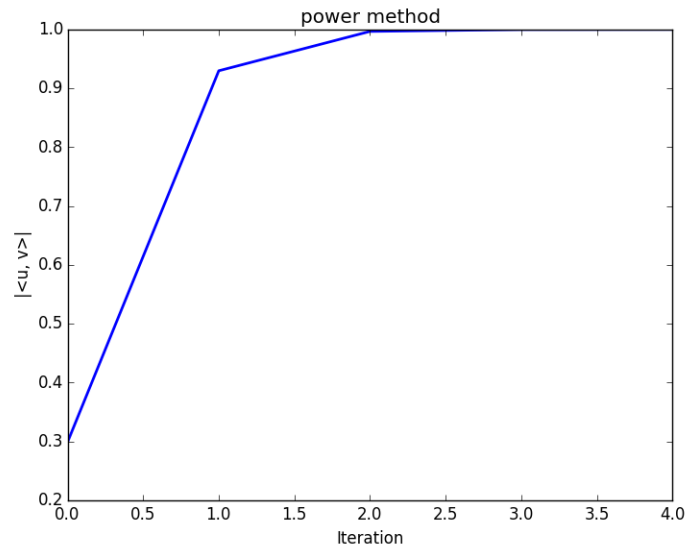


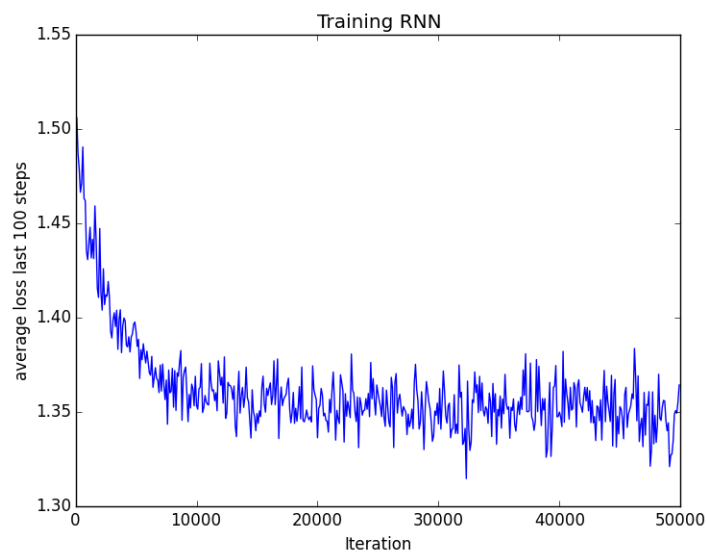
Exercise 10.1

- a)
Largest eigenvalue of M (greatest absolute value) = -12 with corresponding eigenvector $(\frac{-1}{3}, \frac{-2}{3}, \frac{2}{3})$
b) see **10.1.py**
c) The plot:



Exercise 10.2

- a,b,c,d,e,f) see **10.2_bonus.py**
The plot:



Bonus

a) see **10.2_bonus.py**

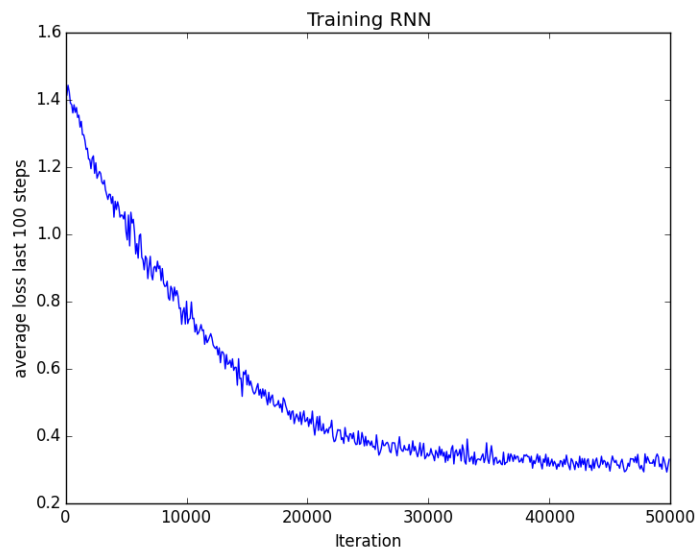
Achieved perplexity when testing on the whole sequence: ≈ 3.83

b) We can see that the perplexity is almost 4, so this model will predict the sequence almost randomly, which is not good. But, this is reasonable because the input data is also almost randomly.

Let's look at the second data set that we created "data2.txt":

"aaaaaaaaa...ccccccccc...ggggggggg...ttttttttt..."

in which each nucleotide is repeated 100 times. We have following result:



Perplexity for this data set: ≈ 1.06 , which is very good, because the input has a clear pattern.