CS685 Quiz 7: *probe tasks*
Released 11/18, due 11/23 on Gradescope (please upload a PDF!)
Please answer both questions in 2-4 sentences each.

1. **Let's say we want to design a probe task to measure whether BERT's token-level representations encode information about whether or not a particular token is a named entity. Why do we have to freeze BERT's representations during the training of our probe network?**
   While training a probe network, we are interpreting the BERT representations and not actually improving the tokens for their original task objective. Hence during training of the probe network we freeze the BERT representations, so that they are not modified and we can learn their interpretation. The only parameters which will be updated are the weights of the classifier which is classifying the token as named entity representation or not.

2. **Now let's say we want to probe whether or not BERT's [CLS] token has encoded the length of an input sentence. Explain how you would design a control task for this probe to address the effect of probe network complexity.**
   As a control task for this probe, we could map random numbers as sentence length to each input sentence before training the probe. We would like the probe network to have high selectivity - i.e. high linguistic accuracy but low control probe accuracy. If our control probe predicts the control task with high accuracy, we will have low selectivity indicating that the model is too complex and is memorizing the control task rather than depending on linguistic properties to solve the probe task.