

CS685 Quiz 1: *Language models*

Released 2/12, due 2/19 on Gradescope (please upload a PDF!)

Please answer both questions in ~1 paragraph each. Make sure to also fill out the AI disclosure!

1. You are given two language models trained on Wikipedia. One is an unsmoothed 5-gram model (i.e., prefixes are four tokens long), while the other is a fixed-window neural language model with an identical prefix size. Which model's estimate of the conditional probability distribution $P(w \mid \text{"chalkboards flap their wings"})$ is likely to be more reasonable and why?

Ans: The fixed window neural language model with an identical prefix size is more likely to give a reasonable estimate of the given conditional probability distribution because "chalkboards flap their wings" seems to be a unique set of words which might not be present in the training data set and the unsmoothed 5-gram model might assign a probability of 0 to many words following the prefix, even though there might be some words following the prefix and hence leading to data sparsity. The neural language model handles this better through smoothing techniques and learns the complexity between the words in the training data set which prevents assigning zero probabilities to words not seen in the training set.

2. You have two datasets that each contain 100K documents. The first contains only English text, while the second contains only Korean text. You train two recurrent language models on each dataset, setting aside 5K examples from each as test sets. The two models have the same architecture and identical numbers of learned parameters. You observe that the English LM achieves a perplexity of 12 on its test set, while the Korean LM achieves a perplexity of 18 on its test set. From these results, is it valid to conclude that your model architecture is better suited for English than Korean? Why or why not?

Ans: Even though minimum perplexity means better probability, it is not valid to conclude that the model architecture is better suited for English than Korean because there may be various other factors that affect this. There might be a difference in the dataset in terms of the quality, like how clean the data is, diversity of the dataset, etc leading to the difference in perplexity. The Korean language in itself could be more complex leading to a higher perplexity. We also don't know if the hyperparameters used are similar or not. Therefore, it might not be valid to determine the model architecture suits best for English only with the perplexity.

AI Disclosure

AI1: Did you use any AI assistance to complete this homework? If so, please also specify what AI you used.

Your answer here

Yes. ChatGPT

(only complete the below questions if you answered yes above)

AI2: If you used a large language model to assist you, please paste **all** of the prompts that you used below. Add a separate bullet for each prompt, and specify which problem is associated with which prompt.

- what factors affect perplexity and model architecture in recurrent language models? (Q2)

AI3: (Free response) For each problem for which you used assistance, describe your overall experience with the AI. How helpful was it? Did it just directly give you a good answer, or did you have to edit it? Was its output ever obviously wrong or irrelevant? Did you use it to get the answer or check your own answer?

- The response was helpful to an extent but it gave me a lot of generalized information and I had to read through to find points useful to help me answer this particular question.