

No.	Contents
1	BBBBCDDDDDD
2	ACDAAABBCC
3	CCDDDCDDCCC
4	ABCDEABCDEABCDE
5	ABCCCCCCCCCBC
6	DEADEADDDDE
7	BEBBBBEEECBBBC
8	ABCAACCBBE
9	AEEEEABCEEA
10	ABCCCCDDDDDD

W1 1. Xây dựng ma trận tf.idf

Từ câu 2 trở đi, sinh viên chọn hàng ứng với chữ số cuối trong MSSV của mình. Ví dụ sinh viên có MSSV 12345678 sẽ làm dữ liệu ứng với hàng 8 (ABCAACCBBE)

W3: AE 2. Với thông tin mạng AE như sau:

ma trận 3x5 từ Input đến tầng ẩn

wxh				
0,287027	0,84606	0,572392	0,486813	0,59
0,902874	0,871522	0,691079	0,18998	0,74
0,537524	0,09224	0,558159	0,491528	0,41

ma trận 5x3 từ tầng ẩn đến output

why		
0,37168	0,974829459	0,830035
0,39141	0,282585823	0,659836
0,64985	0,09821557	0,334287
0,91266	0,32581642	0,14463
0,71	0,27	0,83

Tầng ẩn dùng hàm tanh

Tầng output dùng softmax

Tính loss khi lần lượt xử lý 2 ký tự đầu tiên của chuỗi nhập

- W3: Word2Vec Skip-gram**
- Nếu dùng mạng AE như trên được sử dụng để áp dụng cho mô hình **skip gram** với **window size là 1**, tính loss khi ký tự thứ 2 của chuỗi nhập đang được sử dụng làm input
 - Nếu encoder của mạng AE trên được sử dụng để tiếp tục làm ma trận embedding của mô hình word2vec. Nếu tiếp theo chúng ta dùng một mô hình CNN với 2 filter ứng với 3-gram và bigram như sau **W4: CNN**

3-gram filter

0.6	0.7	0.9
0.2	0.9	0.7
0.3	0.3	0.30

bigram filter

0.1	0.5	0.2
0.3	0.8	0.1

Nếu dùng 1-maxpooling, cho biết feature map sau cùng được tạo ra.