Homework 2: Markov Models of Natural Language

Name: Marlene Lin

Collaborators: black tea, black coffee, blah blah

This homework focuses on topics related to string manipulation, dictionaries, and simulations.

I encourage collaborating with your peers, but the final text, code, and comments in this homework assignment should still be written by you.

Submission instructions:

- Submit `HW2.py` on Gradescope under "HW2 - Autograder". Do **NOT** change the file name. The grade you see is the grade you get for the accuracy portion of your code (so no surprises). The style and readability of your code will be checked by the reader aka human grader.
- Convert this notebook into a pdf file and submit it on GradeScope under "HW2 - PDF". Make sure your text outputs in the latter problems are visible.

Language Models

Many of you may have encountered the output of machine learning models which, when "seeded" with a small amount of text, produce a larger corpus of text which is expected to be similar or relevant to the seed text. For example, there's been a lot of buzz about the new GPT-3 model, related to its carbon footprint, bigoted tendencies, and, yes, impressive (and often humorous) ability to replicate human-like text in response to prompts.

We are not going to program a complicated deep learning model, but we will construct a much simpler language model that performs a similar task. Using tools like iteration and dictionaries, we will create a family of **Markov language models** for generating text. For the purposes of this assignment, an n-th order Markov model is a function that constructs a string of text one letter at a time, using only knowledge of the most recent n letters. You can think of it as a writer with a "memory" of n letters.

```
In [ ]:    # This cell imports your functions defined in HW2.py

           from HW2 import count_characters, count_ngrams, markov_text,random
```

Our training text for this exercise comes from Jane Austen's novel *Emma*, which Professor Chodrow retrieved from the archives at (Project Gutenberg). Intuitively, we are going to write a program that "writes like Jane Austen," albeit in a very limited sense.

```
In [3]:    with open('emma-full.txt', 'r', encoding='utf8') as f:
```

```
        emma = f.read()
    print("done loading")
```

```
done loading
```

Problem 1: Define `count_characters` in HW2.py

Write a function called `count_characters` that counts the number of times each character appears in a user-supplied string `s`. Your function should loop over each element of the string, and sequentually update a `dict` whose keys are characters and whose values are the number of occurrences seen so far. Your function should then return this dictionary.

You may know of other ways to achieve the same result. However, you should use the loop approach, since this will generalize to the next exercise.

Note: while the construct `for character in s:` will work for this exercise, it will not generalize to the next one. Use `for i in range(len(s)):` instead.

Example usage:

```
count_characters("Torto ise!")
{'T': 1, 't' : 1, 'o' : 2, 'r' : 1, 'i' : 1, 's' : 1, 'e' : 1, ' ' : 1,
'!': 1}
```

Hint: Yes, you did a problem very similar to this one on HW1.

In [5]:
```python
# test your count_characters here
s = "Torto ise!"
print(count_characters(s))
```

```
{'T': 1, 'o': 2, 'r': 1, 't': 1, ' ': 1, 'i': 1, 's': 1, 'e': 1, '!': 1}
```

How many times does 't' appear in Emma? How about '!'?

How many different types of characters are in this dictionary?

In [6]:
```python
output = count_characters(emma)
```

In [7]:
```python
print("'t' appears " + str(output['t'])+" times")
print("total diff types = " + str(len(output)))
```

```
't' appears 58067 times
total diff types = 82
```

Problem 2: Define `count_ngrams` in HW2.py

An n *-gram* is a sequence of n letters. For example, `bol` and `old` are the two 3-grams that occur in the string `bold` .

Write a function called `count_ngrams` that counts the number of times each n -gram occurs in a string, with n specified by the user and with default value `n = 1` . Your function should return the dictionary. You should be able to do this by making only a small modification to `count_characters` .

Example usage:

```
count_ngrams("tortoise", n = 2)

    {'to': 2, 'or': 1, 'rt': 1, 'oi': 1, 'is': 1, 'se': 1} # output
```

In [8]:
```
# test your count_ngrams here
count_ngrams("tortoise", n = 2)
```

Out[8]:   `{'to': 2, 'or': 1, 'rt': 1, 'oi': 1, 'is': 1, 'se': 1}`

How many different types of 2-grams are in this dictionary?

In [9]:
```
# write your answer here
print(str(len(count_ngrams(emma,2))) + " types of 2-grams")
```

`1236 types of 2-grams`

Problem 3: Define `markov_text` in HW2.py

Now we are going to use our n -grams to generate some fake text according to a Markov model. Here's how the Markov model of order n works:

A. Compute (n +1)-gram occurrence frequencies

You have already done this in Exercise 2!

B. Pick a starting n -gram

The starting n -gram can be selected at random, or the user can specify it.

C. Generate Text

Now we generate text one character at a time. To do so:

1. Look at the most recent n characters in our generated text. Say that `n = 3` and the 3 most recent character are `the` .

2. We then look at our list of `n+1` -grams, and focus on grams whose first `n` characters match. Examples matching `the` include `them` , `the` , `thei` , and so on.
3. We pick a random one of these `n+1` -grams, weighted according to its number of occurrences.
4. The final character of this new `n+1` gram is our next letter.

For example, if there are 3 occurrences of `them` , 4 occurrences of `the` , and 1 occurrences of `thei` in the n-gram dictionary, then our next character is `m` with probabiliy 3/8, `[space]` with probability 1/2, and `i` with probability `1/8` .

Remember: the *3rd*-order model requires you to compute *4*-grams.

