Homework 3

Jackson Zaunegger

Question 1: How did the choice of initialization of word embeddings affect training of the LM and/or performance of the embeddings in the HW1 classifier? I found using a tensor to represent the integer values of the context words, and a tensor for the integer value of the target word to be the most effective. I tried using vectors that have a size of 1 x vocab size but I found that it took far to long to be feasible to use, so I left that idea.

I found using just the word indecies, the loss would converge quite quickly, in junction with the Adam optimizer and NLLLoss for the loss function. During my tests I found using these with a learning rate of 0.0001 allowed for convergence quickly, and can be minimize the loss further by enlarging the epoch count. I also found loading the dataset in with a DataLoader and using mini-batches drastically improved computation time when running the language model, I was using a batch size of 64 on my tests with the 500 line review file. I would assume that number would need to be increased for the sample with 1000 lines.

Question 2: Explain your choice of loss function, based on a comparison with at least one other loss function. While completing this homeowork, I checked a couple different loss functions to see how they would compare. I tried the NLLLoss() and CrossEntropyLoss(). I found that both performed well when creating the embeddings, and they preformed about the same, but NLLLoss seemed to produce better results but takes a little bit longer to compute. Since I seemed to be getting better results with NLLLoss I decided to keep using that.