Problem Set 3
Srikanth Kilaru

Part 1:
Final Test loss was 0.11786 and accuracy was 96.8%. Please see attached .png file for the screenshot of Part 1
Files: part1.png
Directory – part-1

Part 2:
Question 1:
The difference between the two graphs is that, in the small hidden_size of 8, the validation loss and perplexity decrease and stabilize with the number of training steps, but they first decrease and then increase again with the larger hidden_size of 256. This is because for the given data size the hidden_size of 256 (i.e. vector of size 256) is too big and over fits the data.
Files:
small-1.png
large1.png

Question 2:
Files:
dropout-0.1.png
dropout-0.3.png
dropout-0.5.png

Dropout value	Best_valid_ppl	Test_ppl
0.1	9.49	8.753
0.3	9.145	8.486
0.5	9.039	8.412

The difference between their learning curves is that as the dropout is lower the validation error goes up after reaching an initial low. The higher the dropout, the better the validation accuracy. This is because the high rate of drop out prevents over fitting during the training phase.

Question 3:
The samples with temperature as 0.01 are the best, with temperature = 0.5 producing not so great results and with temperature = 5, producing absolute garbage where they are not even English words. The high temperature produces a softer probability distribution over the classes and makes the RNN more easily excited by samples, resulting in more diversity and also more mistakes. Therefore low temperature produces a more robust prediction.
Files:
PS3-2-3.txt

Part 3:

I used the text version (included in zip as Salinger.txt) of the American classic –
"Catcher in the Rye" by J.D. Salinger.

1. With Default settings:
   Files:(p3-default.png)
   "best_valid_ppl": 3.7410857677459717,
   "test_ppl": 3.4930338859558105
2. With dropout = 0.5, num_layers=5, hidden_size=128, num_unrollings=12
   Files:(p3-best.png)
   "best_valid_ppl": 4.436203479766846,
   "test_ppl": 4.126055717468262,