|  |
| --- |
| Question1  Learning rate=0.5    Learning rate=0.75    Learning rate=1 |
| learning rate' =0.5, num\_steps = 2001 |
| Average loss at step 0 : 290.188232421875 |
| Average loss at step 1000 : 2847.0221324768067 |
| Average loss at step 2000 : 5738.627358200073 |
|  |
| learning rate' =1, num\_steps = 2001 |
| Average loss at step 0 : 276.2760314941406 |
| Average loss at step 1000 : 11542.980423187257 |
| Average loss at step 2000 : 22726.921563049316 |
|  |
| learning rate' =0.75, num\_steps = 2001 |
| Average loss at step 0 : 186.97874450683594 |
| Average loss at step 1000 : 5134.203470703125 |
| Average loss at step 2000 : 9911.349928604126 |
|  |
| Learning Rate is the step size for each update of the coefficients, as words are repositioned in the feature space, nearby words shrink when it has larger learning rate, using a low learning rate in terms of making sure that we do not miss any local minima, loss decreases as learning rate increase in a certain range. it also needs a long time to converge .  Question2  windowsize=16, num\_steps = 2001, learn rate-1    Average loss at step 0 : 276.2760314941406  Average loss at step 1000 : 11542.980423187257  Average loss at step 2000 : 22726.921563049316    Window size=10, num\_steps = 2001, learn rate-1  Average loss at step 0 : 210.13246154785156  Average loss at step 1000 : 8806.808626303673  the position of a word within a given size window does affect the training, for example, a typical window size might be 10, meaning 10words behind and 10 words ahead (20 in total), more . It depends on the how many words using for training ,if window size of 10 can capture the context of a word , but 15 is choose , it will decrease the quality of the learnt model ,and vise versa.  Question3  ebedding=150, num\_steps = 2001, learn rate-1    Average loss at step 0 : 276.2760314941406  Average loss at step 1000 : 11542.980423187257  Average loss at step 2000 : 22726.921563049316    windowsize=100, num\_steps = 2001, learn rate-1  Average loss at step 0 : 186.80801391601562  Average loss at step 1000 : 7152.875771560669  Word Embeddings are the texts converted into numbers and there may be different numerical representations of the same text. If size is increasing , more similar words are found in similar locations.  Question 4  learning rate' =1, num\_steps = 2001    Average loss at step 0 : 276.2760314941406  Average loss at step 1000 : 11542.980423187257  Average loss at step 2000 : 22726.921563049316    learning rate' =1, num\_steps = 1001  Average loss at step 1000 : 10843.388273635865  Increase the number of steps to assert a stricken convergence |