{4+8} \times \dfrac{1+1}{4+8} \times \dfrac{1+1}{1+8} \times \dfrac{0+1}{1+8} \times \dfrac{3+1}{4+8}$

$= \dfrac{4}{12} \times \dfrac{2}{12} \times \dfrac{2}{9} \times \dfrac{1}{9} \times \dfrac{4}{12} = \dfrac{2}{4374}$

Ans 4:-

Training

	1	2	3	4	5	6	
1)	Rio	sport	great	India	Proud	dish	P
			7	8			
2)	India	sport	bad	player			P
	9	10		11			
3)	Sam	injustice		poverty	India		N
	12	13		14			
4)	Rio	games	feel	good	sport		P

Test 5) India injustice Sam Rio ?

Assume

P(India injustice Sam Rio) in -ve class.

$P(N / \text{India injustice Sam Rio})$

$\Rightarrow P(N) \cdot P(\text{India injustice Sam Rio} / N)$

$\Rightarrow P(N) \cdot P(\text{India}/N) \cdot P(\text{Injustice}/N) \cdot P(\text{Sam}/N) \cdot P(\text{Rio}/N)$

$\Rightarrow \frac{1}{4} \times \frac{1+1}{4+14} \times \frac{1+1}{18} \times \frac{1+1}{18} \times \frac{0+1}{18}$

$\Rightarrow \frac{1}{4} \times \frac{2}{18} \times \frac{2}{18} \times \frac{2}{18} \times \frac{1}{18 \times 9} \Rightarrow \frac{1}{18 \times 18 \times 18 \times 9} = \frac{1}{52488}$

P (India injustice Sam Rio) in +ve class.

P(P / India injustice Sam Rio)

P(P) P(India injustice Sam Rio/P)

P(P) · P(India/P) · P(injustice/P) · P(Sam/P) · P(Rio/P).

$\Rightarrow \frac{3}{4} \times \frac{2+1}{15+14} \times \frac{0+1}{29} \times \frac{0+1}{29} \times \frac{2+1}{29}$

$\Rightarrow \frac{3}{4} \times \frac{3}{29} \times \frac{1}{29} \times \frac{1}{29} \times \frac{3}{29} \Rightarrow \frac{27}{2821129}$ — (2)

P(N/Test) > P(P/Test)

So — It is in -ve class

Ans 5:-

$$X = [2, 1, 0, 2, 3.12, 1.2]$$

$$W = [3.5, -6, 1.2, 0.4, 3.5, 0.3]$$

$$b = 0.20$$

Cross Entropy loss:- for +ve class:-

$P(+/x) \Rightarrow P(y=1/x) = \sigma(wx+b)$.

$$\therefore \sum wx + b$$
$$= [2 \times 3.5 - 6 + 0 + 0.8 + 3.12 \times 3.5 + 0.36 + 0.20]$$
$$= 13.28$$

$\rightarrow \sigma(13.28)$

$$\frac{1}{1+e^{-13.28}} = 0.9999.$$

sigmoid function.

$$\sigma(x) = \frac{1}{1+e^{-x}}$$

$$L_{CE} = -[y \log \sigma(wx+b) + (1-y) \log(1-\sigma(wx+b))]$$
$$= -[1 \cdot \log 0.9999\ldots + 0]$$
$$= -\log_e 0.9999\ldots \simeq 0$$

(b) for -ve class :- (y = 0).

→ $L_{CE} = (0 + 1 \times \log [1 - 0.9999])$

$= -\log_e 0.0000001$?

$= 21.6$ ✗

⇒

PPMI Matrix

⑦

Term	Auto	Comp	Money	How	Politics	Σ
1) Car	10	8	1	0	3	22
2) Auto	5	1	0	0	1	7
3) Insu	1	0	4	3	0	8
4) Window	1	2	1	2	1	7
5) Computer	3	2	1	0	0	6
6) Tech.	1	3	1	1	1	7
Σ	21	16	8	6	6	(57)

N = 57

P(context)	21/57	16/57	8/57	6/57	6/57
P(w)					
~~57~~					
22/57	10/57	8/57	1/57	0	3/57
7/57	5/57	1/57	0	0	1/52
8/57	1/57	0	4/57	3/57	0
7/57	1/57	2/57	1/57	2/57	1/57
6/57	3/57	2/57	1/57	0	0
7/57	1/57	3/57	1/57	1/57	1/57

Probability of term to total term.

$$\max\left(\log_2\left(\frac{\frac{10}{57}}{\frac{21}{57}\times\frac{22}{57}}\right), 0\right)$$

$$\max\left(\frac{u_{ij}}{u_i * u_j}\right)$$

NLP Midterm Solution

Magic
DATE 12.02.23
PAGE

Ans(1.) a.)

^[bc]*(a[bc]*){0,3}$

b.)

^q+(pq+)+$

c)

^[0-9].+[a-z A-Z]+$

Ans(3.)

