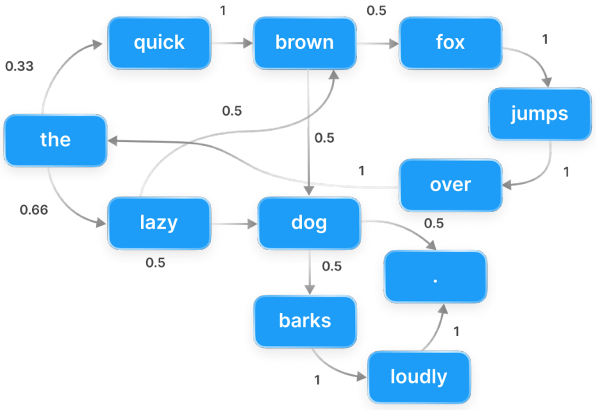
LIN 313 Language and Computers Fall 2025
Practice Midterm

1 N-gram Models

Consider the following graphical representation of a bigram language model.



1. Calculate the model probabilities of these two sentences.
   - The quick brown fox jumps over the quick brown dog.
   - The quick brown fox jumps over the lazy brown dog.
2. Could the following corpus be the training data for this model? Justify your answer

   the quick brown fox jumps over the lazy dog. the lazy dog barks loudly. the lazy brown dog barks loudly.

Part 2 Given the corpus: **"one fish two fish red fish blue fish"**

1. Build a unigram model. That is, build the table of unigram probabilities.

2. How many tokens are in the corpus?

3. How many types are in the corpus?

4. Calculate P("fish") =

5. Calculate P("red") =

6. Using the independence assumption, calculate P("fish fish red")

Part 3 Now build a bigram model for the same corpus. Add ⟨start⟩ and ⟨end⟩ tokens.

1. List all bigrams in the corpus (with start/end tokens).

2. Calculate $P(\text{“}fish''|\text{“}two'') =$

3. Calculate $P(\text{“}two''|\text{“}fish'') =$

2 Naive Bayes Classification

You have trained a sentiment classifier on movie reviews. In your training data of 1000 reviews:

- 600 reviews are positive (POS)

- 400 reviews are negative (NEG)

You've observed the following word frequencies:

Word	Count in POS	Count in NEG
amazing	120	20
terrible	15	100
movie	300	200

Calculate the following probabilities:

1. P(POS) =

2. P(NEG) =

3. $P(\text{“}amazing''|POS) =$

4. $P(\text{“}terrible''|NEG) =$

Part 2 A new review comes in: "amazing movie". Use Naive Bayes to classify it. Show all your work.

1. What is $P(POS|\text{“}amazing'', \text{“}movie'')$?

2. What is $P(NEG|\text{“}amazing'', \text{“}movie'')$?

3. How should we classify this review?

3 Ambiguity in Natural Language

Consider the sentence: **"Grandma Anita sent us a note with the computer."**

1. Provide two different interpretations (readings) of this sentence.

2. What type of ambiguity does this sentence exhibit?

3. Draw parentheses around the sentence to show the difference in constituency.
   - Grandma Anita sent us a note with the computer.
   - Grandma Anita sent us a note with the computer.

Part 2 Winograd Schemas test pronoun resolution. Consider these two sentences:

- In the storm, the tree fell down and crashed through the roof of my house. Now, I have to get it removed.
- In the storm, the tree fell down and crashed through the roof of my house. Now, I have to get it repaired.

1. In the first sentence, what does "it" refer to?

2. In the second sentence, what does "it" refer to?

3. Why is this type of ambiguity challenging for computers but easy for humans? In other words, what does a computer have to know in order to correctly solve this question?

4 Machine Learning Concepts

You are building a spam email classifier. Your test set has 1000 emails: 50 are spam, 950 are not spam. Your classifier produces the following results:

- True Positives (correctly identified spam): 40
- False Positives (incorrectly identified as spam): 20
- False Negatives (spam missed by classifier): 10
- True Negatives (correctly identified as not spam): 930

1. Calculate the accuracy of your classifier.

2. Calculate the precision. What does this tell you?

3. Calculate the recall. What does this tell you?

4. Your boss wants to make sure no important emails get filtered. Should you optimize for precision or recall? Explain.

Part 2 Classify each scenario as supervised or unsupervised learning:

1. Training an N-gram language model on Shakespeare's complete works

2. Building a sentiment classifier with 10,000 labeled tweets

3. A recommendation system that groups users with similar viewing patterns.

Conceptual Questions

1. SHRDLU was an early AI system that worked in a "blocks world."

   (a) What approach to AI did SHRDLU represent (symbolic or subsymbolic)?

