1. **Computational Linguistics**

Consider the famous sentence, "The quick brown fox jumps over the lazy dog."

Draw a reasonable parse tree for the sentence (assuming the existence of reasonable grammar rules). The root of the tree should be S, representing a sentence, and the leaves should be the words of the sentence.

Also express the CFG rules, including the lexical rules, that are implied by the tree.

1. Statistical NLP (conventional and modern)

*Briefly answer* the following questions (with *one or two phrases or sentences for each question*) related to statistical natural language processing.

1. Naïve Bayes systems work well for some text categorization tasks, even though the "naïve" assumption is clearly false. Explain what it means for the assumption to be false for this task, and give a specific example that demonstrates it is false.
2. Consider a conventional, feedforward neural network applied to the task of text categorization, and one sentence is being classified at a time. Assume it has been trained on a corpus with D labeled sentences, and the total size of the vocabulary is V. It is now being used to classify a document with T total tokens and U unique, or distinct, tokens. If a conventional feedforward neural network is being used for the task, what would typically be the number of input nodes? What would be represented by each input node?
3. Now consider text categorization involving d-dimensional word embeddings and a recurrent neural network (either a simple RNN, or a variation such as an LSTM). We have learned that it shouldn't be necessary to pad sentences to ensure they have equal length. When using other types of deep neural networks with word embeddings (such as a feedforward neural networks or a CNN), it typically is necessary to pad the input sentence. Why isn't it generally necessary to pad sentences when using an RNN for text categorization?
4. Now consider a hidden Markov model being used for part-of-speech (POS) tagging. If the tagger is trained using a treebank (a corpus containing labeled examples of POS), what parameters need to be learned?
5. Now consider a simple RNN being used as a POS tagger (in practice, a variation such as an LSTM would more likely be used). If the tagger is trained using a treebank, what parameters need to be learned?