1	2	3	4	5	
<s>	you	are	Rio	</s>	
<s>	Rio	you	are	</s>	
<s>	you	like	yellow	orange	Rio </s>
<s>	you	are	Rio	</s>	

$$P(Rio/are) = \frac{count(are\ rio) + 1}{count(are) + |V|}$$

$$= \frac{2+1}{3+8} = \frac{3}{11}$$

P(<s> you like yellow rio </s>)

$$= P(you/<s>) \cdot P(like/you) \cdot P(yellow/like) \cdot P(rio/yellow) \cdot P(</s>/rio)$$

$$= \frac{4}{12} \times \frac{2}{12} \times \frac{2}{9} \times \frac{1}{9} \times \frac{4}{12}$$

$$= \frac{1}{3} \times \frac{1}{6} \times \frac{2}{9} \times \frac{1}{9} \times \frac{1}{3}$$

$$= 1/54 \times 2/81 = 2/4374$$

Magic
DATE 14/02/23
PAGE

Ans.2)

	District	House	Income	Previous cust.	class
1.	S	D	H	N	N
2.	S	D	H	Y	N
3.	M	D	H	N	R
4.	U	SD	H	N	R
5.	U	SD	L	N	R
6.	U	SD	L	Y	N
7.	M	SD	L	Y	R
8.	S	T	H	N	N
9.	S	SD	L	N	R
10.	U	T	L	N	R
11.	S	T	L	Y	R
12.	M	T	H	Y	R
13.	M	D	L	N	R
14.	U	T	H	Y	R

Entropy: (S)

For class → $-\frac{10}{14}\log\frac{10}{14} - \frac{4}{14}\log\frac{4}{14}$

→ $-\left(\frac{5}{7}\log\frac{5}{7} + \frac{2}{7}\log\frac{2}{7}\right)$

→ 0.8632

Information gain:

Attribute ① dist.

Gain(S, Dist.)

→ $0.8632 - \left(\frac{5}{14} \times \text{Entropy}(S) + \frac{4}{14} \times \text{Entropy}(U) + \frac{5}{14} \times \text{Entropy}(M)\right)$

Entropy(S) → 3N 2R class

→ $-(3/5 \log 3/5 + 2/5 \log 2/5)$

→ $-(0.6 \log 0.6 + 0.4 \log 0.4)$

→ 0.9710

Entropy(U) → 4R 1N

→ $-(4/5 \log 4/5 + 1/5 \log 1/5)$

→ $-(0.80 \log 0.80 + 0.20 \log 0.20)$

→ ~~0.043~~ 0.722

Entropy(4) → 4R 0N

→ 0

IG(S, dist) → $0.8632 - (5/14 \times 0.9710 + 5/14 \times 0.722)$

→ $0.8632 - 0.6046$

→ 0.2586

Attribute2 HouseType

IG(S, House) ↗

$$0.8632 - \left(\frac{4}{14} \times \text{Entropy}(D) + \frac{5}{14} \cdot \text{Entropy}(SD) + \frac{5}{14} \cdot \text{Entropy}(T) \right)$$

↗

Entropy(D) ↗ 2N 2R

↗ $-(2/4 \log 2/4 + 2/4 \log 2/4)$

↗ $-\log(1/2)$

↗ (1)

Entropy(SD) ↗ 4R 1N

↗ $-(4/5 \log 4/5 + 1/5 \log 1/5)$

↗ $-(0.8 \log 0.8 + 0.2 \log 0.2)$

↗ 0.722

Entropy(T) ↗ 4R 1N

↗ 0.722

↗ $0.8632 - \left(\frac{4}{14}(1) + \frac{5}{14}(0.722) + \frac{5}{14}(0.722) \right)$

↗ $0.8632 - 0.8014$

↗ 0.0618

Attributes Income

IG(S, income)

→ 0.8632 − (7/14 × Entropy(H) + 7/14 × Entropy(L))

→ Entropy(H) → 3N 4R

→ −(3/7 log 3/7 + 4/7 log 4/7)

→ 0.9853

Entropy(L) : 1N 6R

→ (−1/7 log 1/7 − 6/7 log 6/7)

→ 0.5917

∴ IG(S, Income) →

0.8632 − (1/2 × (0.9853) + 1/2 × (0.5917))

→ 0.8632 − 0.7885

→ 0.0747

Attribute4 (Prev. Cust)

IG(S, Prev. Cust.) ?

$0.8632 - \left(\frac{6}{14} \times \text{Entropy}(Y) + \frac{8}{14} \times \text{Entropy}(N)\right)$

Entropy(Y): 2N 4R

? $-\left(\frac{2}{6}\log\frac{2}{6} + \frac{4}{6}\log\frac{4}{6}\right)$

? 0.9183

Entropy(N): 2N 6R

$-\left(\frac{2}{8}\log\frac{2}{8} + \frac{6}{8}\log\frac{6}{8}\right)$

? $-\left(\frac{1}{4}\log\frac{1}{4} + \frac{3}{4}\log\frac{3}{4}\right)$

? 0.8113

IG(S, Prev. Cust.) ?

