Q:- Suppose you are given a following sentences:-

Jay Read Harry Potter.
Gauri Read a different Novel.
She read a Novel by Rushdie.

i) Learn a Bigram language Model using this data with add-1 smoothing.

ii) Using the language model learnt in Part(i) above, estimate the probability for the sentence.
" Jay Read a Novel "

⇒ Solution

<S> Jay Read Harry Potter </S>
<S> ~~Sam~~ Gauri Read a different Novel </S>
<S> She read a Novel by Rushdie </S>.

Bigram Probabilities

→ P(Jay | <S>) = 1/3
→ P(Gauri | <S>) = 1/3
→ P(She | <S>) = 1/3.
→ P(</S> | Potter) = 1/1 = 1
→ P(</S> | Novel) = 1/2
→ P(</S> | Rushdie) = 1/1 = 1.
→ P(Read | Jay) = 1/1 = 1
→ P(Harry | Read) = 1/3.
→ P(Potter | Harry) = 1/1 = 1
→ P(Read | Gauri) = 1/1 = 1
→ P(a | Read) = 2/3
P(different | a) = 1/2
P(Novel | different) = 1/1 = 1
P(read | she) = 1/1 = 1
P(Novel | a) = 1/2
P(by | Novel) = 1/2
P(Rushdie | by) = 1/1 = 1.

Add-1 Smoothing

Add-1 in numerator.
and Total words in vocabulary. $|V|$.
unique.

⇒ In given corpus there are 11 total unique words.

∴ Bigram Probabilities with add-1 smoothing are.

$P(\text{Jay} \mid \langle s \rangle) = \frac{1+1}{11+3} = \frac{2}{14}$.

$P(\text{Gave} \mid \langle s \rangle) = \frac{1+1}{11+3} = \frac{2}{14}$.

$P(\text{She} \mid \langle s \rangle) = \frac{1+1}{11+3} = \frac{2}{14}$.

$P(\langle /s \rangle \mid \text{Potter}) = \frac{1+1}{1+11} = \frac{2}{12}$.

ii) Probability for "Jay Read a Novel"

⇒ $P(\text{Jay} \mid \langle s \rangle) * P(\text{Read} \mid \text{Jay}) * P(a \mid \text{Read}) * P(\text{Novel} \mid a) * (\langle /s \rangle \mid \text{Novel})$

⇒ $\frac{2}{14} * \frac{1+1}{1+11} * \frac{2+1}{3+11} * \frac{1+1}{2+11} * \frac{1+1}{2+11}$

⇒ $\frac{2}{14} * \frac{2}{12} * \frac{3}{14} * \frac{2}{13} * \frac{2}{13}$

⇒ $\simeq 0.0001$