① One Hot Encoding

Dataset	Text	O/P
S1	The food is Good	1
S2	The food is bad	0
S3	Pizza is Amazing	1

Vocabulary → Size ⇒ (7)
↳ no. of unique words

{ The food is Good
bad
Pizza Amazing. }

Vector	The	food	is	Good	bad	Pizza	Amazing
The ⇒	1	0	0	0	0	0	0
food ⇒	0	1	0	0	0	0	0
is ⇒	0	0	1	0	0	0	0

⇓ one hot encoding

$$S1 = \begin{bmatrix} [1\,0\,0\,0\,0\,0\,0], \\ [0\,1\,0\,0\,0\,0\,0], \\ [0\,0\,1\,0\,0\,0\,0], \\ [0\,0\,0\,1\,0\,0\,0] \end{bmatrix}$$

$$S2 = \begin{bmatrix} [1\,0\,0\,0\,0\,0\,0], \\ [0\,1\,0\,0\,0\,0\,0], \\ [0\,0\,1\,0\,0\,0\,0], \\ [0\,0\,0\,0\,1\,0\,0] \end{bmatrix}$$

$$S3 = \begin{bmatrix} [0\,0\,0\,0\,0\,1\,0], \\ [0\,0\,1\,0\,0\,0\,0], \\ [0\,0\,0\,0\,0\,0\,1] \end{bmatrix}$$

Note :

Simple Mean

- Kisi bhi categorical value (Textya label) Ko 0 aur 1 Ke form me convert kr dete h TaKi Machine learning Model Samjh ske!!
- Mtlb har categorical Ke liye eK column banta hai aur jis category Ka data hota hai us column me 1 aata hai baaKi Mai 0.

Advantages :-

(1) Easy to Implement with python.
[sklearn , one hot encoder]

(2) Intuition is Easy

Disadvantages :-

(1) Sparse Matrix → Overfitting.

(2) ML Algorithm → fixed size I/P.

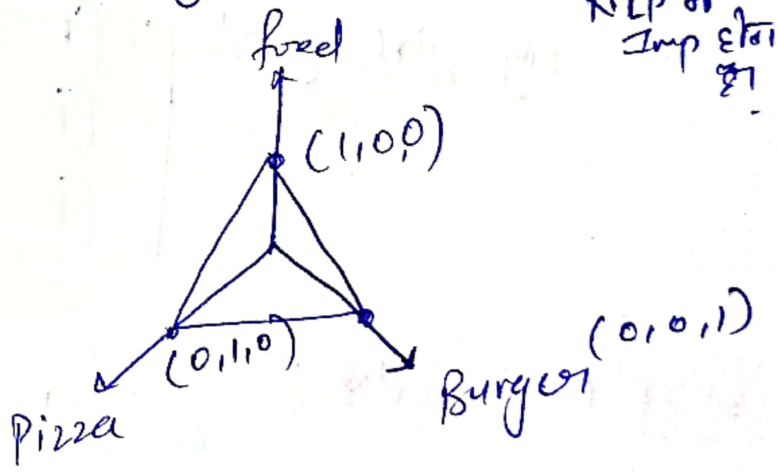
ent

f_1	f_2	f_3	O/P

(3) No Semantic Meaning is getting captured. ← NLP की Imp Step है।

ent: food Pizza burger

[1 0 0]
[0 1 0]
[0 0 1]



(4) Out of Vocabulary [OOV]

Scanned with OKEN Scanner