Chayan Kochar – 2019114008
NLP Assignment1

- As the LMs were used both on trained and tested data, the results justified the biasness. The perplexity of trained data would obviously be very low(as it was the data which was used to train!)
- Thus we see that avg perplexity of LMs used upon trained datasets being < 10 while those that were acted upon test dataset yielded perplexity of around 400-800
- In the health corpus, it was fairly observed that the perplexity swung a lot, like for some sentences around 50 while for some around 500 or more. I think one important reason is being the presence of unknowns like in proper nouns(maybe like name of disease) – leading to higher PP, whereas because of the repetitive nature of the text, rest sentences had lower PP.

Kneyser vs Witten Bell: Both the alogorithms have their own pros and cons. Though here I implemented the basic Kneyser-Ney algo(having absolute discount along with interpolation), still kneyser-ney outperformed witten-bell for majority of the results.

Importantly, I also felt that witten-bell outperforms Kneyser-ney in higher orders and vice versa. That is because when we see perplexity of trained data, we know it is going to be performed in the highest order(4 here), whose scores are less for witten bell, but when we see perplexity of test data, we know we might encounter unknown context, unknown words, etc because of which we may have used interpolation(hence low order).

Average Perplexity scores:
LM1-test : 402.2521714216223
LM1-train: 6.192819843864215
LM2-test: 437.9941301530257
LM2-train: 4.410316655745156
LM3-test: 252.6005053605036
LM3-train: 5.919801567504577
LM4-test: 255.12205455736816
LM4-train: 4.314931352833273