What you should do

Write a function `markov_text` that generates synthetic text according to an `n` -th order Markov model. It should have the following arguments:

- `s` , the input string of real text.
- `n` , the order of the model.
- `length` , the size of the text to generate. Use a default value of 100.
- `seed` , the initial string that gets the Markov model started. I used `"Emma Woodhouse"` (the full name of the protagonist of the novel) as my `seed` , but any subset of `s` of length `n` or larger will work.

Demonstrate the output of your function for a couple different choices of the order `n` .

Expected Output

Here are a few examples of the output of this function. Because of randomness, your results won't look exactly like this, but they should be qualitatively similar.

```
markov_text(s, n = 2, length = 200, seed = "Emma Woodhouse")

    Emma Woodhouse ne goo thimser. John mile sawas amintrought will on
    I kink you kno but every sh inat he fing as sat buty aft from the
    it. She cousency ined, yount; ate nambery quirld diall yethery,
    yould hat earatte
```

```
markov_text(s, n = 4, length = 200, seed = "Emma Woodhouse")

    Emma Woodhouse!"—Emma, as love,              Kitty, only this person
    no infering ever, while, and tried very were no do be very friendly
    and into aid,    Man's me to loudness of Harriet's. Harriet
    belonger opinion an
```

```
markov_text(s, n = 10, length = 200, seed = "Emma Woodhouse")
```

Emma Woodhouse's party could be acceptable to them, that if she
ever were disposed to think of nothing but good. It will be an
excellent charade remains, fit for any acquainted with the child
was given up to them.

Notes and Hints

Hint: A good function for performing the random choice is the `choices()` function in the `random` module. You can use it like this:

```python
import random

options = ["One", "Two", "Three"]
weights = [1, 2, 3] # "Two" is twice as likely as "One", "Three" three
times as likely.

random.choices(options, weights)

    ['One'] # output
```

The first and second arguments must be lists of equal length. Note also that the return value is a list -- if you want the value *in* the list, you need to get it out via indexing.

Hint: The first thing your function should do is call `count_ngrams` above to generate the required dictionary. Then, handle the logic described above in the main loop.

```python
In [4]:   import random
          # test your markov_text here
          markov_text(emma, n = 4, length = 200, seed = "Emma Woodhouse")
```

Out[4]: 'Emma Woodhouse, but in scarcel how much afrain. If you callence of plent, and h
         er all the relapsingly told utter, and I wanded exceptable, unperson her and ple
         ased by no in cottom one received he been'

Problem 4

Try out your function for varying values of `n` . Write down a few observations. How does the generated text depend on `n` ? How does the time required to generate the text depend on `n` ? Do your best to explain each observation.

(extra credit) What do you think could happen if you were to repeat this assignment but in unit of words and not in unit of characters? For example, 2-grams would indicate two words, and not two characters.

(extra credit) What heuristics would you consider adding to your model to improve its prediction performance?

```python
In [11]:  # write your observations and thoughts here
          import timeit
```

```python
import matplotlib.pyplot as plt
time = []
nrange = range(10,20)
for i in nrange:
    print("n = " + str(i))
    print('\n')
    %timeit print(markov_text(emma, i, 200, "Emma Woodhouse"))
```

n = 3

Emma Woodhouse trium, ther friended you much onceded head been unear to heat a b
ack at it would keepine you, (form not dive in howed thing from good—hause. A su
bjections to whened; by by talls, disple
Emma Woodhouse or I had lassemediates—that come which it could not period see hi
s jeall palpable had be spectly, name once object the quited ther ever frientryi
ng litter fortrick andmama say sylla sto
Emma Woodhouse, ther are said nother painsolutions of its for with the imagitate
d for humour. Perhappeak to escred began at the creeable look playful the is not
could nevery drew inion, and Emma known
Emma Woodhouse; "anot admirablent—"My frominent cle Weston was ever unlive joyme
n's extresparade hear to be came inted, arriagerned its that am subjection engt
h, who was of he, and of preturage. Harri
Emma Woodhouse mights wint, in as not they we he chame, marry bearsh of man hono
ugh them likes begantly any old prestoping.—Oh! does ther such such out delancea
ll know a draugh.
CHAPTER II
CHAPTER IX

Emma Woodhouse's if I am nextrade forte. Did had to ming a befor suffere only se
e on her; an been a could gone so party, up.—Ther was thous in the deal it witho
uld obline is sourse. Herent, in to ther
Emma Woodhousekeep morned the the her fast withosed that shed as an is such to t
o king sent injust at a good could regread not lause," recomma andall up a day b
oys, Miss Bates always, Harried.

"We ar
Emma Woodhouse two eviced her suggle of truth, his righbury."

Emma an done."

The "pering of be to minemembraid,

My died a marriendeprobable—but the hoped,

"What himself.

He was a gratious were ve
1.2 s ± 9.32 ms per loop (mean ± std. dev. of 7 runs, 1 loop each)
n = 4

Emma Woodhouse, that have not atter."

"You disappoint in speaking proposal a greatural in matched her quite and proof
the address friend unanswer; and himself."

"We came to affectly—I am quick direc
Emma Woodhouse was really low it sever not like Mr. Miss Bates, as Emma, "but sh
e disgraciously mightley's eyeing away. He next be a likely to be a merriet Smit
h—

"The have her and their gratitude as
Emma Woodhouse you wrong too."

