Question # 1

(a) Index encoding: Assign unique indices to each word. What are the indices for "Cat", "Dog", and "Bird"?

"Cat" : 0

"Dog" : 1

"Bird" : 2

(b) One-hot encoding: Assume one-hot encoding with vectors of length 3 (for simplicity). Provide the one-hot encoding representation for each of the three words. "Cat", "Dog", "Bird".

"Cat" : [1, 0, 0]

"Dog" : [0, 1, 0]

"Bird" : [0, 0, 1]

(c) Encode the following documents using Bag of words (BoW) technique:

Documents:-

1) "I like natural language processing. It is fascinating to see how computers can understand and generate human-like text".

2) "Natural language processing involves the interaction between computers and humans using natural language. It is a field of study that focuses on making computers understand and process language".

3) "I find the field of natural language processing very interesting. The ability of machines to comprehend and generate human-like language is a significant advancement in technology."

Vocabulary: {"I", "like", "natural", "language", "processing", "It", "is", "fascinating", "to", "see", "how", "computers", "can", "understand", "and", "generate", "human-like", "text", "involves", "the", "interaction", "between", "humans", "using", "a", "field", "of", "study", "that", "focuses", "on", "making", "process", "find", "very", "interesting", "ability", "machines", "comprehend", "significant", "advancement", "in", "technology"}.

Document 1: [1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]

Document 2: [0, 0, 2, 3, 1, 1, 1, 0, 0, 0, 0, 2, 0, 1, 2, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]

Document 3: [1, 0, 1, 2, 1, 0, 1, 0, 1, 0, 0, 0, 0, 0, 1, 1, 1, 0, 0, 2, 0, 0, 0, 0, 1, 1, 2, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1]

(d) Encode the following documents using term frequency - inverse Doument frequency (TF - IDF) scheme:

Documents:-

1). "I like natural language processing."

2). "Natural language processing is interesting".

3). "I find natural language processing fascinating".

Word	Document 1	Document 2	Document 3
I	1	0	1
like	1	0	0
natural	1	1	1
language	1	1	1
processing	1	1	1
is	0	1	0
interesting	0	1	0
find	0	0	1
fascinating	0	0	1

Question #2 :-

1) Lisa is my pet dog

Lisa : [1, 0, 0, 0, 0]

is : [0, 1, 0, 0, 0]

my : [0, 0, 0, 1, 0]

pet : [0, 0, 1, 0, 0]

dog : [0, 0, 0, 0, 1]

weights [0.2, 0.3, 0.1, -0.2, 0.5]

Output layer weights

= [0.6, -0.4, 0.2, 0.1, 0.3]

→ Forward Pass

"is" target word

"Lisa" "my" context word

Input vector = [1, 0, 0, 0, 0] | [0, 0, 0, 1, 0] = [1, 0, 0, 1, 0]

Hidden Layer Activation = relu $(\Sigma i w)$

= [0.13, 0.49, 0.14, 0, 0.93]

= sigmoid $(\Sigma h_i w_i)$

Output vector = [0.74, 0.15, 0.35, 0.13, 0.61]

→ Loss

= - sum (target-vector × log (output-vector))

= - sum (1 × log (0.74) + 0 + 0 + 0 + 0)

= - log (0.74)

= -0.322

Back Propagation:

$\frac{\partial L}{\partial w\text{-output}} = (T-O) \times \text{hidden-layer-activation}$

$= ([1,0,0,0,0] - [0.74, 0.15, 0.35, 0.13, 0.61])$
$\times [0.13, 0.49, 0.11, 0, 0.93]$
$= [-0.21, 0.11, -0.22, 0.00, -0.09]$

$\frac{\partial L}{\partial w\text{-hidden}} = [1,0,0,0,0] - [0.74, -0.15, 0.35, 0.13, 0.67]$
$\times [1,0,0,1,0] \times [0.2, 0.3, 0.1, -0.2, 0.5]$
$= [0.148, 0, 0, -0.26, 0]$

$w\text{-output} = w\text{-output} - \alpha \frac{\partial L}{\partial w\text{-output}}$

$= [0.6, -0.4, 0.2, 0.1, 0.3] - 0.01[-0.21, 0.11, -0.22, 0, 0.93]$
$= [0.5, -0.4, 0.2, 0.1, 0.274]$

$w\text{-hidden} = w\text{-hidden} - \alpha \frac{\partial L}{\partial w\text{-hidden}}$

$= [0.2, 0.3, 0.1, -0.2, 0.5] - 0.01[0.148, 0, 0, -0.26, 0]$
$= [0.186, 0.3, 0.1, -0.226, 0.5]$

Forward:-

target "my"

context "is", "pet"

input $= [0,1,0,0,0] \,|\, [0,0,1,0,0]$
$= [0,1,1,0,0]$

hidden = ReLU (w_i^o)

= ReLU [0, 0.3, 0.1, 0, 0]

= [0, 0.3, 0.1, 0, 0].

output = softmax $(h w_{out}^{i})$

= softmax ([0, 0.3, 0.1, 0, 0] - [0.5, 0.4, 0.2, 0.1, 0.2]

= [0, -0.12, 0.02, 0, 0]

2) Skip Gram:

W_{ih} = [[0.4, 0.2, 0.3][0.5, 0.6, 0.7]]

w_ho = [0.8, 0.9, 0.1]

=> **forward**

h = tanh $(W_{ih} + b)$

For "Lisa"

h-Lisa = tanh([[0.4, 0.2, 0.3].[0.5, 0.6, 0.7]]. [1, 0, 0, 0, 0]

+ [0, 0, 0]

= tanh ([0.4, 0.5])

= (0.481, 0.631)

h-is = tan [0.46, 0.674]

h-my = [0.517, .687]

h-pet = [0.562, 0.731]

h-dog = [0.557, 0.719]

Calculate Output:

$Y = \text{softmax}(w\_ho \times h + b\_o)$

For Lisa

$Y\_lisa = \text{softmax}(w\_ho \times h\_lisa + b\_o)$

$= \text{softmax}([[0.8, 0.9, 0.1] \times [0.481, 0.613] + [0,0]])$

$= [0.376, 0.624]$

$Y\_is = [0.382, 0.618]$

$Y\_my = [0.418, 0.582]$

$Y\_pet = [0.413, 0.587]$

$Y\_dog = [0.386, 0.614]$

Calculate Loss:-

$\text{Loss} = -\log(y\_i)$

For Lisa

$= -\log(0.624)$

$= -0.523$

$\text{Loss}\_is = 0.535$

$\text{Loss}\_my = -0.604$

$\text{Loss}\_pet = -0.599$

$\text{Loss}\_dog = -0.519$

$\text{Total Loss} = \Sigma L = 2.780$

Back Propagation:-

For Lisa

$\frac{\partial L}{\partial \theta} = [-0.376, -0.624]$

$\frac{\partial \text{Loss}}{\partial h} = [0.8, 0.9, 0.1] \times [-0.376, -0.624]$

$= [-0.28, -0.534]$

$$\frac{\partial Loss}{\partial in} = \delta(in) \times \partial 1$$

$= \text{sigmoid}([0.481, 0.613]) \times [-0.28, -0.534]$

$= [0.135, 0.188]$

Question # 3:-

It is raining

It : [1,0,0,0,0]

is : [0,1,0,0,0]

raining : [0,0,0,1,0]

Hidden State Initiazation

h_0 = [0,0]

Time step 1:

$h_1 = \tanh(U \times [1,0,0,0,0] + W \times h\_0)$

$= \tanh [0.1, 0.2, 0.1, 0.5]$

$= [0.1026, 0.1973]$.

Time step 2

$h_2 = \tanh(U \times [0,1,0,0,0] + W \times h\_0)$

$= [0.99, 0.96]$

$h_3 = [0.87, 0.50]$

$h_4 = [0.94, 0.61]$

$h_5 = [0.975, 0.69]$.

Output.

$$[0.10 \times 0.1, 0.19 \times 0.6] + [0.4 \times 0.1, 0.096, 0.6]$$
$$+ [0.87 \times 0.1, 0.6 \times 0.5]$$
$$= [0.1, 0.54] + [0.04, 0.34] + [0.087, 0.30]$$
$$= [0.29, 0.66].$$
$$= 0.29 + 0.66$$
$$= \frac{1}{1+e^{-0.95}}$$
$$= 0.72.$$

Loss:-

$$-y\log(\hat{y}) - (1-y)\log(1-\hat{y})$$
$$= 0.024.$$

Backpropagation:-

Question 3

is:

Input: [0, 1, 0, 0, 0]

$w \cdot h_{t-1} = [0.1, 0.2, 0, 0.2][0.76, 0.76]$

$= [0.76, 0.152]$

$h_t = \tanh[1, 1] + [0.76, 0.152]$

$\approx [0.8, 0.8]$

warm:

$v \cdot x_t = [1, 1][0, 0, 0, 1]$

$= [0, 1]$

$w \cdot h_{t-1} = [0.16, 0.87]$

Outside:

$v \cdot x_t = [1, 1][0, 0, 1, 0, 0]$

$= [0, 1]$

$w \cdot h_{t-1} = [0.1, 0.2, 0, 0.2][0.16, 0.87]$

$= [0.18, 0.37]$

$h_t = \tanh([0, 1] + [0.181, 0.374])$

$\approx [0.179, 0.88]$

Output :-

$\text{Sigmoid}(v \cdot h_3) = [0.1, 0.6][0.179, 0.88]$

$= 0.557$

Loss:-

$\text{Loss} = -y \log \hat{y} - (1-y) \log(1-\hat{y})$

$= 0.856$

$$= -y\log(\hat{y}) - (1-y)\log(1-\hat{y})$$

$$= 0.024$$

Back Propagation:-

b)

It : [1, 0, 0, 0, 0]

is : [0, 1, 0, 0, 0]

warm: [0, 0, 0, 0, 1]

outside: [0, 0, 1, 0, 0]

It:

$h_{t-1} = [0, 0]$ for time step 1

$$U.x_t = [1, 1][1, 0, 0, 0, 0]$$
$$= [1, 1]$$

$$w.h_{t-1} = [0.1, 0.2, 0, 0.2][0, 0]$$
$$= [0, 0]$$

$$h_t = \tanh[1, 1] + [0, 0]$$
$$\approx (0.76, 0.76)$$

$$V_{new} = V - \text{learning rate} \cdot \frac{\partial Loss}{\partial V} = [0.10039, 0.60]$$

$$U_{new} = U - \text{learning rate} \cdot \frac{\partial Loss}{\partial U} = [1.0014, 0.0002, 0.1084, 0.2012]$$

$$W_{new} = W - \text{learning rate} \cdot \frac{\partial Loss}{\partial W} = [0.12, 0.6, 0.347, 0.21]$$