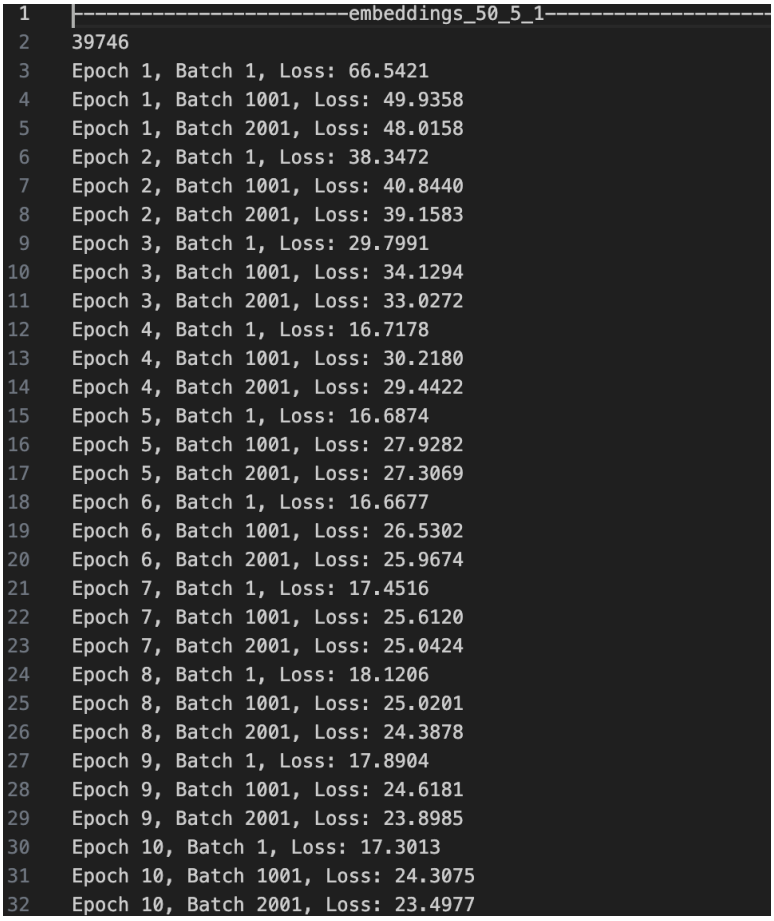
A2-write-up

Requirement 3

A screenshot of loss change

The following is an example with the hyperparameters:

emb_sizes=50, k_values=5, window_sizes=1, epoch=10

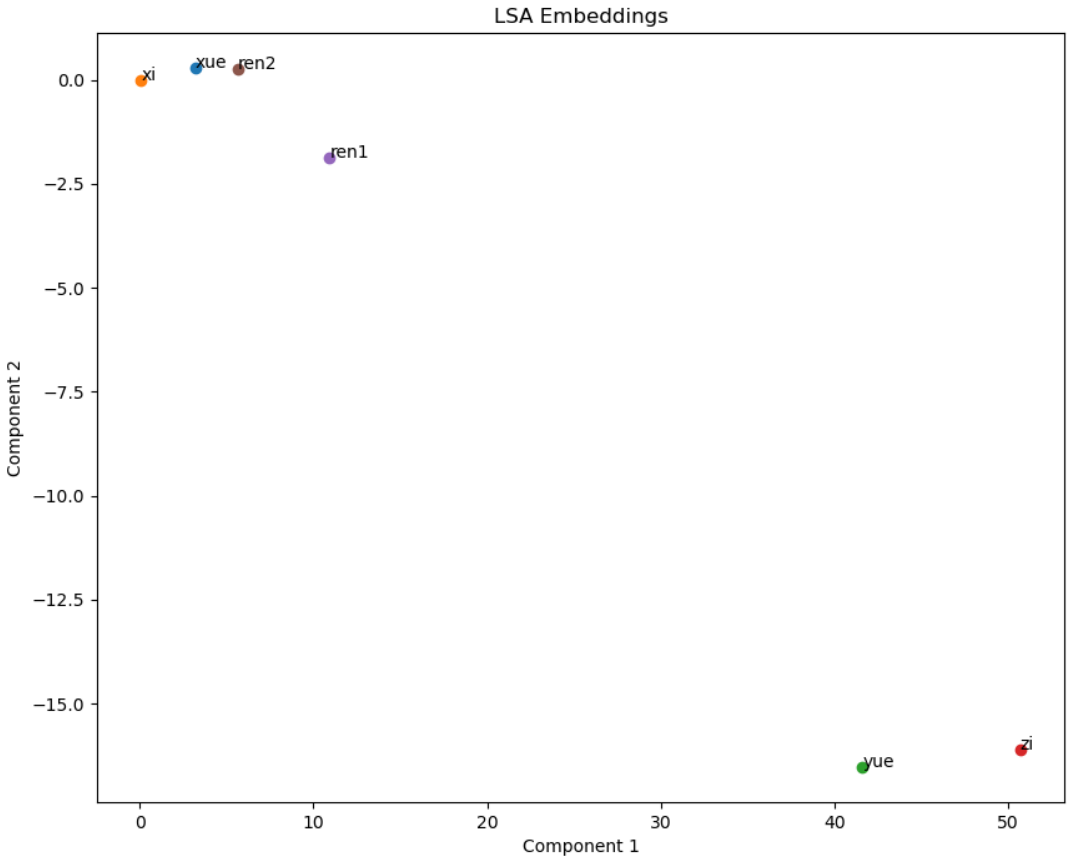


```
----------embeddings_50_5_1----------
39746
Epoch 1, Batch 1, Loss: 66.5421
Epoch 1, Batch 1001, Loss: 49.9358
Epoch 1, Batch 2001, Loss: 48.0158
Epoch 2, Batch 1, Loss: 38.3472
Epoch 2, Batch 1001, Loss: 40.8440
Epoch 2, Batch 2001, Loss: 39.1583
Epoch 3, Batch 1, Loss: 29.7991
Epoch 3, Batch 1001, Loss: 34.1294
Epoch 3, Batch 2001, Loss: 33.0272
Epoch 4, Batch 1, Loss: 16.7178
Epoch 4, Batch 1001, Loss: 30.2180
Epoch 4, Batch 2001, Loss: 29.4422
Epoch 5, Batch 1, Loss: 16.6874
Epoch 5, Batch 1001, Loss: 27.9282
Epoch 5, Batch 2001, Loss: 27.3069
Epoch 6, Batch 1, Loss: 16.6677
Epoch 6, Batch 1001, Loss: 26.5302
Epoch 6, Batch 2001, Loss: 25.9674
Epoch 7, Batch 1, Loss: 17.4516
Epoch 7, Batch 1001, Loss: 25.6120
Epoch 7, Batch 2001, Loss: 25.0424
Epoch 8, Batch 1, Loss: 18.1206
Epoch 8, Batch 1001, Loss: 25.0201
Epoch 8, Batch 2001, Loss: 24.3878
Epoch 9, Batch 1, Loss: 17.8904
Epoch 9, Batch 1001, Loss: 24.6181
Epoch 9, Batch 2001, Loss: 23.8985
Epoch 10, Batch 1, Loss: 17.3013
Epoch 10, Batch 1001, Loss: 24.3075
Epoch 10, Batch 2001, Loss: 23.4977
```