"Ah! Miss Batest out he habits only as her—nother silent to the chat herself to-day—about.—I do not the eagerned thing gives, indeed. I can doors three more to the old
Emma Woodhouse hear. 'Dear engaged that Hartfield on she doubt her ridicular, Miss Fairfax in he was a marriet till be may charge no defined the Mr. Elton's been, all has before to be, or natured towa
Emma Woodhouse's expectably; "but Mr. Weston was not be more could of allowings. The end very pretter." In he money five you know the Cole road has he saved a truth faults; and know, to them."

How ve
Emma Woodhouse of anywhere no aggrandmama. Your thoughtened her, she law any quite sure this soon beauty of so opportablish views, I used though to her geography not his know are me glad them; and poo
Emma Woodhouse, slow, Knight not do. His a delayed not be question. I suppositive thing off dissued to say quite for a crowd. Mr. Frank your obscuriosition it with his own without her?—I was a good en
Emma Woodhouse. "I am rathere at Harriet, smiles of humour object to Jane?—for people. She will be exclaim, would do young lister; that's been now you have my brings too go; and respectionate founds a
2.61 s ± 25.8 ms per loop (mean ± std. dev. of 7 runs, 1 loop each)
n = 5

Emma Woodhouse attention, and the hint of us; Mr. Knight." They were deep mourners of fancying any morning two minutes. And as son;—"but you will not to Mrs. Cole, that he evening, "that is—I did not
Emma Woodhouse's feeling his have him and listence of this was a principle of the refresh pleading her on her husband.

"It was soon as he courself. The must long silent; and good school was you were
Emma Woodhouse with Mrs. Elton was asked me without a little dear, did not thing but her first sums, would have nothing that was stood behaviour before shew her," said he, "in talked of thinking, or g
Emma Woodhouse designs.

Harriet. The fears in half and the hour. You must be intent natural to be promise. But I must not taken if she fears, with her; but in one of their accounts."

"Well, settling
Emma Woodhouse in a warm you," said thing only added, and of all was soon. Those which all this master at they were any day; and pleasure her think of Mr. Weston, who brought at the most that supplant
Emma Woodhouses I absolutely to hear no ended that shall not, and of her so bear these can have overcome to a very of his complete dinner, promote to her."

"Ungratulation the were am I told you—just
Emma Woodhouse have looked of Bath, and as to ask her own sake, I would fancied in Bath; and whether common silence him her situations you very feared very fit to be very like of it before, a very thi
Emma Woodhouse) appear value simplicity of Mrs. Churchill we exercise a man can this man.—I do not help it."

"What rose and still she was extreme polite. He had a little zigzags of the very obliged t
4.93 s ± 107 ms per loop (mean ± std. dev. of 7 runs, 1 loop each)
n = 6

Emma Woodhouse country familiar, ignorant, and master. You could, whole paper is persuaded away, thank him rather and she asked, produce an augmentatious now. And so I always kindness, from her, which
Emma Woodhouse, shall be quite surprizes be expense of his air and such a trial of friended, in for Emma could be so. I think of man, but so good a letter, for

any thing to you begged her his journey
Emma Woodhouse, we may not pleasure you soon it was a character. The style he wi
ll probably in the practice of ill-fame. Mrs. Weston venture, so particularly co
ntented;" and that the uneasy. I have a
Emma Woodhouse, and all the doors. Whenever she ought to any enjoyment of what y
ou say to your own. I can easy credulity. He was close by her ungratefully by th
e bleak part of them, and from Scotland,
Emma Woodhouse?—Till the most respectable, all extravagance, with Jane was going
to the housemaid that."

"I shall endeavour of beauty of Hartfield as you made her so natured as if, like
to you like H
Emma Woodhouse. I never married many cold—and attention of, however, it is total
ly regret; but still do so many feelings of each.

The evening, the more of little disagree with me."

The joke; he once
Emma Woodhouse's perfectly as soon continued, "I dare to settled. There is presu
mptuous treated by Mrs. Weston doat upon the world I could ask; but to this mome
nt at Hartfield told me if you thinks it
Emma Woodhouse, the back, that respecting, of their husband, she had foresight!—
especially. Soft is to be the achieved.

The different was before I can be no long as her two to ask me what you shall l
8.39 s ± 103 ms per loop (mean ± std. dev. of 7 runs, 1 loop each)
n = 7

Emma Woodhouse would have cooled himself of the daily happiness in marrying away
by her own sheets; an excellent precaution; was devoted to write verses and I wi
ll be an arrangement, that Miss Fairfax
Emma Woodhouse, (speaking pleasure.

The ladies of so sudden trial of this is a peculiar object of my friendship, wer
e more for me. William Larkins will be satisfied with the greatest satisfaction.
—He
Emma Woodhouse"—he stopt by Mrs. Weston, on his entirely her instrumental, she c
ould not strong. The service—and how much handsome preserved moment comfortable
praises, always thoughts of maternal kin
Emma Woodhouse's. My mother's side to talk to him hither with me."

"Emma," said Harriet rational agitations of pleasures of life."

"She meaning whist club establish a musical knowledge of the issue.
Emma Woodhouse's feelings. She did not as likely to give all the evening, and ca
rrying any thing so warm admiration of never dusty, and for me.—Oh! and I must c
ontaining his servants, and minutes, per
Emma Woodhouse," she really are hearing him to talk—and yet it is not at all a t
iresome to look taller than she is with a blush.

"None; not allowing for a string?' and somebody had ever be at home an
Emma Woodhouse, believe that I did it against accept your health, not infection.
I could be made themselves; or were spacious smile—"much less than the occasion
s. It is a very improvided at supposed t
Emma Woodhouse to move when satisfied with such risks.—Liable absence.

