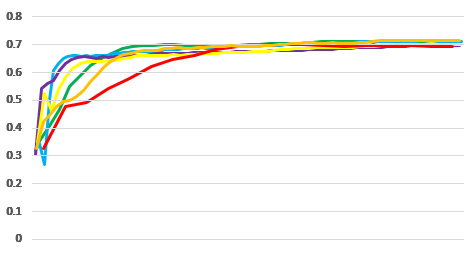
The results below uses the default learning rate of 0.0001 given in the code.

In the first graph, the color red, orange, yellow, green, blue and purple represent mini-batch size 1, 10, 30, 50, 100, 1000 respectively. Generally, it appears that the more mini-batches the algorithm uses, the more efficient it is. The purple line (mini-batch size 1000) reaches the peak in the first place, along with the red line (mini-batch size 1) being the slowest.



In the second graph, the x-axis represents the training set size 1000, 10000, 20000, 50000 and 100000 whereas the y-axis represents the accuracy. The result uses the default setting with 30 mini-batches. The features are selected as the most frequent 1000 words appeared in the training set using FreqDist. It appears that as the training set size goes up, the accuracy goes up too. But the downside is that as the training set size grows up, the runtime also increases a lot.

