

Assignment 2: language modelling

In this assignment, you will construct several language models and test them. Look at the provided notebooks for language model lectures.

There are four accompanying text files of books published by the famous Chinua Achebe. You are expected to concatenate these books into a dataset for training a language model.

Create a train, test and validation datasets from the dataset. Using the training set, build a language model for generating English text and evaluate your model using perplexity and BLEU (Bilingual Evaluation Understudy) metric.

Calculate the perplexity and BLUE of your test set. Use your language model to generate sentences by randomly sampling from the models.