

PART 01

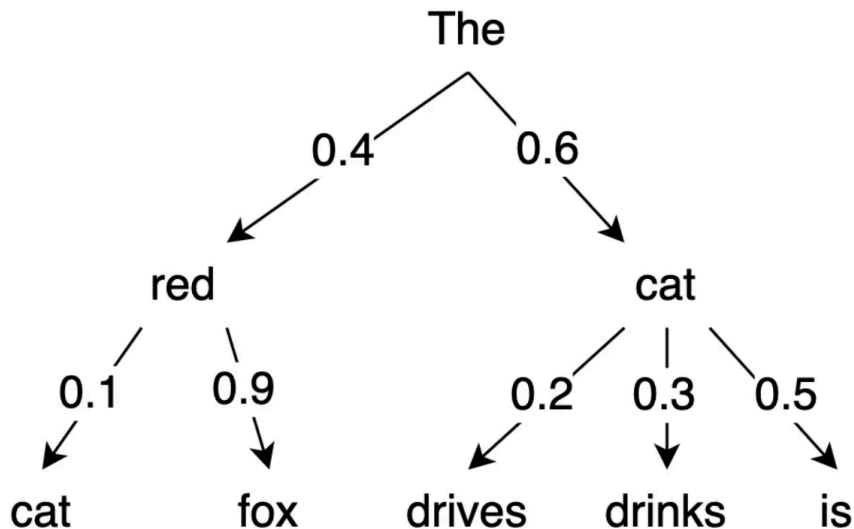
Prompt Engineering

Temperature

$$p(x_i) = \frac{e^{\frac{x_i}{T}}}{\sum_{j=1}^V e^{\frac{x_j}{T}}}$$

What happens when T approach 0? How about infinity?

Top-p



How does the following work on the left example:

1. Greedy search
2. beam search with $n_beam=2$
3. Top-k sampling
4. Top-p sampling

Max Length, Stop Sequences

Predict common sense results of the following actions.

==

Action: I didn't water the plant for 3 weeks.

Result: The plant died.

==

Action: I went to school.

Result: I got a diploma.

==

Action: I left the AC on all day.

Result: I got a high utility bill.

==

Action: I helped my neighbors when their car broke down.

Result: My neighbors were grateful.

==

Action: I put the ice cream outside for an hour.

Result: **The ice cream melted.**

==

Frequency Penalty

“Write a poem where every word starts with Z”

Frequency Penalty = 0	Frequency Penalty = 2
Zebras zigzagging zealously, Zephyrs zipping, zesty, zestfully. Zodiac’s zenith, zeal’s zodiac, Zinnias zigzag, zircon’s zodiac. Zeppelin zooming, zigzag	Zealous zephyrs zoom, zigzagging zestily, Zinnia zones, zenith zeppelins’ zone precisely. Zenith's ziggurats zealously zap, Zirconium zebras zip-zap on Zanz

Basic Prompt

Prompt

The sky is

Output:

blue



The sky is blue on a clear day. On a cloudy day, the sky may be gray or white.

Prompt:

Complete the sentence:
The sky is

Output:

so beautiful today.

Prompt Formatting

```
This is awesome! // Positive  
This is bad! // Negative  
Wow that movie was rad! // Positive  
What a horrible show! //
```

```
<Question>?  
<Answer>  
<Question>?  
<Answer>  
<Question>?  
<Answer>  
<Question>?
```

Elements of Prompts

Instruction - a specific task or instruction you want the model to perform

Context - external information or additional context that can steer the model to better responses

Input Data - the input or question that we are interested to find a response for

Output Indicator - the type or format of the output.

PART 02

Huffman Coding

Compression

Quality = 100



Quality = 50



Quality = 10



Quality = 5



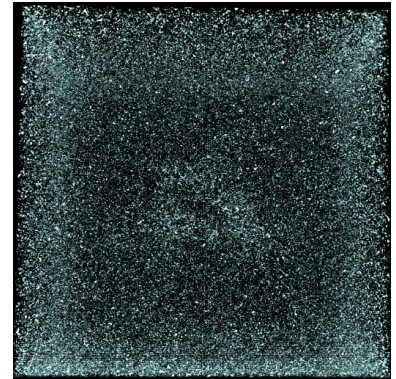
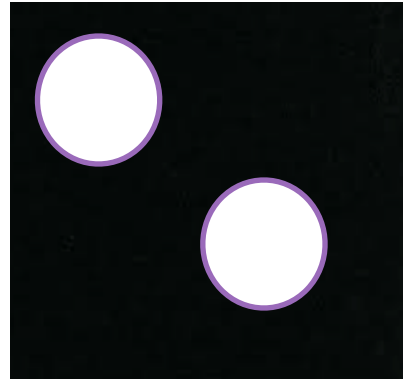
Less Compression



More Compression

Which is harder to compress?

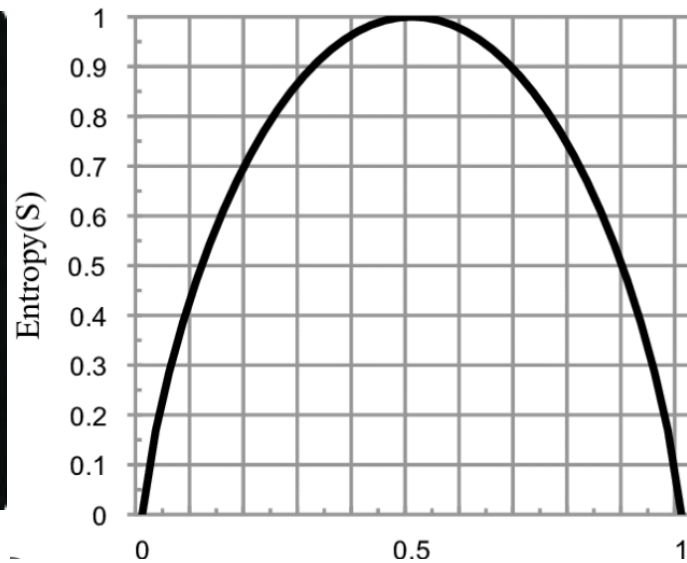
Quality = 100



Less Compression

Entropy

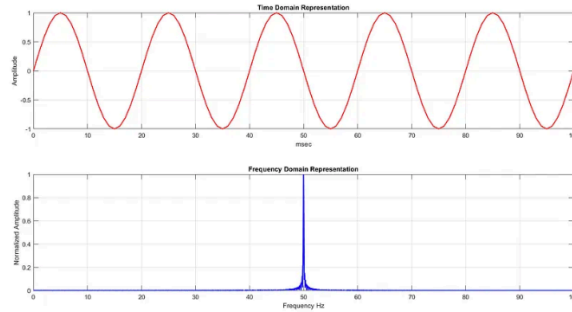
$$H = - \sum p(x) \log p(x)$$



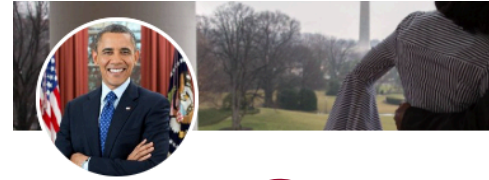
Redundancy



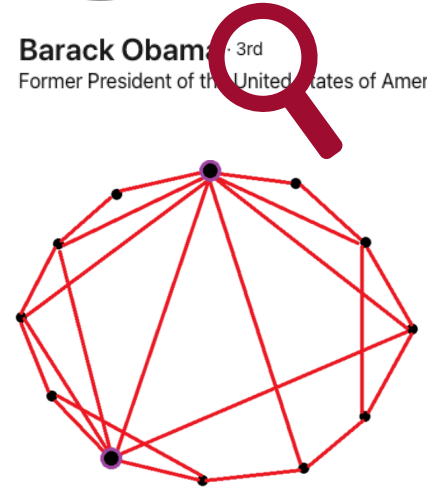
Smoothness/locality prior for image



Periodic prior for sound



Barack Obama, 43rd
Former President of the United States of America



Small-world / six-degree
prior for social networks

Redundancy in English Text & Entropy of English

Example Rules:

1. i before e except after c.

Hippie, Fries, Field. cake

2. q must always be followed by a u

Quick, quiche, question, quarrel

3. grammar

Cannot do “subject subject” as a sentence

4. dictionary

Hufamomina is not a word

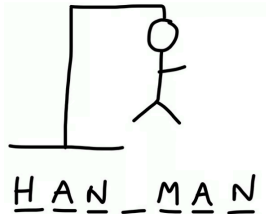
Shannon tried calculating English’s entropy using n-gram $p(x_{\{n\}} | x_{\{<n\}})$. Is this a good approach? O complexity?

Another approach: take a set of English words (8000), calculate the word-level entropy based on the subset. Then divide by average number of characters to get Character-level entropy.

Maximum entropy: ~4.7 bit/letter

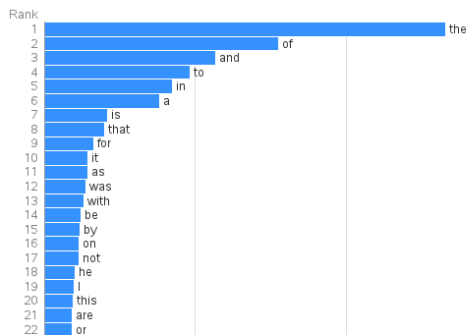
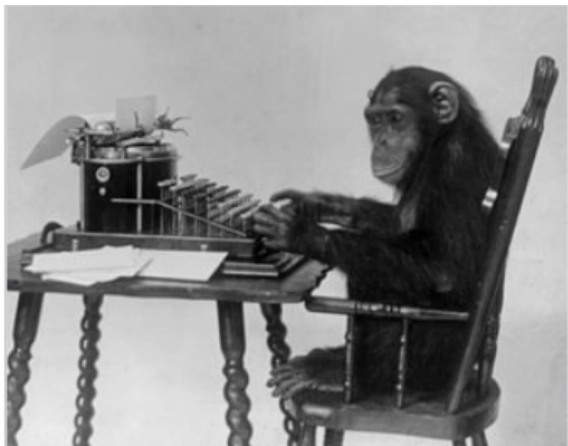
Approximated entropy of English: 2.63 bit/letter (why?)

Is English redundant? Is it good or bad?



Infinite Monkey Theorem

Almost surely, he would type up Shakespeare.



Zero-order approximation	XFOML RXKHRJFFJUJ ALPWXFWJXYJ FFJEYVJCQSGHYD QPAAMKBZAACIBZLKJQD
First-order approximation	OCRO HLO RGWR NMIELWIS EU LL NBNSEBYA TH EEI ALHENHTTPA OOBTTVA NAH BRL
Second-order approximation	ON IE ANTSOUTINYS ARE T INCTORE ST BE S DEAMY ACHIN D ILONASIVE TUCOOWE AT TEASONARE FUSO TIZIN ANDY TOBE SEACE CTISBE
Third-order approximation	IN NO IST LAT WHEY CRATICT FROURE BIRS GROCID PONDENOME OF DEMONSTURES OF THE REPTAGIN IS REGOACTIONA OF CRE
First-order word approximation	REPRESENTING AND SPEEDILY IS AN GOOD APT OR COME CAN DIFFERENT NATURAL HERE HE THE A IN CAME THE TO OF TO EXPERT GRAY COME TO FURNISHES THE LINE MESSAGE HAD BE THESE
Second-order word approximation	THE HEAD AND IN FRONTAL ATTACK ON AN ENGLISH WRITER THAT THE CHARACTER OF THIS POINT IS THEREFORE ANOTHER METHOD FOR THE LETTERS THAT THE TIME OF WHO EVER TOLD THE PROBLEM FOR AN UNEXPECTED

Compression relies on redundancy

Fixed length v.s. variable length encoding

	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>
Freq in '000s	45	13	12	16	9	5
a fixed-length	000	001	010	011	100	101
a variable-length	0	101	100	111	1101	1100

Which scheme uses fewer bits to encode the corpus?

