

Name _____

Homework 4

COMP 435 – Machine Learning
Spring 2019

Due Wednesday, May 5 (by 11:55 pm)

Collaboration Policy

You may work individually or in pairs for this project.

Handin and Grading

Submit your generated story, kmeans plots and analysis, and your code to mygcc.

Problems

1. Green Eggs and Ham text generator.

You can grab the text from here:

<https://www.clear.rice.edu/comp200/resources/texts/Green%20Eggs%20and%20Ham.txt>

Step 1, choose your states (words punctuation etc.)

Step 2, read the file and create a transition matrix.

Step3

You will write a loop that samples from a first order Markov Model to simulate an AI bot that has learned to talk only from reading green eggs and ham. The steps are as follows.

- Given a word (optionally and its index)
- Sample the next word from the transition matrix
 - o Go to the row in the transition matrix that matches the word
 - o generate a random number between 0 and 1
 - o step over the cells in the row of the transition matrix that matches that word, creating a cumulative distribution as you go (or create the CDF separately in a preprocessing step and iterate over it) and grab the word at where the random number falls.
 - o Print out that word

Step 4

Print out the following:

- 1) The list of tokens you used as states for your transition matrix (i.e. Sam I . am) there should be ~50 of them (depending how you chose to handle punctuation).
- 2) The transition matrix (don't worry about formatting too much – I should be able to tell where one row ends and a new begins though).
- 3) 3 paragraphs, each with the same starting word (your choice for starting word) and 150 words per paragraph. (these should be different even though you start with the same word each time, your program should not be deterministic)

Please submit your code to mygcc