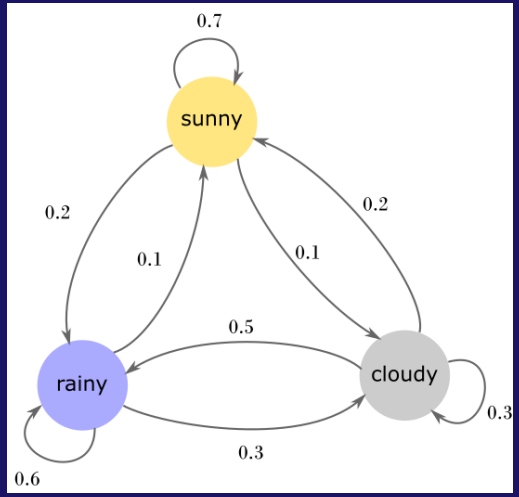
Markov Chain Text Generator

AP Computer Science Principles

Three-state Markov Chain / Markov State diagram



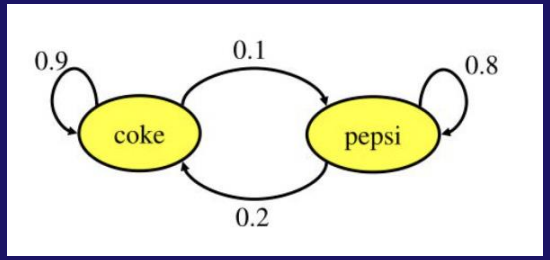
DO NOW

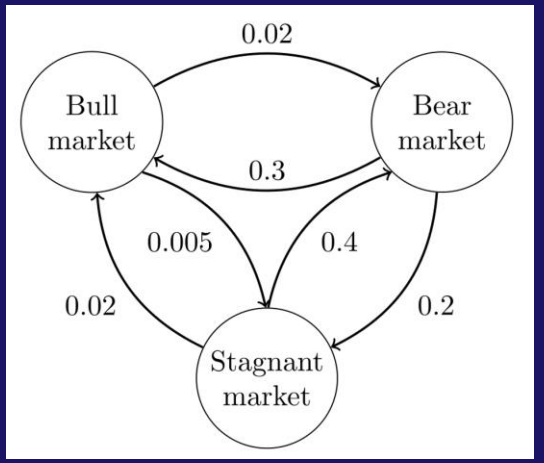
Analyze the given diagram.

Given that the following statements are true, determine more true statements:

- If it is sunny today, there is a 10% chance it'll be cloudy tomorrow.
- If it is sunny today, there is a 20% chance it'll be rainy tomorrow.

MARKOV CHAIN APPLICATIONS





- **Text Generation**
- **Financial modelling** and forecasting (including trading algorithms)
- **Logistics**: modelling future deliveries or trips
- **Search Engines**: PageRank can seen as modelling a random internet surfer with a Markov Chain

Excerpt from *Green Eggs and Ham*

WOULD YOU LIKE THEM IN A HOUSE?
WOULD YOU LIKE THEN WITH A MOUSE?

I DO NOT LIKE THEM IN A HOUSE.
I DO NOT LIKE THEM WITH A MOUSE.
I DO NOT LIKE THEM HERE OR THERE.
I DO NOT LIKE THEM ANYWHERE.
I DO NOT LIKE GREEN EGGS AND HAM.
I DO NOT LIKE THEM, SAM-I-AM.

WOULD YOU EAT THEM IN A BOX?
WOULD YOU EAT THEM WITH A FOX?

NOT IN A BOX. NOT WITH A FOX.
NOT IN A HOUSE. NOT WITH A MOUSE.
I WOULD NOT EAT THEM HERE OR THERE.
I WOULD NOT EAT THEM ANYWHERE.
I WOULD NOT EAT GREEN EGGS AND HAM.
I DO NOT LIKE THEM, SAM-I-AM.

Let's predict what words are most likely to follow "WOULD" by observing ***what words follow*** WOULD in the text and ***how often***.

Excerpt from *Green Eggs and Ham*

WOULD YOU LIKE THEM IN A HOUSE?
WOULD YOU LIKE THEN WITH A MOUSE?

