

COSC 410LA S25 Lab 5: Character-Level Language Modeling

This Lab is due **Friday, Mar 7, 2025 at 11:59PM**.

Introduction

The goal of this lab is to introduce you to fitting a character level language model to data and generating text from it. You will expand your toolkit to the use of Keras for fitting neural networks.

Here are some learning objectives for this lab:

1. Properly handle text data for language modeling
2. Train an LSTM character level language model
3. Implement two approaches to language generation

There are 3 parts to this lab. The table below highlights what you are expected to do in each part.

| Part | Description | Write code? | Write responses? |
|------|--|-------------|------------------|
| 1 | Preprocess data | Yes | No |
| 2 | Fitting model to the data | No | Yes |
| 3 | Implement two approaches to generation | Yes | Yes |

Provided Files

- `Lab5.ipynb`
- `pride-and-prejudice.txt`, the text of *Pride and Prejudice* by Jane Austen.
- `pride_model.keras`, a pretrained model

Work through `Lab5.ipynb` with your partner(s). Then submit `Lab5.ipynb` with the cell outputs saved to Gradescope.