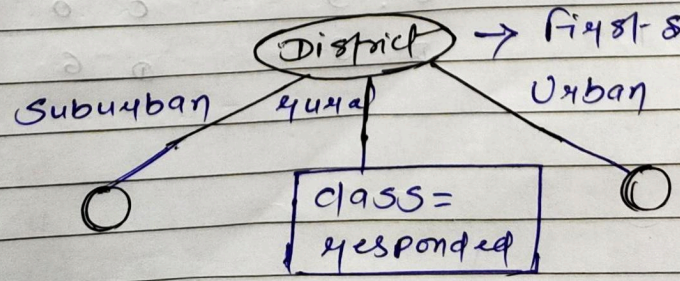
$0.8632 - \left(\frac{6}{14} \times 0.9183 + \frac{8}{14} \times 0.8113\right)$

? $0.8632 - 0.8571$

? 0.0061

Information gain from District attribute is Max.

So Dist. will be selected as first splitting Node



Ans 4.)

n	13	12	11	10	9	9	9	8	7	6	5	4	3	3	2
e	12	11	10	9	8	8	8	7	6	5	4	3	2	2	3
h	11	10	9	8	7	7	7	6	5	4	3	2	1	2	3
o	10	9	8	7	6	6	6	5	4	3	2	1	2	3	4
c	9	8	7	6	5	5	5	4	3	2	1	2	3	4	5
Space) _	8	7	6	5	4	4	4	3	2	1	2	3	4	5	6
m	7	6	5	4	3	3	3	2	1	2	3	4	5	6	7
a	6	5	4	3	2	2	2	1	2	3	4	5	6	7	8
i	5	4	3	2	1	1	1	2	3	4	5	6	7	8	9
l	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10
l	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11
i	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12
W	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13
#	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
	#	W	i	l	l	l	i	a	m	_	c	o	h	o	n

Magic
DATE 12.02.23
PAGE

Ans.5) x y

	Exp.	Sal.	x^2	xy
	3	20	9	60
	8	40	64	320
	9	50	81	450
	13	60	169	780
	3	20	9	60
	6	30	36	180
	11	40	121	440
	21	80	441	1680
	1	10	1	10
	16	70	256	1120
Σ	91	420	1187	5100

$n = 10$

$b = \frac{n\Sigma xy - \Sigma x \cdot \Sigma y}{n\Sigma x^2 - (\Sigma x)^2}$

$\frac{10 \cdot (5100) - 91 \cdot 420}{10 \cdot (1187) - (91)^2}$

$\frac{12780}{3589} \approx \frac{3.56}{}$

$a = \frac{\Sigma y \Sigma x^2 - \Sigma x \Sigma xy}{n(\Sigma x^2) - (\Sigma x)^2}$

$\frac{420 \times 1187 - 91 \times 5100}{10(1187) - (91)^2}$

$\frac{34440}{3589} \approx \frac{9.595}{}$

linear reg. eqn

$y = a + bx$

For 17th of exp.

$9.595 + 3.56(17)$

≈ 70.115

$\simeq \boxed{70}$

Ans. 6)

Train:

	Class
Rio sport great india proud	P
India sport bad player	N
Sam injustice poverty india	N
Rio games feel good sport	P

Test:

India injustice sam Rio class = ?

Vocab. $|V| = 13$

P(India injustice sam rio) in -ve class:

$P(N/\text{India injustice sam rio}) \propto$

$\propto P(N) \cdot P(\text{India injustice sam rio}/N)$

$\propto P(N) \cdot P(\text{India}/N) \cdot P(\text{injustice}/N) \cdot P(\text{sam}/N) \cdot P(\text{rio}/N)$

$\propto \frac{1}{4} \times \frac{2}{17} \times \frac{2}{17} \times \frac{2}{17} \times \frac{1}{17} \propto \frac{4 \times 2}{4 \times 17^4} \propto \frac{2}{17^4} \propto \frac{2}{4913 \times 17}$

$\propto 2.035 \times 10^{-4}$

2.394×10^{-5}

$P(\text{India}/N) \Rightarrow \frac{\text{count(India in -ve)} + 1}{\text{count(-ve)} + |V|}$

$\propto \frac{2}{4+13} \propto \frac{2}{17}$

$P(\text{injustice}/N) \propto \frac{2}{17}$

$P(\text{sam}/N) \propto \frac{2}{17}$ $P(\text{rio}/N) \propto \frac{1}{17}$

count(-ve) = (4)