Mrs. Weston found the nephews taken, the prospect of shade, you know they parte
d. A word was the very differently. He had undert
11.6 s ± 87.8 ms per loop (mean ± std. dev. of 7 runs, 1 loop each)
n = 8

Emma Woodhouse, I do not know that I should not have heard it asserted. The heat
he was no obscurity, nothing of this sort, you know, thinks herself.—In ten minu

tes, and on the fogs of December. But I
Emma Woodhouse. I absolutely seriously good to outweigh every moment all the hor
ses. He had, by gentle encouraging you Mr. K.—But I never could say no more than
any thing, a suspicion entering the bar
Emma Woodhouse was silent people they talked of, and when Mr. Martin.

High in the candour and a half ago. He had gone three long months!—to be allowed
them that it was to bear a different sort, but e
Emma Woodhouse, full of the fancy for having gone to leave town in the country t
o some, and a very earnestly lamented the name of manner; and I have a great dea
l."

"And I do assure her, very retired
Emma Woodhouse, who readily acceded to ask my opinion of things, and nothing to
encounter any confidence of family that I have happened. Harriet, poor dear Isab
ella were standing or knowledge; and, we
Emma Woodhouse's ear; and she have chosen for the envious feelings in such restr
ain him all these deficient.—She did it happened; a little about Highbury, she i
s a person than modern ease of common co
Emma Woodhouse felt no unwilling to lose you—just as accomplishment!" said Mrs.
Elton's very pleasant nature—it is impossible to myself, I will give it houseroo
m till Colonel and Mrs. Campbell are not
Emma Woodhouse must not married wrong—and not use his caring for the proper disc
rimination to his son to you, who had been too dreadful loss to you; and to feel
a strong a sense of ill-usage,) which m
14.6 s ± 169 ms per loop (mean ± std. dev. of 7 runs, 1 loop each)
n = 9

Emma Woodhouse, you do not like morning, we have seen her? Indeed you are a grea
t deal of health. Where features, "I am very sorry to part, and considered it at
first," said she. "I certainly has not
Emma Woodhouse, he will not be surprized—but it does not read?"

"Oh yes!—that is, I dare say, an excellent servant; and a Harriet expressed mann
er has any things do not imagine it just now, meant not
Emma Woodhouse is so obliging, and spending their box of letters? It used to sta
y at home by herself.—The first, were a recommendation to Miss Bates, "then I ne
ed not bring the path.... The only taste
Emma Woodhouse, with a degree of excellent education. Living constantly passed s
uspicion of—and in strange things to call her father. Had he begun at Mr. Cole n
ever to expose it in any degree that mad
Emma Woodhouse revived him a very handsome. I thought so. And he is a gentle sel
fishness of a rather glad that there should have been applying all that I must n
ot appear—but no, she would not deny tha
Emma Woodhouse's visits; now she feared, on Mr. Knightley; and, to her more than
any other part towards her friend in Windsor, to whom, a few weeks at least thes
e three minutes, Mr. Woodhouse or Mr. K
Emma Woodhouse—if you are much better sense from any allusion, independence or i
ndependent on the subject, did not, could not think, without a mother in affecti
on was not regular, equal, kindly interc
Emma Woodhouse was persuade away. But why must she considerate!—But however he m
arried.

Jane Fairfax might be considerable influence failed, he then mentioned it to mea
n any other people charming! Yo
17.5 s ± 200 ms per loop (mean ± std. dev. of 7 runs, 1 loop each)

In [7]:
```python
# write your observations and thoughts here
nrange = range(10,20)
for i in nrange:
    print("n = " + str(i))
    print('\n')
```

```
    %timeit print(markov_text(emma, i, 200, "Emma Woodhouse"))
 markov_text(emma,n=10,length=200,seed="Emma Woodhouse")
```

n = 10

Emma Woodhouses were fixed, till roused by Miss Woodhouse was safely covered, wh
en every body, was interest much oftener than Emma had not added any lasting war
mth, that she could not put them all to
Emma Woodhouse, but your kindness and reserve—such apparent sullenness, she had
received a parcel from her, excepting one, have we seen either Mr. or Mrs. Elto
n. Only one page more. I shall have no do
Emma Woodhouse soon followed appeared that London was not think of her?—Is not s
he?"

"I have not forgotten one matter of fact, but here she is very rarely of use to
Mr. Woodhouse, what can any body.
Emma Woodhouse, do talk to me and make her instrument is indifferent ways of bei
ng prepared—and her eyes which had made up his mind to take leave. This has been
used to sit and watch the due time, wal
Emma Woodhouse, who did not look like the same. And when you first in consequenc
e I do not understand how they got there. Seven miles were to be early, as the l
atest interesting to each other. From ou
Emma Woodhouse came into the doubtful room, observed, "always overcareful for ev
ery body's mouth, and dated Sept. 28th—and began, 'My dear Madam,' but I forget
what passed between them, Emma's feeling
Emma Woodhouse, I wish her no evil.—No, let them run about a great many errands
for Mrs. Weston, "as can well be; and while meanly exerting her father and Mrs.
Weston's good wine, and felt sure that h
Emma Woodhouse described? Hazle eyes excepted, two years more might marry any bo
dy else—and the situation, seemed like a young man, I believed every body's wish
es for the world, fastening in the world
19.8 s ± 104 ms per loop (mean ± std. dev. of 7 runs, 1 loop each)
n = 11

Emma Woodhouse, do you think I ought to add, with the additional softening claim
of a lingering illness of his heart and sweet temper, to understand that Frank a
dmired her complexion. My dear Mrs. Elt
Emma Woodhouse, or any other young creature of the room, he would have lasted lo
nger. Emma was very disagreeable—I shall never change, he could not add. His per
ception of any thing like it, he will be
Emma Woodhouse," said Harriet, contemplation of his life!"

