Assignment => 2

Gitesh Shastri
2K15/IT/028

NLP

(Q) For your training corpus of Previous assignment.

1.) Create a BOW feature matrix

2.) Probability of test sentence for bigram language mode.

S1: <S> The resort was good </S>

S2: <S> The staff was really good</S>

S3: <S> Great food was there in resort </S>

S4: <S> Great food services </S>

Bag of Words →

(<S> The) (The resort) (was good) (resort was) (good </S>) (The staff) (staff was) (was really) (really good) (<S> Great) (great food) (food was) (was there) (there in) (in resort) (resort </S>) (food services) (services </S>)

BOW feature matrix →

	(<S> The)	(The resort)	(was good)	(resort was)
1	1	1	1	1
2	1	0	0	0
3	0	0	0	0
4	0	0	0	0

Gidesh Shrestri 2k15/IT/028

	(good <s>)	(the staff)	(staff was)	(was really)	(really good)	(good <s>)
1	1	0	0	0	0	0
2	1	1	1	1	1	0
3	0	0	0	0	0	1
4	0	0	0	0	0	1

	(great food)	(food was)	(was there)	(there in)	(in resort)	(resort <s>)
1	0	0	0	0	0	0
2	0	0	0	0	0	0
3	1	1	1	1	1	1
4	1	0	0	0	0	0

	(food services)	(services <s>)
1	0	0
2	0	0
3	0	0
4	1	1

Test Sentence →

1.) <s> The resort foods really good <s>

$P(\text{The}/\langle s\rangle) \cdot P(\text{resort}/\text{The}) \cdot P(\text{was}/\text{resort}) \cdot P(\text{really}/\text{was}) \cdot P(\text{good}/\text{really}) \cdot P(\langle s\rangle/\text{good})$

$= \frac{2^1}{4^2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{3} \times \frac{1}{1} \times \frac{2^1}{2_1}$

$= \frac{1}{24}$