P(India injustice sam 410) in +ve class

$$P(P/\text{India injustice sam 410})$$

$$\propto P(P) \cdot P(\text{India Injustice sam 410}/P)$$

$$\propto \frac{3}{4} \times P(\text{India}/P) \cdot P(\text{injustice}/P) \cdot P(\text{sam}/P) \cdot P(\text{410}/P)$$

$$\propto \frac{3}{4} \cdot \left(\frac{2+1}{14+13}\right) \cdot \left(\frac{0+1}{14+13}\right) \cdot \left(\frac{0+1}{14+13}\right) \left(\frac{2+1}{27}\right)$$

(count+(+ve) = (14))

$$\propto \frac{3}{4} \cdot \frac{3}{27} \cdot \frac{1}{27} \cdot \frac{1}{27} \cdot \frac{3}{27}$$

$$\propto \frac{1}{(27)^3 \times 4} \propto \underline{1.270 \times 10^{-5}}$$

So we can clearly see

$2.394 \times 10^{-5} > 1.270 \times 10^{-5}$

$P(N/\text{Test}) > P(P/\text{Test})$

So

given test sentence is in Negative class.

Ans:7)

$x = [4, 1, 0, 2, 3.12, 4.2]$

$w = [3.5 \quad -6 \quad -1.2 \quad 0.4 \quad 3.5 \quad 0.3]$

$b = 0.15$

Cross entropy loss:

a) For +ve class ($y=1$)

$$P(+|x) \Rightarrow P(y=1/x) = \sigma(wx+b)$$

$$\sum wx + b \Rightarrow (14 - 6 + 0 + 0.8 + 10.92 + 1.26) + 0.15$$

$$\Rightarrow 8.8 + 11.07 + 1.26$$

$$\Rightarrow \sigma(21.13)$$

$$\Rightarrow \frac{1}{1+e^{-21.13}} \Rightarrow 0.99999..3341$$

LCE ?

$$-[y \log \sigma(wx+b) + (1-y) \log(1-\sigma(wx+b))]$$

$$\Rightarrow -\log(0.999..)$$

$$\Rightarrow \simeq 0$$

b) For -ve class ($y=0$)

$$P(-/x) \Rightarrow 1 - \sigma(wx+b)$$

$$\Rightarrow 0.00000000\cancel{0}067$$

$$LCE \Rightarrow -\log(0.00000000\cancel{0}067)$$

$$\Rightarrow 21.12$$

Magic
DATE 12.02.23
PAGE

Ans.8.)

$x_1 = 4$ (+ve lexicon)

$x_2 = 3$ (-ve lexicon)

$w_1 = 1.2$

By default $w_2 = -3.1$ (Missed -ve sign!! ☹)

$(x_0 = 1)$ $b = 1.5$ learning-rate $= 0.1$

$y = 1$

$$\theta_{t+1} = \theta_t - \eta \Delta L(F(x,\theta), y)$$

$$\nabla_{w,b} = \begin{bmatrix} \frac{\partial L_{CE}(\hat{y}, y)}{\partial w_1} \\ \frac{\partial L_{CE}(\hat{y}, y)}{\partial w_2} \\ \frac{\partial L_{CE}(\hat{y}, y)}{\partial b} \end{bmatrix} \Rightarrow \begin{bmatrix} (\sigma(wx+b) - y)\, x_1 \\ (\sigma(wx+b) - y)\, x_2 \\ (\sigma(wx+b) - y)\, x_0 \end{bmatrix}$$

$x_0 \Rightarrow 1$

$\Rightarrow wx + b \Rightarrow w_1 x_1 + w_2 x_2 + b$

$\Rightarrow 4.8 + 9.3 + 1.5$

$\Rightarrow +15.6$ (-3)

$\sigma(+15.6) \Rightarrow 0.999999832$ $\quad -3 \Rightarrow 0.047425$

$$\theta_1^{(1)} = \begin{bmatrix} 1.2 \\ 3.1 \\ 1.5 \end{bmatrix} + (-)\, 0.1 \times \begin{bmatrix} -0.00000068 \\ -0.00000051 \\ -0.00000017 \end{bmatrix} \begin{matrix} -3.8103 \\ -2.0577 \\ -0.9525 \end{matrix}$$

$$\Rightarrow \begin{bmatrix} 1.200000068 \\ 3.100000051 \\ 1.500000017 \end{bmatrix} \Rightarrow \begin{bmatrix} 1.58103 \\ 3.38577 \\ 1.59525 \end{bmatrix}$$