Miss Fairfax and her father so well satisfied without their being left in solita
ry female walking, and in the most natural than one crowd, a
Emma Woodhouse, I was absolutely thrown away. Emma knew it to have behaved, had
the young man who told her he loved and esteemed Isabella was too much agitated
to encounter each other at once, that up
Emma Woodhouse-ing me!—But upon my honour, there is but one married woman in the
world is not the power of doing justice to. Mrs. Weston, "and on others, very li
ttle tea if you please, Mr. Churchill p
Emma Woodhouse, full of trivial communicated; the door between the do-nothing an
d the do-all. If I had followed, of reckoning up exactly what he and all his fat
her had lived with the woman he had want
Emma Woodhouse-ing me!—But upon my honour, there seemed a principle of separatin
g allowed herself in the considerate, or benevolent nerves till it were proved a
gainst Emma Woodhouse, of Hartfield's at
Emma Woodhouse, rather warmly, "you are quite odious. What are Harriet Smith was
a girl of such steadiness of character was not a large hand, but very clear and
decided it; and it will be quite unnece
21.9 s ± 208 ms per loop (mean ± std. dev. of 7 runs, 1 loop each)
n = 12

Emma Woodhouse, how you do forget!"

"My dear Emma," said Mr. Knightley. We invalids think we are privileged people."

"My dear Isabella's sake; and who consequently to be endured, though their feeli
n
Emma Woodhouse-ing me!—But upon my honour, there is nothing to be done in a mome
nt, was increasing coldness, he seemed resolved to remain with him, than with an
y other; and, therefore, set off to-morr
Emma Woodhouse, how could she have been completely frightened enough about the c
ountry. She believed she should never allow for Mr. Knightley, "whoever may have
had this hot walk to no purpose."

Emma
Emma Woodhouse, and the ostler had been of some use; but it is not at all afrai
d. I should certainly a very cheerful look and manner has any chance with Harrie
t possible but by note; no church for her
Emma Woodhouse, for such surprising good fortune of the engagement at present. I
n the daily circle—but that is nothing in the world, to have created so steady a
nd persevering attentions your heart at
Emma Woodhouse, tell me what I ought never to have kept—I know that in fortune a
nd affectionate father—certainly as to the guilt of those who handle it. Mr. Kni
ghtley and Harriet, when they were half
Emma Woodhouse, you need not be at any pains to reconcile me to the match at Ran
dalls. There does seem a charming introduction to what is lost. He came back, he
told us what John ostler had stood out
Emma Woodhouse, that supposing—that if—strange as it may deserve him; and now th
at I understand—dearest Jane's prospects—that is, I do not know whether the inco
nvenience, an oppression for ever.

The
23.2 s ± 137 ms per loop (mean ± std. dev. of 7 runs, 1 loop each)
n = 13

Emma Woodhouse-ing me!—But upon my honour my performance on the pianoforte, and
of her voice! I have heard so much of her) as in a line of society above him. I
was very much pleased, and who were full
Emma Woodhouse-ing me!—But upon my honour my performance is mediocre to the last
degree. You, Miss Woodhouse, I wish you would not have me if I were to ask her—a
nd I am very much obliged to you for it
Emma Woodhouse, how do you do?—Very well, I thank you. I could not comprehend i
t. She was as happy as yourself."

"Yes, she would be pleased to wait a moment, or he would think any thing of Miss
Fairf
Emma Woodhouse, I wish her no evil.—No, let them be ever so happy together. I am
amused by one part of the subject required; concluding, however, was severe; and
by the ten days' end, her nephew's let
Emma Woodhouse, I could deserve nothing wrong in his notions, less of the match
I felt to be all on her side; and had not the smallest doubt that, could he see
his little effusion honoured as I see it
Emma Woodhouse, he was obliged, in spite of the compliments to Mrs. Bates's, he
may get there from the Crown in good time, the Randalls party agreed to it; but
added, that she should be gratified by t
Emma Woodhouse, and the only source whence any thing like it? Your neighbours th
ese ten years."

"No, upon no account proceed.

Emma's first resolution was to keep her with them." Mr. Elton looked up,
Emma Woodhouse-ing me!—But upon my honour my performance is mediocre to the last
degree. You, Miss Woodhouse, who had previously made up his mind, and does not s
eem in the least. What can it possibly

24.3 s ± 105 ms per loop (mean ± std. dev. of 7 runs, 1 loop each)
n = 14

Emma Woodhouse, handsome, clever, and rich, with a comfortable home for her, and
trying to understand each other now, without the smallest remark; and, excepting
her own partner, there was no recoveri
Emma Woodhouse-ing me!—But upon my honour," said he very seriously, "it does no
t. It is not in the smallest wish for your marrying Mr. Martin. While you were i
n the smallest doubt (nor have I now) tha
Emma Woodhouse, and only wanted invitation, to give the appearance of every thin
g was due; a separation for the present—to entreat her to come in; my mother wil
l be so very happy to walk into the same
Emma Woodhouse, whatever your penetrating eyes may fancy—I am sick of England—an
d would leave it to-morrow, if I could."

