Matthew Oakley
Dr. Rivas
Deep Learning W/ Tensorflow
12/10/2017
Final Exam

The strategy I took was just random in trying to find the best parameters and see what had the greatest impact on the test perplexity. I believe that the most influential parameters is the hidden size modifier. This is because when both hidden sizes had 8 epochs before the learning rate decayed the 400 hidden size had a much higher perplexity (1). I believe that early on the number of epochs at before the learning decay are least influential. This is seen in (2) where 0 and 2 epochs before decay have similar perplexities. Throughout test 1 training perplexity gets lower while valid perplexity gets worse. This could be because it is overfitting for the training dataset. The sentence is, "The crowd came the chief priests, the disciples had put him on the sabbath, so that they were afraid.<eos>Now after he" this is not bad when compared to earlier sentences where it would repeat words. The quality of the sentence is not tied to perplexity because an example is, "The King of the scribes came and put his clothes from a right side; and they were sitting around at him" this is where test perplexity is above 2000.

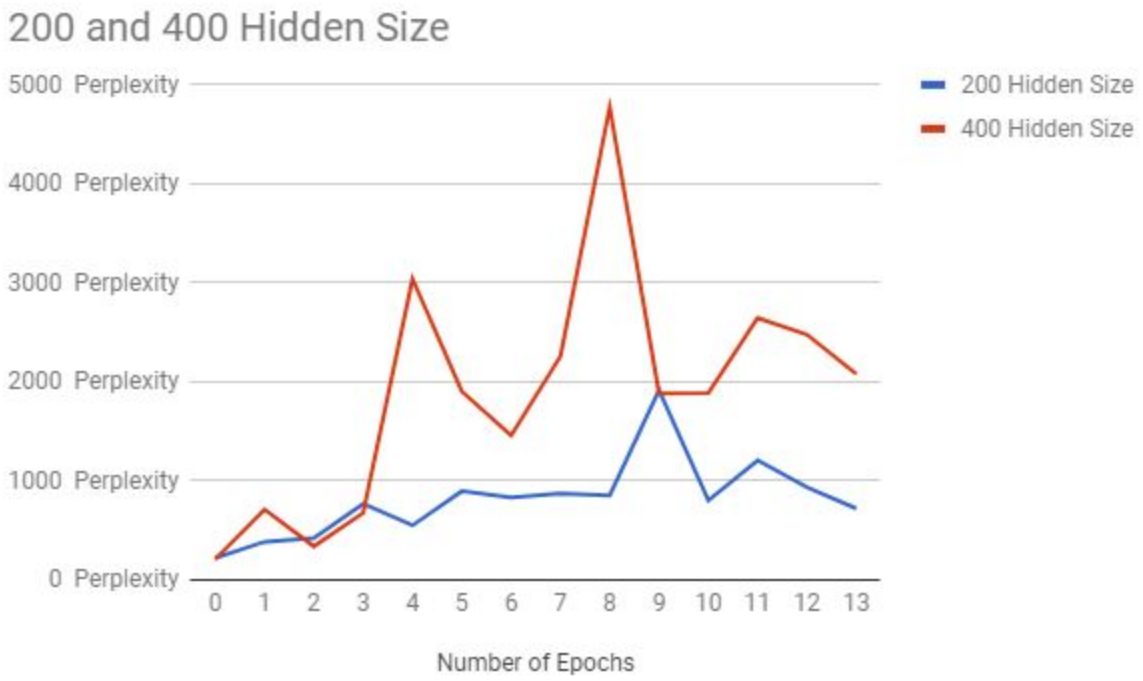

Figure 1. This shows the test perplexity of the mark dataset when increasing the number of epochs before learning rate decay. There was also a change of the hidden size to see if that had an impact of test perplexity.

Number of Epochs	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**Avg**
200 Hidden Size	220.055	384.296	422.365	769.619	550.128	895.282	832.959	870.538	854.14	1916.426	802.332	1208.109	934.827	720.887	812.997
400 Hidden Size	208.781	711.205	339.151	676.541	3039.729	1904.609	1460.274	2257.235	4773.31	1883.699	1886.376	2645.806	2477.017	2081.33	1881.790
600 Hidden Size	207.516														
Avg	212.117	547.751	380.758	723.080	1794.929	1399.946	1146.617	1563.887	2813.725	1900.063	1344.354	1926.958	1705.922	1401.109	

Figure 2. This is a chart with the test perplexity for each test. The best column then had its hidden size tested again to see if it lowered test perplexity. It did lower test perplexity but took much longer to run.

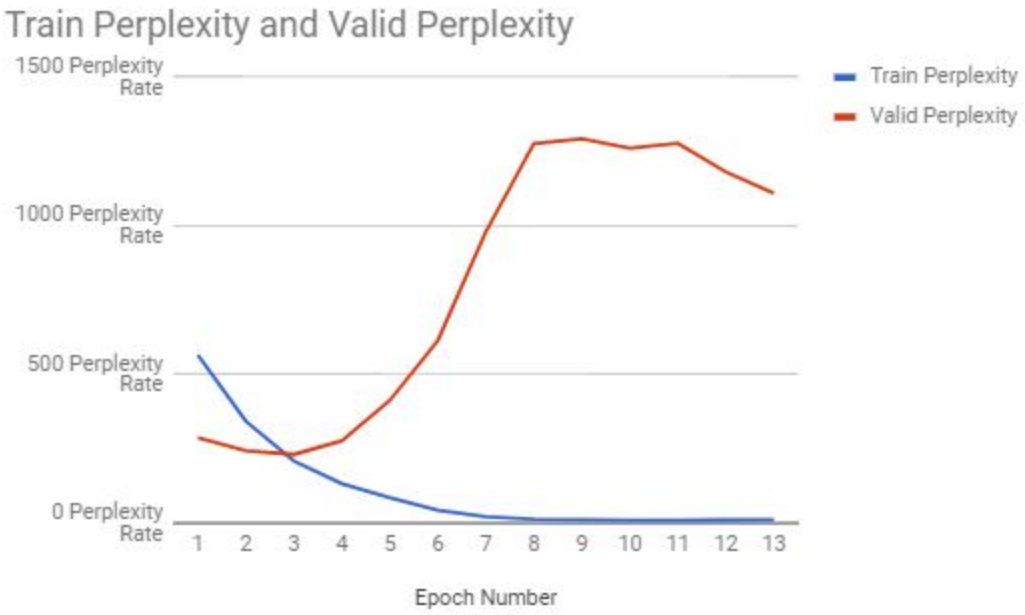

Figure 3. This is the Train and Valid Perplexity rate for 200 for hidden size and epochs before learning rate decay being at 4.