CS 505 – Spring 2022 – Assignment 5 (100 pts) – Transformers for Language Modeling and Classification

Problems due 11:59PM EST, April 29, 2022.

In this assignment, you will learn about implementing Transformers using HuggingFace library for language modeling and text classification, and use the deep learning framework Pytorch. The assignment will ask you to follow tutorials written in Google Colab while answering some questions and modifying the tutorials. You have 2 weeks to finish this assignment.

Submission Instructions Read the submission instructions fully as they are updated.

- You will export your answers for each problem from a jupyter notebook (not Colab) instance into a pdf. After you export your answers into pdf for each problem, please merge them in order and submit a single pdf. Pdf's from Colab are not as readable. Please provide a link to your Colab at the beginning of your submission along with your name and your collaborators.
- When submitting on Gradescope, make sure you assign the correct page for a given part, otherwise the grader won't be able to see your answer.
- It is helpful to make your answers as visible and easily readable as possible. You can bold your answers and/or alter their colors.
- Please indicate names of those you collaborate with. While collaboration is allowed, only submit your own code and answers.
- Every late day will reduce your score by 20. After 2 days (i.e., if you submit on the 3rd day after due date), it will be marked 0.

Problem 1. *(60 pts)* ***Generate text from pretrained models and fine-tune models using Datasets library****.*

Make a copy of this tutorial in Google Colab in your own Drive: click File → Save a copy in Drive and answer questions and write code as directed in the tutorial.

Problem 2. *(10 pts)* ***Fine-tune models for text generation using raw text.***

Make a copy of this tutorial in Google Colab in your own Drive and answer questions and write code as directed in the tutorial.

Problem 3. *(30 points)* ***Fine-tune Transformers for text classification.***

Make a copy of this tutorial in Google Colab in your own Drive and answer questions and write code as directed in the tutorial.