I DO NOT LIKE THEM IN A HOUSE.
I DO NOT LIKE THEM WITH A MOUSE.
I DO NOT LIKE THEM HERE OR THERE.
I DO NOT LIKE THEM ANYWHERE.
I DO NOT LIKE GREEN EGGS AND HAM.
I DO NOT LIKE THEM, SAM-I-AM.

WOULD YOU EAT THEM IN A BOX?
WOULD YOU EAT THEM WITH A FOX?

NOT IN A BOX. NOT WITH A FOX.
NOT IN A HOUSE. NOT WITH A MOUSE.
I WOULD NOT EAT THEM HERE OR THERE.
I WOULD NOT EAT THEM ANYWHERE.
I WOULD NOT EAT GREEN EGGS AND HAM.
I DO NOT LIKE THEM, SAM-I-AM.

Let's predict what words are most likely to follow "WOULD" by observing ***what words follow*** WOULD in the text and ***how often***.

WOULD appears 7 times.
YOU follows 4/7, or 57%, of the time.
NOT follows 3/7, of 43%, of the time.

TEXT GENERATION

WOULD appears 7 times.
YOU follows 4/7, or 57%, of the time.
NOT follows 3/7, of 43%, of the time.

So, most likely after WOULD, the text generator will follow with YOU.

In other words, given one-word prefix WOULD, suffix word YOU is most likely to occur.

We are finding patterns and applying them.

Excerpt from *Green Eggs and Ham*

WOULD YOU LIKE THEM IN A HOUSE?
WOULD YOU LIKE THEN WITH A MOUSE?

I DO NOT LIKE THEM IN A HOUSE.
I DO NOT LIKE THEM WITH A MOUSE.
I DO NOT LIKE THEM HERE OR THERE.
I DO NOT LIKE THEM ANYWHERE.
I DO NOT LIKE GREEN EGGS AND HAM.
I DO NOT LIKE THEM, SAM-I-AM.

WOULD YOU EAT THEM IN A BOX?
WOULD YOU EAT THEM WITH A FOX?

NOT IN A BOX. NOT WITH A FOX.
NOT IN A HOUSE. NOT WITH A MOUSE.
I WOULD NOT EAT THEM HERE OR THERE.
I WOULD NOT EAT THEM ANYWHERE.
I WOULD NOT EAT GREEN EGGS AND HAM.
I DO NOT LIKE THEM, SAM-I-AM.

We can also predict what suffix words are most likely to follow two-word prefixes, like WOULD YOU.

WOULD YOU appears 4 times.
LIKE follows 2/4, or 50%, of the time.
EAT follows 2/4, of 50%, of the time.

OVERVIEW

We can repeat this process for all the words in the text, and generate a new text based on the probabilities of word order:

https://projects.haykranen.nl/markov/demo/

- How will changing "order" affect the output text?
- What order number generates the most realistic output text?

ASSIGNMENT

What if we tried this with multiple Dr. Seuss texts? Will the results be more or less realistic?

With your pair programming partner, select your favorite...

- Text from an author
- Poems from a poet
- Songs from an artist
- Series of tweets (you may have to manually scrape off Twitter)

and generate some realistic machine-generated output. Use an amalgamation of sources for input data.

Be prepared to share!

Example:

Generated with *Green Eggs and Ham*, order 4:

YOU DO NOT LIKE THEM WITH A GOAT.
I WOULD EAT THE RAIN.
NOT IN A HOUSE. NOT IN A HOUSE.
I DO NOT LIKE THEM IN A CAR?
EAT THEM IN A BOX.
I DO NOT WITH A MOUSE.
I WILL SEE.
I DO NOT LIKE GREEN EGGS AND HAM.
I DO NOT EAT THEM IN A BOX.
I WOULD YOU LIKE THEY ARE SO GOOD, YOU SAY.
TRY THEM IN A BOX.
AND I WILL EAT THEM HERE OR THEM, SAM-I-AM.
I DO NOT LIKE THEM HERE OR THEM IN A HOUSE.
I WILL NOT, WITH A FOX.
NOT LIKE THERE.
SAY! I WILL SEE.
NOT IN A CAR. AND THEM! EAT THEM IN A HOUSE.
AND I WILL EAT THEM, IN THE DARK!
WOULD YOU LIKE THEM, SAM-I-AM.