Requirement 5:

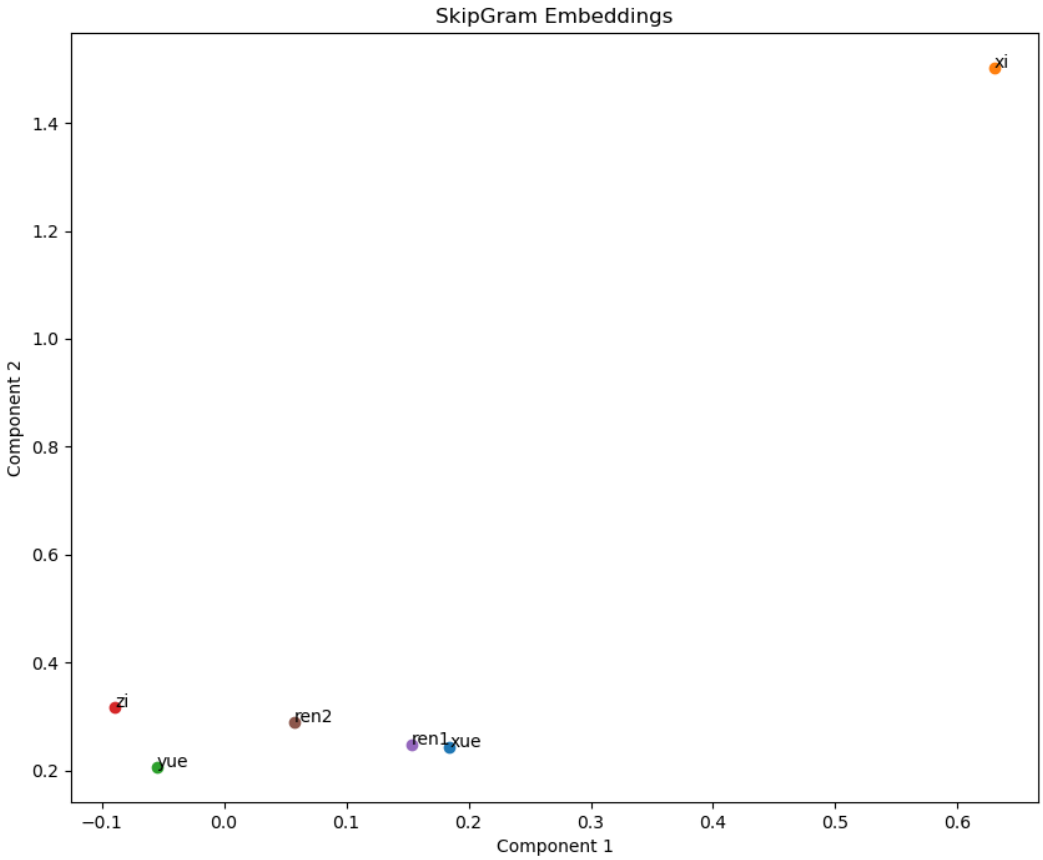
Compare your favorite embedding plot with the one we obtained from the LSA Lab.

LSA



Skip-Gram

emb_sizes=50，k_values=5，window_sizes=1



Briefly describe the difference in your write-up：

The obvious difference is that In the context of applying the Skip-Gram model, the character “习" has longer distance with other characters compare to LSA model. Even “学” is far away from “学” with Skip-Gram model.

The main reason is that Skip-Gram embeddings capture semantic similarities. Since "习" is used uniformly across diverse contexts, it lacks strong semantic associations with particular words. Additionally, the meaning of "习" might be broader than just "学". Therefore, its embedding could reflect this breadth, causing it to be distant from specific words, even those that are semantically related.

Compare to LSA, “学” and “习” are very close. This is because LSA focuses on the co-occurrence of words across the entire collection of documents, rather than within specific contexts or windows. Since "学" and "习" frequently co-occur in the same or similar documents, LSA would capture this global co-occurrence pattern and represent them as being close in the embedding space.