"You are in luck.—Your only blunder was confined to my ear, when you imagined
Emma Woodhouse, and defying Miss Smith with her; a most welcome addition to ever
y source and every expression of the scene being laid at Mr. Cole's, or balls at
the Crown. The difference which I never
Emma Woodhouse, where the waters do agree, it is quite wonderful how she does he
r hair!—No hairdresser from London I think could.—Ah! Dr. Hughes I declare—and M
rs. Hughes for a moment. She could now l
Emma Woodhouse-ing me!—But upon my honour," said he very seriously good principl
es, admire it:—elegance, which, whether of person or of mind, she saw so little
in Highbury. The less worthy females wer
Emma Woodhouse, I would rather have a stranger preferred than one's very particu
lar friend always at hand, I, probably, never shall again; and the conviction se
emed real; he looked as if he felt it. S
25.4 s ± 72.5 ms per loop (mean ± std. dev. of 7 runs, 1 loop each)
n = 15

CHAPTER I

Emma and Harriet were going to separate in preparation for the regular four o'cl
ock dinner, the hero of the evening. The dews of a summer evening are what I wou
ld not expose any body to. Ho
CHAPTER I

Emma and Harriet belonged to Frank Churchill to be at all deserving Emma, was th
ere so much fond solicitude, so much keen anxiety for her doing right, which no
other creature, and that she
CHAPTER I

Emma Woodhouse, who would be so very painful a conclusion of the first series of
reflection. This was the fact. My right to place myself in a situation of extrem
e difficulty and uneasiness,
CHAPTER I

Emma Woodhouse, and her more active, talking daughter, almost ready to sink unde
r the agitation of this moment. The dread of being supposed ashamed. You are not
striving to look taller than
CHAPTER I

Emma and Harriet went together; Miss Bates and her niece, with the Eltons?—Here
is quite a separate puzzle."

Upon her speaking her wonder aloud on that part of the subject. I planned the m
CHAPTER I

Emma and Harriet had been walking together, he had so often come and walked by h
er, and talked so very delightful to have you all meet so!—Our little room so ho
noured!"

She was in the adjo
CHAPTER I

Emma and Harriet had been walking together, he had so often come and walked by h
er, and talked so very delightful apples, and Mrs. Wallis does them full justice
—only we do not have them bak
CHAPTER I

Emma Woodhouse, who had previously determined how far she would proceed, on any
application of the sort; and it would be incumbent on her to avoid any encourage
ment to begin.—She did not, c
25.8 s ± 159 ms per loop (mean ± std. dev. of 7 runs, 1 loop each)
n = 16

CHAPTER I

Emma and Harriet had done nothing which woman's friendship and woman's feelings
would not justify.

He had frightened her a little about the seeming incongruity of gentle manners a
nd a conc
CHAPTER I

Emma Woodhouse, I could deserve nothing from either. Acquit me here, and procure
for me, when it is allowable, the acquittal and good wishes of that said Emma Wo
odhouse, I could deserve not
CHAPTER I

Emma and Harriet went together; Miss Bates and her niece, with the Eltons; the g
entlemen on horseback. Mrs. Weston remained with Mr. Woodhouse, not thinking the
better of her other complain
CHAPTER I

Emma Woodhouse, I could deserve nothing from either. Acquit me here, and procure
for me, when it is allowable, the acquittal and good wishes of that said Emma Wo
odhouse, whom I regard with
CHAPTER I

Emma Woodhouse, I could deserve nothing from either. Acquit me here, and procure
for me, when it is allowable, the acquittal and good wishes of that said Emma Wo
odhouse, whom she really wan
CHAPTER I

Emma Woodhouse, whom I regard with so much brotherly affection for you, he is so
far from making flourishes, that any other fine lady ever did. Mrs. Churchill li
ved, I suppose there are. Bu
CHAPTER I

Emma and Harriet professed very high expectations of extraordinary virtue from t
hose for whom education had done so little; entered into the idea with thorough
enjoyment, and Mrs. Weston. B
CHAPTER I

Emma Woodhouse, handsome, clever, and rich, with a comfortable home for her, and
Emma imagined a very sufficient income; for though the event of the 26th ult., a
s you will conclude, immedia
25.9 s ± 378 ms per loop (mean ± std. dev. of 7 runs, 1 loop each)
n = 17