How do we encode “bad”?

How do we decode 11000101? (Is the decoding unique? Why?)

Designing Unique Prefixes with Prefix Tree

Idea:

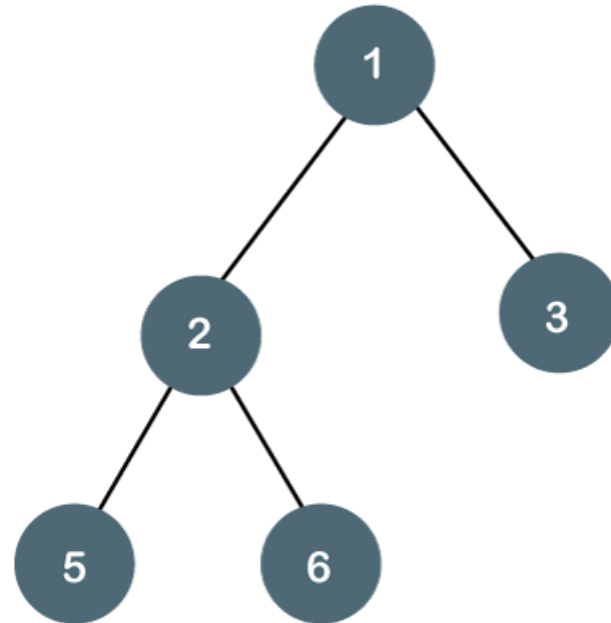
Greedy bottom-up construction of tree
Read encoding based on path from root to leaves.

Why it works:

Each token traces the path of a leaf.
A leaf has no children.
So each token has unique prefix.

Why it's efficient:

Greedy algorithm always look for least frequent token
Less frequent tokens become leaves earlier
Less frequent tokens -> longer path -> longer code



Walkthrough

$a/20, b/15, c/5, d/15, e/45$

1. Pick two least frequent words
2. Use them as leaves of a subtree
3. Merge frequency on their common parent
4. Add common parent back to list
5. Repeat

Walkthrough

$a/20, b/15, c/5, d/15, e/45$

The tokens here (a,b,c,d,e) can be bigrams!
e.g., a = cat | the, b = on | was

1. Pick two least frequent words
2. Use them as leaves of a subtree
3. Merge frequency on their common parent
4. Add common parent back to list
5. Repeat

References

<https://medium.com/@lazyprogrammerofficial/what-is-temperature-in-nlp-lms-aa2a7212e687>

<https://medium.com/nlplanet/two-minutes-nlp-most-used-decoding-methods-for-language-models-9d44b2375612>

<https://docs.ai21.com/docs/when-the-generation-stops>

<https://towardsdatascience.com/guide-to-chatgpts-advanced-settings-top-p-frequency-penalties-temperature-and-more-b70bae848069>

<https://www.promptingguide.ai/introduction/basics>

<https://www.mathworks.com/help/images/jpeg-image-deblocking-using-deep-learning.html>

https://www.researchgate.net/figure/Top-The-CNN-trained-for-the-task-of-full-face-detection-Bottom-The-CNN-trained-for-the_fig2_308944615

https://cs.stanford.edu/people/eroberts/courses/soco/projects/1999-00/information-theory/entropy_of_english_9.html

<https://home.cse.ust.hk/faculty/golin/COMP271Sp03/Notes/MyL17.pdf>