

1. Explain precision and recall
2. What is perplexity? What is its place in NLP?
3. Explain the vanishing gradients problem.
4. How do you decide on batch size? What are some of the trade-offs to consider?
5. BPE encoding example
6. Compare and contrast Naive Bayes and regression classifiers. What factors would lead you to use one over the other?
7. A historian is interested in grouping a set of documents by theme. They have 200 documents, and they have hand-classified a dozen into three groups, but would like to automate the process for the rest of the documents. Talk through how you would approach the problem. What kinds of tools and techniques would you use?
8. Give an example of a bag-of-words model.
9. Calculate the minimum string edit distance between two words.
10. You are web-scraping a modest amount of data (50,000 posts) from a knitting-pattern sharing site, in order to build a pattern recommendation program. The site supports both discussion threads and reviews of individual patterns. What kind of data would you collect, and how would you store it?
11. You are interested in studying this meme:  
  
nobody:  
literally no one:  
not a single person in the world:  
me: (x)  
  
How would you identify instances of this meme?
12. Assess the difficulty of building a program to solve crossword puzzles. What kind of model would you build? How would you format the data for the model?
13. What loss function would you use for training a neural network recipe classifier?
14. Discuss different strategies for getting word embeddings.

15. Consider a neural network with 12 standard Transformer layers. What other information would you need from me to estimate the number of weights in the model?
16. Explain the difference between cross-attention and masked self-attention
17. If you had two language models, how would you decide which one is better?
18. What are some strategies you might use to pick hyperparameters?