CHAPTER I

Emma Woodhouse, handsome, clever, and rich, with a comfortable home for her, and
Emma imagined a very sufficient income; for though the event of the disclosure b

```
ear an equal resemblance!—Bu
CHAPTER I

Emma Woodhouse, handsome, clever, and rich, with a comfortable home and happy di
sposition, seemed to unite some of the best society in the place. A line from me
would bring you a little hos
CHAPTER I

Emma Woodhouse, handsome, clever, and rich, with a comfortable home and happy di
sposition, seemed to unite some of the best society in the place. A line from me
would bring you a little hos
CHAPTER I

Emma Woodhouse, handsome, clever, and rich, with a comfortable home and happy di
sposition, seemed to unite some of the best society, that with all the gentlenes
s of his address, true elegan
CHAPTER I

Emma Woodhouse, handsome, clever, and rich, with a comfortable home and happy di
sposition, seemed to unite some of the best blessings of existence; what could b
e increasing Emma's dislike.
CHAPTER I

Emma Woodhouse, handsome, clever, and rich, with a comfortable home for her, and
Emma imagined a very sufficient income; for though the event of the disclosure b
ear an equal resemblance!—Bu
CHAPTER I

Emma Woodhouse, handsome, clever, and rich, with a comfortable home and happy di
sposition, seemed to unite some of the best blessings of existence; and had live
d nearly twenty-one years in
CHAPTER I

Emma Woodhouse, handsome, clever, and rich, with a comfortable home for her, and
Emma imagined a very sufficient income; for though the event of the disclosure b
ear an equal resemblance!—Bu
26 s ± 155 ms per loop (mean ± std. dev. of 7 runs, 1 loop each)
n = 18


CHAPTER I

Emma Woodhouse, handsome, clever, and rich, with a comfortable home for her, and
Emma imagined a very sufficient income; for though the vicarage of Highbury was
not large, he was known to h
CHAPTER I

Emma Woodhouse, handsome, clever, and rich, with a comfortable home and happy di
sposition, seemed to unite some of the best society in the place. A line from me
would bring you a little hos
CHAPTER I

Emma Woodhouse, handsome, clever, and rich, with a comfortable home and happy di
sposition, seemed to unite some of the best society in the place. A line from me
would bring you a little hos
CHAPTER I

Emma Woodhouse, handsome, clever, and rich, with a comfortable home and happy di
sposition, seemed to unite some of the best blessings of existence; and had live
d nearly twenty-one years in
CHAPTER I

Emma Woodhouse, handsome, clever, and rich, with a comfortable home for her, and
Emma imagined a very sufficient income; for though the vicarage of Highbury was
not large, he was known to h
```

```
CHAPTER I

Emma Woodhouse, handsome, clever, and rich, with a comfortable home and happy di
sposition, seemed to unite some of the best blessings of existence; and had live
d nearly twenty-one years in
CHAPTER I

Emma Woodhouse, handsome, clever, and rich, with a comfortable home for her, and
Emma imagined a very sufficient income; for though the vicarage of Highbury was
not large, he was known to h
CHAPTER I

Emma Woodhouse, handsome, clever, and rich, with a comfortable home and happy di
sposition, seemed to unite some of the best blessings of existence; what could b
e increasing Emma's dislike.
26.7 s ± 853 ms per loop (mean ± std. dev. of 7 runs, 1 loop each)
n = 19


CHAPTER I

Emma Woodhouse, handsome, clever, and rich, with a comfortable home and happy di
sposition, seemed to unite some of the best blessings of existence; what could b
e increasing Emma's dislike.
CHAPTER I

Emma Woodhouse, handsome, clever, and rich, with a comfortable home for her, and
Emma imagined a very sufficient income; for though the vicarage of Highbury was
not large, he was known to h
CHAPTER I

Emma Woodhouse, handsome, clever, and rich, with a comfortable home for her, and
Emma imagined a very sufficient income; for though the vicarage of Highbury was
not large, he was known to h
CHAPTER I

Emma Woodhouse, handsome, clever, and rich, with a comfortable home for her, and
Emma imagined a very sufficient income; for though the vicarage of Highbury was
not large, he was known to h
CHAPTER I

Emma Woodhouse, handsome, clever, and rich, with a comfortable home for her, and
Emma imagined a very sufficient income; for though the vicarage of Highbury was
not large, he was known to h
CHAPTER I

Emma Woodhouse, handsome, clever, and rich, with a comfortable home and happy di
sposition, seemed to unite some of the best blessings of existence; and had live
d nearly twenty-one years in
CHAPTER I

Emma Woodhouse, handsome, clever, and rich, with a comfortable home for her, and
Emma imagined a very sufficient income; for though the vicarage of Highbury was
not large, he was known to h
CHAPTER I

Emma Woodhouse, handsome, clever, and rich, with a comfortable home and happy di
sposition, seemed to unite some of the best blessings of existence; what could b
e increasing Emma's wretchedn
27 s ± 391 ms per loop (mean ± std. dev. of 7 runs, 1 loop each)
```

Out[7]: 'Emma Woodhouse, as you are where they had quitted the house again exceedingly d
isappointed. Mrs. Weston, "if it be not inconvenient for even half a day.\n\n"I
told you he would think well of the little '

Observation

- As the number of n goes up from 2 to 20, the times for each run generating a fake-text of length 200 increase almost linearly with the value of n.

- As for the accuracy of the fake-text, there is a general trend of having more accurate speelings of words, complete sentence structures, correct punctuations, and even some local word-to-word coherence as the value of n goes up. (Although there isn't really a good quantifier defined for the 'accuracy' of the generated fake-text.)

- The sentences are perhaps syntactically correct, but they do not necessarily make any sense.

- As the value of n goes even higher, it starts to resemble the original text perhaps a little bit too much. Reasonable, because we are matching set phrases now with the larger length of n, and phrase like "a lingering illness" probably only show up once or twice in the entire book. Then, it is just copy-text being created instead of fake-text and the same lines repeat.

```
In [ ]:
#extra credit I: using regex to match words.
txt = '''
Holden Caulfield, a depressed 17-year-old, lives in a sanitorium in California a

Holden recalls the events of a single night during the previous year's Christmas

Back at his dorm, Holden's dorm neighbor, Robert Ackley, who is unpopular among

In a taxicab, Holden asks the driver whether the ducks in the Central Park lagoo

The next morning, Holden, becoming increasingly depressed and in need of persona

Holden decides to meet his old classmate, Carl Luce, for drinks at the Wicker Ba

When his parents return home, Holden slips out and visits his former and much-ad

Losing hope of finding belonging or companionship in the city, Holden impulsivel

Holden finally alludes to encountering his parents that night and "getting sick"
'''
```

```
In [ ]:
#modified markov

def markov_word2(s, n, length, seed = "Emma Woodhouse"):
    """"Generate fake text according to an n-th order Markov model, with data fro
    Args:
        s: str. the text from which to learn grams.
        n: positive int. the order of the Markov model.
        length: positive int. the number of synthetic characters to generate. sh
        seed: str. should have a default value. input the sequence of words sepa
    Returns:
        The output string fake_text. fake_text starts with the seed.
        length of fake_text = length of seed + argument 'length'
    """
    #calculate n+1 word dictionary
```

```python
        word_dict = count_nwords(s,n+1)
        seedl = re.findall(r'\w+|[,.!?:;]+', seed)
        #check validity of seed
        if (len(seedl) < n):
            seed = random.choice(list(word_dict.values()))
        #initialize fake_text as a list of words
        fake_text = re.findall(r'\w+|[,.!?:;]+', seed)
        while len(fake_text) < length:
            for key,val in word_dict.items():
                keyl = re.findall(r'\w+|[,.!?:;]+', key)
                if keyl[0:n] == fake_text[-n:]:
                    fake_text.append(random.choice(keyl))
        return fake_text
```

