

Introduction to NLP-Assgn1

Report

Name: Aryan Jain

Roll Num.: 2019101056

Below are the Perplexity scores of each language model:

For LM1 on training data: 3.650240370362834

For LM1 on test data: 71.80724468334694

For LM2 on training data: 2.8002774347447263

For LM2 on test data: 75.2870179571669

For LM3 on training data: 3.3459745048912746

For LM3 on test data: 86.23235262305535

For LM4 on training data: 2.4959502258768103

For LM4 on test data: 93.27244853793543

→ *Kneyser-Ney smoothing seemed to work better than Witten-Bell smoothing on test set, whereas vice versa happened in training set.*

→ *Perplexity score is better on training data than on test data for both the algorithm.*

Below are the graphical representation of the Perplexity values on training and testing data repectively.