Magic
DATE 16-02-23
PAGE

Ans ③

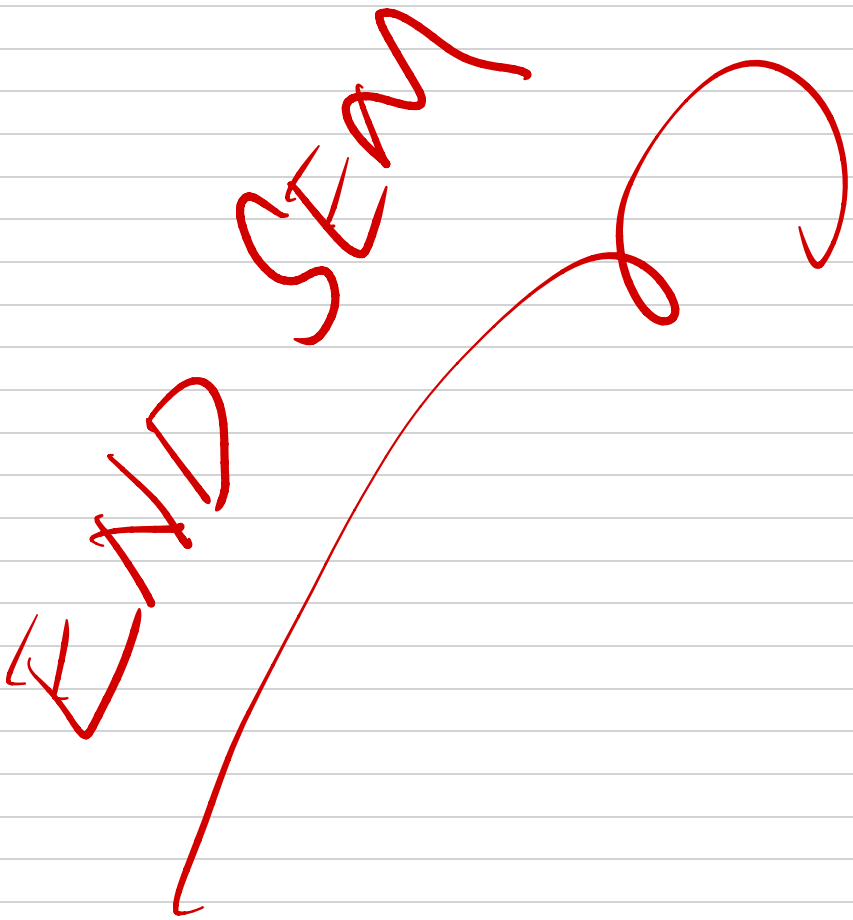
	Content / Term	Auto.	comp.	Money	House	Politics	Σ
1	Car	2	0	1	0	3	6
2	Auto	3	1	0	0	1	5
3	Insu.	1	0	4	3	0	8
4	Win.	1	2	1	2	1	7
5	Comp.	0	2	1	0	0	3
6	Tech.	1	3	1	1	1	7
	Σ	8	8	8	6	6	36

N = 36

P(content)	8/36	8/36	8/36	6/36	6/36
P(W)					
6/36	2/36	0	1/36	0	3/36
8/36	3/36	1/36	0	0	1/36
8/36	1/36	0	4/36	3/36	0
7/36	1/36	2/36	1/36	2/36	1/36
3/36	0	2/36	1/36	0	0
7/36	1/36	3/36	1/36	1/36	1/36

PPMI

	Auto	comp	Money	Housem.	Politics
Car	0.585	0	0	0	1.585
Auto	1.433	0	0	0	0.263
Insu	0	0	1.17	1.17	0
Win	0	0.362	0	0.7776	0
comp.	0	1.585	0.585	0	0
Tech	0	0.9475	0	0	0


END SEAN

Page No.:
Date: / /

NLP

Syllabus

1. Word Embedding [PPMI, tf-idf, word to vector]
2. NER / Information / Relation Extraction.
3. Parsing (Basics, CKY parsing, CNF)
4. Coreference Resolution.
5. Question Answering.
6. Word semantics
7. Text summarization.
8. Machine Translation.

80% from these and 20% from before.

① @ CKY Algo:-

0 I 1 eat 2 sushi 3 with 4 chopsticks 5 with 6 you 7

	1	2	3	4	5	6	7
0	NP	·	S	·	S	·	S
1		Verb	VP	·	VP	·	VP
2			NP	·	NP	·	NP
3				Prep	PP	·	PP
4					NP	·	NP
5						Prep	PP
6							NP

①(b) → CNF:-

$A \rightarrow \epsilon$

$A \rightarrow AA$

$A \rightarrow a$

Given;

$S' \rightarrow S$

$S \rightarrow ASB$

$A \rightarrow aASA \mid a \mid \epsilon$

$B \rightarrow SbS \mid A \mid bb$

① Remove $\in$ productions :-

$S' \rightarrow S$

$S \rightarrow ASB \mid SB$

$A \rightarrow aASA \mid a \mid aS \mid aSA \mid aAS$

$B \rightarrow SbS \mid A \mid bb \mid \in$

$S' \rightarrow S$

$S \rightarrow ASB \mid SB \mid AS \mid S$

$A \rightarrow aASA \mid a \mid aS \mid aSA \mid aAS$

$B \rightarrow SbS \mid A \mid bb$

② Remove unit productions

$S' \rightarrow ASB \mid SB \mid AS$

$S \rightarrow ASB \mid SB \mid AS$

$A \rightarrow aASA \mid a \mid aS \mid aSA \mid aAS$

$B \rightarrow SbS \mid bb \mid aASA \mid a \mid aS \mid aSA \mid aAS$

③ Remove useless productions:

$S' \rightarrow XB \mid SB \mid AS$

$S \rightarrow XB \mid SB \mid AS$

$A \rightarrow ZA \mid a \mid XS \mid UA \mid YX$

$B \rightarrow MS \mid TT \mid ZA \mid a \mid YS \mid UA \mid YX$

$X \rightarrow AS$

$Y \rightarrow a$ ✓

$Z \rightarrow YX$

$U \rightarrow YS$

$T \rightarrow b$ ✓

$M \rightarrow ST$

Sol3 (a) :- Mean Reciprocal Rank

Q :	1	2	3	4	5	6	7

(2) Remove unit products:

$S' \rightarrow ASB \mid SB \mid AS$

$S \rightarrow ASB \mid SB \mid AS$

$A \rightarrow aASA \mid a \mid aS \mid aAS \mid aSA$

$B \rightarrow SbS \mid bb \mid aASA \mid a \mid aS \mid aAS \mid aSA$

(3) Remove useless productions.

$S' \rightarrow XB \mid SB \mid AS$

$S \rightarrow XB \mid SB \mid AS$

$A \rightarrow ZA \mid a \mid YS \mid YX \mid UA$

$B \rightarrow MS \mid TT \mid ZA \mid a \mid XS \mid YX \mid UA$

$X \rightarrow AS$

$Y \rightarrow a$

$Z \rightarrow YX$

$U \rightarrow YS$

$T \rightarrow b$

$M \rightarrow ST$

Sol 3(a) :- Mean Reciprocal Rank Method:

Q No:-	1	2	3	4	5	6	7
Correct Ans No :	4	3	6	2	8	1	2

$$\text{Mean Reciprocal Rank} = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{\text{rank}_i}$$

$$\text{MRR} = \frac{1}{7}\left[\frac{1}{4} + \frac{1}{3} + \frac{1}{6} + \frac{1}{2} + \frac{1}{8} + \frac{1}{1} + \frac{1}{2}\right] \Rightarrow \frac{1}{7}\left[\frac{69}{24}\right] = \frac{23}{56}$$

$\Rightarrow$ Ans

Sol 3(b) :- Rouge-2 score:

Human	System
Water Spinach ✓	Water spinach ✓
Spinach is ✓	Spinach is ✓
is a ✓	is a ✓
a commonly	a leaf
commonly eaten ✓	leaf vegetable ✓
eaten leaf	vegetable commonly
leaf vegetable ✓	commonly eaten ✓
vegetable of	eaten in
of Asia ✓	in tropical
	tropical areas
	areas of
	of Asia ✓

(Human Bigrams) = 9

(System Bigrams) = 12

(common bigrams) = 6

Recall = $\frac{6}{9}$ Precision = $\frac{6}{12}$

$$f_1 \text{ score} = \frac{2PR}{R+P} = \frac{2 \times \frac{6}{12} \times \frac{6}{9}}{\frac{6}{9} + \frac{6}{12}}$$

Rouge 1 :- Pre Re f_1

$\frac{10}{13}$ $\frac{10}{10}$

$$= \frac{2 \times \frac{10}{13} \times 1}{1 + \frac{10}{13}} = \frac{.4}{7}$$

$$\Rightarrow \frac{20}{}$$

Sol 4:- Cosim Similarity.

	Doc1	Doc2	Doc 3	Doc4	Q
new	1	0	0	0	0
home	1	1	1	1	1
sales	1	1	1	1	1
top	1	0	0	0	0
forcasts	1	0	0	0	0
rise	0	1	0	1	0
in	0	1	2	0	0
july	0	1	1	1	0
increase	0	0	1	0	0
is	0	0	0	0	1
very	0	0	0	0	1
bad	0	0	0	0	1

Similarity Matrix

	D_1	D_2	D_3	D_4	Q
D_1	1	0.4	0.316	0.6	0.4
D_2		1	0.79	0.8	0.4
D_3			1	0.474	0.316
D_4				1	0.4
Q					1

Using Cosine Similarity = $\frac{D_1 \cdot D_2}{|D_1|\,|D_2|}$

⇒ formula, fill similarity matrix.

$\lambda = 0.3$

MMR = arg max [λ $Sim_1(D_i, Q) - (1-\lambda)\, Sim_2(D_i, D_j)$]

First iteration:-

Let S is empty set.

∵ max pairwise similarity within S = 0

So,

MMR = arg max (Sim (d_i, q)).

∵ d_1 has max similarity with d_2, d_4, pick any one

S = {d_1}.

Second Iteration:-

find max distance of an element in S to given d_i,

sim (d_1, d_i)

for d_2,

sim (d_1, d_2) = 0.4

sim (d_2, q) = 0.4

MMR = 0.4 x 0.3 − 0.7 x 0.4

= 0.12 − 0.28 = −0.16

Similarly for d_3, d_4,

for d_3,

$$Sim(d_1, d_3) = 0.316$$
$$Sim(d_3, q) = 0.316$$

$$MMR = 0.3 \times 0.316 - 0.7 \times 0.316$$
$$= -0.1264$$

for d_4,

$$Sim(d_1, d_4) = 0.6$$
$$Sim(d_4, q) = 0.4$$

$$MMR = 0.3 \times 0.4 - 0.7 \times 0.6$$
$$= -0.3$$

Here, for d_2, max MMR is there,

So, $S = \{d_1, d_2\}$.

Third Iteration :-

find $\max \begin{cases} sim(d_i, d_1) & \text{for 1st part} \\ sim(d_i, d_2) & \text{for 2nd part.} \end{cases}$

$$\max\left(\underbrace{Sim(d_1, d_3)}_{0.316}, \underbrace{sim(d_2, d_3)}_{0.79}\right) = 0.79$$

$$sim(d_3, q) = 0.316$$

$$MMR = 0.3 \times 0.316 - 0.7 \times 0.79$$
$$\Rightarrow -0.4582.$$

for d_4,

$$\max \left[\underset{0.6}{sim(d_1, d_4)} ; \underset{0.8}{sim(d_2, d_4)} \right) = 0.8$$

$$Sim\ (d_4, q) = 0.4$$

$$MMR = 0.3 \times 0.4 - 0.7 \times 0.8$$

$$= -0.44.$$

MMR for d_4 is max &.,

$$S = \{ d_1, d_2, d_4 \}.$$

These are 3 sentences in summary set.

Sol6:- Hidden Markov Model → HMM.

·Initial Probability :-

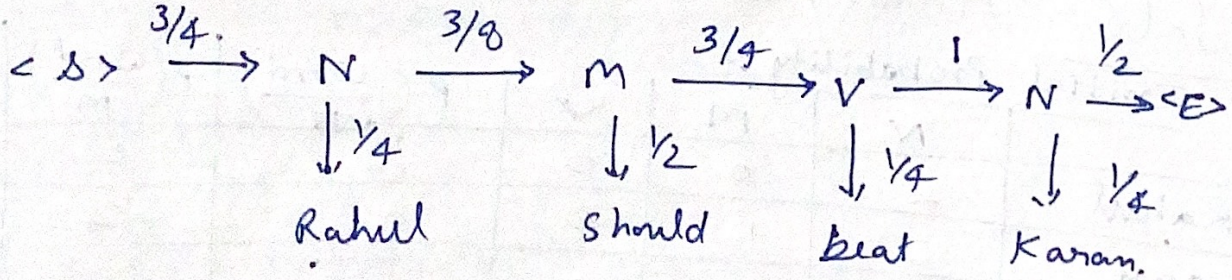
	N	M	V	P(word/N)	P(word/M)	P(word/V)
Rahul	2	0	0	1/4	0	0
can	0	1	0	0	1/4	0
play	0	0	2	0	0	1/2
chess	2	0	0	1/4	0	0
Karan	2	0	0	1/4	0	0
Should	0	2	0	0	1/2	0
Clean	0	0	1	0	0	1/4
table	1	0	0	1/8	0	0
will	0	1	0	0	1/4	0
shayam	1	0	0	1/8	0	0
beat	0	0	1	0	0	1/4
	8	4	4			

Using HMM,

P(N M V N / Rahul should beat Karan).

	N	M	V	<E>
<S>	3/4	1/4	0	0
N	0	3/8	1/8	4/8
M	1/4	0	3/4	0
V	4/4	0	0	0

"Rahul should beat Karan"



$$P = \frac{3}{4} \times \frac{1}{4} \times \frac{3}{8} \times \frac{1}{2} \times \frac{3}{4} \times \frac{1}{4} \times 1 \times \frac{1}{4} \times \frac{1}{4} \times \frac{1}{2}$$

$$= \frac{27}{2^{15}} = 0.000823357.$$

Sol 5(b) :-

For, Sim_{path} :

$Path_{len}$ (Hill, Shore)

$\Rightarrow 1 + 3 \Rightarrow 4$

$Sim_{path} = \frac{1}{4} = 0.25$

for Sim_{jensik} :

$-\log P(LCS(Hill, shore))$

$-\log_2 P(geo\text{-}form)$

$-\log_2(0.00176) = 9.15$

for Sim_{Lin} :

$$\frac{2 \log P(LCS(Hill, shore))}{\log P(Hill) + \log P(shore)}$$

$$\Rightarrow \frac{2 \log P(geoform)}{\log P(Hill) + \log P(shore)}$$

$$\Rightarrow \frac{2 \ln(0.00176)}{\ln(0.0000193) + \ln(0.0000836)}$$

$\Rightarrow 0.6259.$

Sol 5 (a) :-

Transition Probability :-

	jouer	croquet	la	grillon.
play	1/3	1/3	1/3	1/3
cricket	1/3	1/3	1/3	1/3
the	1/3	1/3	1/3	1/3
team	1/3	1/3	1/3	1/3

Alignment Probability :-

Play | jouer : $\frac{1}{3} \times \frac{1}{3} = \frac{1}{9}$

cricket | croquet

the | la : $\frac{1}{3} \times \frac{1}{3} = \frac{1}{9}$

cricket | grillon

cricket | croquet : $\frac{1}{3} \times \frac{1}{3} = \frac{1}{9}$

team | equipe

play — jouer : $\frac{1}{3} \times \frac{1}{3} = \frac{1}{9}$

cricket — croquet

the — la : $\frac{1}{3} \times \frac{1}{3} = \frac{1}{9}$

cricket — grillon

cricket — croquet : $\frac{1}{3} \times \frac{1}{3} = \frac{1}{9}$

team — equipe

Normalization = $\frac{\frac{1}{9}}{\frac{1}{9} + \frac{1}{9}} = \frac{1}{2}$

Weighted translation count →

	jouer	croquet	la	grille	equipe
play	1/2	1/2	0	0	0
the	0	0	1/2	1/2	0
Cricket	1/2 × 1/3	1/3 (1/2 + 1/2)	1/2 × 1/3	1/2 × 1/3	1/2 × 1/3
team	0	1/2	0	0	1/2

Normalization

~~Alignment~~

~~Normalized Too~~ :-

Alignment Probability :-

$\frac{1}{2} \times \frac{1}{3} = \frac{1}{6}$

$\frac{1}{2} \times \frac{1}{6} = \frac{1}{12}$

$\frac{1}{2} \times \frac{1}{6} \times \frac{1}{12}$

$\frac{1}{2} \times \frac{1}{6} \times \frac{1}{12}$

$\frac{1}{3} \times \frac{1}{2} = \frac{1}{6}$

$\frac{1}{8} \times \frac{1}{2} = \frac{1}{12}$

$\frac{1}{6} + \frac{1}{12} = \frac{3}{6} = \frac{1}{2}$

$\frac{1}{6} + \frac{1}{12} = \frac{3}{12} = \frac{1}{4}$

2a)

Dan Jurafsky

LESK'S ALGORITHM

Sense Bag: *contains the words in the definition of a candidate sense of the ambiguous word.*

Context Bag: *contains the words in the definition of each sense of each context word.*

E.g. "On burning ***coal*** we get ***ash***."

Ash

- Sense 1
  Trees of the olive family with pinnate leaves, thin furrowed bark and gray branches.
- Sense 2
  The ***solid*** residue left when ***combustible*** material is thoroughly ***burn***ed or oxidized.
- Sense 3
  To convert into ash

Coal

- Sense 1
  A piece of glowing carbon or ***burn***t wood.
- Sense 2
  charcoal.
- Sense 3
  A black ***solid combustible*** substance formed by the partial decomposition of vegetable matter without free access to air and under the influence of moisture and often increased pressure and temperature that is widely used as a fuel for ***burn***ing

In this case Sense 2 of ash would be the winner sense.

Sol:- PPMI

context / Term	Auto	comp	Money	House	Polling	Σ
1 car	10	8	1	0	3	22
2. Auto	5	1	0	0	1	7
3. Insu	1	0	4	3	0	8
4. window	1	2	1	2	1	7
5. Comp	3	2	1	0	0	6
6. Tech	1	3	1	1	1	7
Σ	21	16	8	6	6	57

N = 57

P context / P Term	21/57	16/57	8/57	6/57	6/57
22/57	10/57	8/57	1/57	0	3/57
7/57	5/57	1/57	0	0	1/57
8/57	1/57	0	1/57	3/57	0
7/57	1/57	2/57	1/57	2/57	1/57
6/57	3/57	2/57	1/57	0	0
7/57	1/57	3/57	1/57	1/57	1/57

$$PPMI(w_1, w_2) = \max\left(\log_2 \frac{p(w_1, w_2)}{p(w_1) \cdot p(w_2)}, 0\right)$$

$$PPMI(\overline{car}, auto) = \max\left(\log_2 \frac{\frac{2}{57}}{\frac{22}{57} * \frac{21}{57}}, 0\right)$$

$$PPMI(w, c) = \max\left(\log_2 \frac{p(w,c)}{p(w)\,p(c)}, 0\right)$$

PPMI :-

	Auto	Comp	Money	Gous	Relate
car					
auto					
insu					
wire					
com					
Fee					