1.
    a) First create a list of every word in the document and decide the presence of each word in the list. If a word is present, represent with a “1” and if a word is absent represent it with a “0”
    b)
        i) Since there are almost 200,000 words in the English language not including slang, this representation creates a list of almost 200,000 dimensions which is impractical. Each document can not practically be that large.
2. 0 and 1 in this RNN are upper and lower bounds formed from the sigmoid function. The network is computing a linear function and attempting to classify an input as a 0 or a 1. The output resulting from an execution of this RNN would be a confidence level. For example, an output of 0.5 would mean the RNN is simply making a guess, 50/50 between the 0 or 1. An output of 0.08 would mean the RNN is confident the classification is 0 while an output of 0.89 would mean the RNN is confident in a classification of 1.
3.
4. In the context of Recurrent Neural Networks, regularization can be used to help prevent overfitting and thus increase accuracy when the RNN is put to use.
5. Teacher forcing is more accurate than the model output because instead of continuing with the model's errors made in a previous step, the model uses a ground truth as an input. This allows models to learn from errors it’s making in real time, making it stay close to the ground-truth sequence. Also, since loss is computed at every node, you don’t have to wait for a final output to train. Training can be performed in parallel at every node output.