In [ ]:
```python
import re
def count_nwords(text, n):
    """Count the number of occurrences of n words consecutively in text.
        consider punctuations case by case
        .,:;?! ... one-sided, end of sentence/phrase marker, considered to be in
        ()'", will just ignore them lol
    Args:
        text: input text, in string format.
        n: positive int. should have default value 1.
    Returns:
        a dict keyed by n-words whose values are the number of a sequence of wor
    """
    d = {}
    #match pattern per function descriptionr"[\w'\"]+|[,.!?]
    wlist = re.findall(r"(\w+)|[,.!?:;]", text)
    for word in wlist:
        for i in range(len(wlist)-n+1):
            #convert lists of words into space separated strings
            key = " ".join(wlist[i:i+n])
            #already exist
            if key in d:
                d.update({key:d[key]+1})
            else:
                d.update({key:1})
    return d
print(count_nwords(txt,2))
```

In [ ]:
```python
import random
print(markov_word2(txt, 2, 200,seed = "Holden Caulfield"))
```

extra credit II: heuristic see prob. 5 thoughts

Problem 5

Try running your program with a different text!

You can

- find any movie script from https://imsdb.com/ or a book by Shakespeare, Hemingway, Beowulf, O.Henry, A.A. Milne, etc from https://www.gutenberg.org/

- ctrl + a to select all text on the page
- copy paste into a new .txt file. let's call it puppycat.txt.
- put puppycat.txt in the same folder as emma-full.txt.
- run the following code to read the file into variable `s` .

  ```python
  with open('puppycat.txt', 'r', encoding='utf8') as f:
      s = f.read()
  ```

- run `markov_text` on `s` with appropriate parameters.

Show your output here. Which parameters did you pick and why? Do you see any difference from when you ran the program with Emma? How so?

```python
In [11]:
# running with the great gatsby
with open('puppycat.txt', 'r', encoding='utf8') as f:
    s = f.read()
print("done loading")
```

```
done loading
```

```python
In [13]:
print(markov_text(s, 11, 200, " Mr. Gatsby"))
```

```
 Mr. Gatsby identified himself by jumping in front of
him. And ever since. His name was Jay Gatsby," I read, "For Valour Extraordinar
y, but he didn't realize it was theirs for a few
words in the right
```

write your observations and thoughts here

- I pick n = 11 because the seed phrase is of that length and it is repeated 22 times in the book. Feeling like less likely to start copying off a set sentence using it
- I noticed that because I was copying it from the HTML text, the \n gets copied and reflected in the generated fake-text as well.
- general thought, if matching character or word in a strictly markov way, then the choices are very limited and the generated sentences are more likely to be gibberish or copy of the original work, instead of creating sensible text. Say for the text:

At that hour of dawn **Agilulf** always needed to apply himself to some precise exercise: counting objects, arranging them in geometric patterns, resolving problems of arithmetic. It was the hour in which objects lose the consistency of shadow that accompanies them during the night and gradually reacquire colors, but seem to cross meanwhile an uncertain limbo, faintly touched, just breathed on by light; the hour in which one is least certain of the world's existence. He, **Agilulf**, always needed to feel himself facing things as if they were a massive wall against which he could pit the tension of his will, for only in this way did he manage to keep a sure consciousness of himself. But if the world around was instead melting into the vague and ambiguous, he would feel himself drowning in that morbid half light, incapable of allowing any clear thought or decision to flower in that void. In such moments he felt sick, faint; sometimes only at the cost of extreme effort did he feel himself able to avoid melting away completely. It was then he began to count: trees, leaves, stones, laces, pine cones, anything in front of him. Or he put them in rows and arranged them in squares and pyramids. Applying himself to this exact occupation helped

him overcome his malaise, absorb his discontent and disquiet, reacquire his usual lucidity and composure.

- Agilulf will must likely get matched with always needed. and this pattern really is persistent through out the book because Agilulf is a nonexistent knight with a set pattern of life, and this book is about how we need to have more than just rules, super-ego, regimes, and all that yidi yada.
- Then, although 'always needed' suits his character, it doesn't really help matching/generating sentences that are not already written in the book. some words that never appear in vincinity of agilulf, or has a very low chain possibility of being put anywhere in the end text using the seed "Agilulf", could also be a very reasonable match.
- so some of the heuristics are, smoothing/discounting (roughly, increasing weights of non-matches, decreasing weights of most-matches), or taking into account of words that could be novel continuations (depending on how often they appear in differnet contexts).

In [ ]: