**Nguyễn Đỗ Quốc Duy – 1810867**

**Câu 1:** Bảng tf của các từ:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | A | B | C | D | E |
| 1 | 0 | 2/3 | 1/6 | 1 | 0 |
| 2 | 1 | 1/2 | 3/4 | 1/4 | 0 |
| 3 | 0 | 0 | 1 | 5/6 | 0 |
| 4 | 1 | 1 | 1 | 1 | 1 |
| 5 | 1/11 | 2/11 | 1 | 0 | 0 |
| 6 | 2/7 | 0 | 0 | 1 | 3/7 |
| 7 | 0 | 1 | 1/4 | 0 | 1/2 |
| 8 | 1 | 1 | 1 | 0 | 1/3 |
| 9 | 1/2 | 1/6 | 1/6 | 0 | 1 |
| 10 | 1/6 | 1/6 | 2/3 | 1 | 0 |

Bảng Df của các từ:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | A | B | C | D | E |
|  | 7 | 8 | 9 | 6 | 5 |

Ma trận tf.idf của các từ:

idf = log2(N/df)

tf.idf = tf.log2(N/df)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | A | B | C | D | E |
| 1 | 0 | 0.214 | 0.0253 | 0.7369 | 0 |
| 2 | 0.515 | 0.16 | 0.114 | 0.184 | 0 |
| 3 | 0 | 0 | 0.152 | 0.614 | 0 |
| 4 | 0.515 | 0.322 | 0.152 | 0.7369 | 1 |
| 5 | 0.04678 | 0.059 | 0.152 | 0 | 0 |
| 6 | 0.147 | 0 | 0 | 0.7369 | 3/7 |
| 7 | 0 | 0.322 | 0.038 | 0 | 0.5 |
| 8 | 0.515 | 0.322 | 0.152 | 0 | 1/3 |
| 9 | 0.257 | 0.537 | 0.0253 | 0 | 1 |
| 10 | 0.086 | 0.537 | 0.1013 | 0.7369 | 0 |

**Câu 2:**

Hot encoding của các ký tự:

A: 10000

B: 01000

C: 00100

D: 00010

E: 00001

Ma trận sau khi encode chuỗi thứ 8 ABCAACCBBE

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1 | 0 | 0 | 0 | 0 |
| 0 | 1 | 0 | 0 | 0 |

Sau khi đi qua tầng ẩn:

|  |  |
| --- | --- |
| 0,287027 | 0,84606 |
| 0,902874 | 0,871522 |
| 0,537524 | 0,537524 |

Sau khi dùng hàm tanh:

|  |  |
| --- | --- |
| 0.2793961 | 0.68900 |
| 0.717694 | 0.70214 |
| 0.491111 | 0.49111 |

Sau khi đi đến output:

|  |  |
| --- | --- |
| 1.166497 | 1.30455 |
| 0.636222 | 0.792153 |
| 0.416227 | 0.680884 |
| 0.55986 | 0.928628 |
| 0.799771 | 1.086396 |

Sau khi dùng soft max:

(ký tự A, B)

|  |  |
| --- | --- |
| 0.303135 | 0.27581 |
| 0.178378 | 0.165226 |
| 0.143152 | 0.147827 |
| 0.165263 | 0.189386 |
| 0.210072 | 0.221751 |

Cross entropy:

Cột 1: -1log2(0.303135) = 1.721967

Cột 2: -log2(0.165226) = 2.59748

**Câu 3:**Với window size = 1, ký tự B nằm giữa 2 ký tự A và C nên

loss(B):

|  |
| --- |
| 1.85825 |
| 2.758018 |

**Câu 4:**

Encode của chuỗi: ABCAACCBBE

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 0.279396175 | 0.689006 | 0.517114 | 0.279396 | 0.279396 | 0.517114 | 0.517114 | 0.689006 | 0.689006 | 0.529896 |
| 0.717694392 | 0.702147 | 0.598675 | 0.717694 | 0.717694 | 0.598675 | 0.598675 | 0.702147 | 0.702147 | 0.629145 |
| 0.49111144 | 0.491111 | 0.50661 | 0.491111 | 0.491111 | 0.50661 | 0.50661 | 0.491111 | 0.491111 | 0.388473 |

Sau khi dùng 3-gram filter:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 2.712255 | 2.790224 | 2.677333 | 2.644326 | 2.595601 | 2.746826 | 2.86348 | 2.924671 |

Sau khi dùng 2-gram filter:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1.302539 | 1.202931 | 1.159456 | 1.192095 | 1.169744 | 1.137106 | 1.269901 | 1.335726 | 1.219328 |

Áp dụng 1 max-pooling, feature map sau cùng được tạo ra:

|  |
| --- |
| 2.924671 |
| 1.335726 |