

## CS542 (Fall 2023) Initial Self-Assessment

Due August 27, 2023

Answer all the questions below to the best of your understanding. Do not use the internet to look up answers. "I don't know" is a perfectly acceptable answer. This homework is *not* graded but will help tailor the class to your level as much as possible.

Your answers must be legible. Do not put your name anywhere on this assignment. You may print and write on this PDF and scan it, or you may answer the questions in a separate text file and submit that.

1. What is the relationship between regular expressions and finite state automata?

Hmmm, regular expressions are expressions that are used to search a document for a match. Finite state automata sounds like a robot that runs until it hits a specific state, so I guess we can say that whatever program that is takes in the regular expression and runs is a form of finite state automata? But I assume that any program that reaches a final state wouldn't be a finite state automata so I am guessing the automata needs to have a search function or something.

2. What is an *n-gram*?

I was about to say I don't know when it hit me that it may just mean an n-dimensional polygon? Otherwise, I don't know, but I imagine some pretty graphic is used alongside whatever this is.

3. Write down Bayes' Theorem.

Oh gosh... ummmmm, I don't recall but if I remember correctly enough, it is sorta like Bayesian statistics? So it is a theorem that utilizes the expected outcome based on historical data to influence its decision?

4. Define precision, recall, and F-score.

**Precision** is the measurement of how close your results are (for the same measurement) to each other. So, pretty much the smaller the stddev, between your points the better. (This is coming from my chemistry background but I think it is the same).

**Recall**, I presume is something to do with how much memory you allow an agent to store and therefor have the access to 'recall' past states.

**F-score**, some statistical measure, that is all I know. Sooooo... I don't know.

5. What is a “gradient” in machine learning?

A gradient, from my understanding is the slope between the point where a function (NN or whatnot) evaluates itself and the target. So it attempts to follow this gradient in small steps like a marble until it hits some local minimum (which could be a global). It is hard to actually explain this since I guess I just save this as a visual. In one word I would say a gradient is a slope.

6. What is hidden in a Hidden Markov Model?

I don't know. I am guessing it is some NN using a Markov Model or something but I need a refresher on Markov decision processes. I know they are good at handling randomness though.

7. What “type” is a *context-free grammar* in the Chomsky hierarchy?

I don't know.

8. What does it mean to “parse” a sentence?

Break the sentence into its base parts, and then maybe tokenize them or something?

9. Who said “You shall know a word by the company it keeps?”

(a) (Bonus for bragging rights) Who is frequently (and incorrectly) credited with the above quote?

I don't know, but I have a feeling this mayyyy have been mentioned in the first lecture. But I am not sure. And maybe, the bonus question is Mark Twain?

10. Define “neural network.”

A network of consisting of inputs, hidden layers, and outputs (the neurons). The neurons are fully connected (I presume not in every case but whatever) by weights. These weights are adjusted until a specific input activates a specific output neuron. It is a bunch on linear algebra.

11. What is a convolution in a CNN?

CNNs are NNs that are able to look at chunks of input at a time to help it identify and learn specific features that it may have otherwise been infeasible to learn in the traditional manner. This batch-like scanning can help train networks, especially in vision, to learn to identify traits of an image, like horizontal, diagonal, and vertical lines, etc.

12. What is the vanishing gradient problem?

From what I remember, the vanishing gradient problem is when your problem space is too insane or your chosen methodology is inadequate for the problem such that the gradient exponentially increases and all hope is lost that run. I feel like it isn't when the gradient hits zero because that is sort of a goal in some regards.

13. What is the curse of dimensionality?

Ooph, I remember Ross Beveridge or Chuck Anderson mentioning this but I can't recall what it means.

14. What is an "embedding"?

From what I have come to infer of the term 'embedding' is that it is when you put a program in a chip or environment where the hardware is set or something like that. Like making a ML algo that runs on a raspberry pi or something.

15. What is an "encoder"? What is a "decoder"? (in the context of neural networks/Transformers)

I have only heard the term transformer a lot but don't know about them. My guess is that a neural network can learn to encode an input and another neural network can be used to pretty much perform the opposite function and decode an encoded input.

16. What is an "autoregressive" model?

A model that continuously updates its regression curve with every new input? I am not sure. I know I should know this though.

17. According to the developers of the Transformer architecture, what is all you need?

I don't know

## **Submission Instructions**

Please submit your answers (in PDF format - printed and scanned images are OK) to the drop box on Canvas.