   (b) Give two reasons why SHRDLU-style systems ultimately failed to scale to real-world language tasks.

   (c) What kind of mistakes do LLMs make that SHRDLU-style systems do NOT make?

2. Consider **Garbage in, Garbage out** principle in the context of training and evaluating language models. Describe two concrete examples of this principle in action in the context of training and evaluating language models. You can discuss examples from class, from your own experience, or construct a hypothetical.

3. Give an example of when the **independence assumption** for Naive Bayes fails in real language. Provide a short textual example and explain why the assumption fails.

4. The AI field experienced an "AI Winter" in the 1970s-80s. Name two reasons why early AI systems failed and funding dried up.

5 The Perceptron

You are using a perceptron to decide whether to adopt a dog from a shelter. The features are:

- x_1: Is it small? (1 = yes, 0 = no)
- x_2: Is it friendly? (1 = yes, 0 = no)
- x_3: Does it need lots of exercise? (1 = yes, 0 = no)

Your weight vector is: $\mathbf{w} = [3, 5, -2]$
Your bias is: $b = -4$
The activation function is: $\text{sign}(z) = \begin{cases} 1 & \text{if } z > 0 \\ 0 & \text{otherwise} \end{cases}$

Part 1 You meet a dog with features $\mathbf{x} = [1, 1, 1]$ (small, friendly, needs lots of exercise).

1. Calculate the weighted sum: $\mathbf{w} \cdot \mathbf{x} + b$

2. What does the perceptron output? Should you adopt this dog?

Part 2 You meet another dog with features $\mathbf{x} = [0, 1, 0]$ (large, friendly, doesn't need much exercise).

1. Calculate the weighted sum and output.

2. Based on your weights, which feature is most important to your decision? How can you tell?

6 Language

List and briefly explain how each of the following facts exemplifies one of these properties of human language: **arbitrariness, systematicity, recursion, productivity**.

1. The word for "thank you" in Arabic is "shukran".

2. The word "google-able" (as in, "This question is easily googleable.")

3. You say po-tay-to, I say po-tah-to.

4. It's possible to make a sentence that is infinitely long.

7 Speech

Roman Jakobson identified six functions of language. For each utterance below, identify the PRIMARY function being used. Choose from: **referential, emotive, conative, phatic, metalingual, poetic**.

1. "Wow! That's an incredible sunset!"

   Function: _____

   Explanation:

2. “Hey, are you still there? Can you hear me?”
   Function: _____
   Explanation:

3. “The meeting starts at 3 PM in Room 204.”
   Function: _____
   Explanation:

4. “Close the door, please.”
   Function: _____
   Explanation:

5. “What I mean by ‘algorithm’ is a step-by-step procedure for solving a problem.”
   Function: _____
   Explanation:

6. A commercial slogan: “Snap, Crackle, Pop!”
   Function: _____
   Explanation:

8 Conversation

Paul Grice proposed four maxims that guide cooperative conversation: **Quality, Quantity, Relation (Relevance), and Manner**. When a speaker obviously flouts one of the maxims, their utterance often has the effect of implicating something left unsaid directly, that the listener must infer.

For each scenario, identify which maxim is being violated or flouted, and explain the effect.

1. Your roommate asks: "Do you know where my keys are?" You respond: "They're either in your room, the kitchen, the living room, the bathroom, or possibly in your backpack, jacket pocket, or maybe you left them at the library, or perhaps they fell under the couch..."

   Which maxim? _____

   Explanation:

2. A professor writes a recommendation letter for a philosophy PhD program that says only: "Ms. Johnson attended all her classes and turned in assignments on time."

   Which maxim? _____

   Explanation:

3. Someone asks: "How was your date last night?" You reply: "Well, the restaurant had nice lighting."

   Which maxim? _____

   Explanation:

4. Your clumsy friend trips and falls. You say: "Wow, you're so graceful!"

   Which maxim? _____

   Explanation: