ILLINOIS

IE534/CS547: Deep Learning **(Due: Mar-11-2021)**

Homework #5

Instructor: Richard B. Sowers

Read all the instructions below carefully before you start working on the assignment, and before you make a submission.

- **This is a group homework, every group only submit ONE solution on Compass .** Please include the names of all the group members.
- **Due time is at 11:59pm** at the due date. **No late submission!**
- All students are expected to abide by the Honor Code
- All date-times will be in Champaign-Urbana
- Please put your typed solution in a PDF format. For code, you should submit a google colab notebook link with viewers permission to instructors and TAs in your solution PDF file.

Problem 1. Coding Question (30 points)

Training, Testing and Validation: Two dimensional polynomials.

Consider a dataset consisting of (X, Y, Z) records. The location of the dataset is in "piazza/Resources/General Resources/HW5_data.csv". We want to fit a two-dimensional polynomial model to the data. Specifically, consider

$$Z_n = \sum_{\alpha,\beta} c_{\alpha,\beta} X_n^{\alpha} Y_n^{\beta}$$

where α and β in the sum range over all nonnegative integers such that $\alpha + \beta \leq D$ (D being the degree of the polynomial), and where D is a hyperparameter.

Below are a few steps to follow:

- Split the dataset into training, validation and testing with a ratio of 0.6:0.2:0.2 (use the first 60% for training and last 20% for testing).
- Generate polynomial features up to degree D (consider the **PolynomialFeatures** fuction in sklearn). Be careful about the intercept.
- Fit a linear regression model between Z and the polynomial features on training data.
- Given $D \in \{2, 3, 4, 5, 6, 7\}$, find the best D that minimize the mean squared error (MSE) for the validation data.
- Calculate the testing MSE with the best D.

What to submit:

- You should submit a **PDF file** with a google colab notebook link with viewers permission to instructors and TAs.
- In addition to the code for each of the previous steps, you should print the following things at the end of your code:
  - The best D (10 points).
  - The training, validation and testing error for the best D (10 points).
  - The coefficients for the best D (10 points).