ENEL 645,
Assignment 1- Linear Regression (15 %)
Due Time: 18-May-2023, 23:59
Instructor: Abbas Mahbod

Description:

One of the most important scopes of ENEL 645 (Spring Semester) is to deal with real-world ML problems. To address this, we will start our first assignment with one of the most up-to-date datasets offered by Open Calgary.

The city of Calgary's open data is a website in which various types of datasets are available, and different tools can be applied for exploring and interacting with the data. The first assignment is to apply a linear regression model to Community Crime and Disorder Statistics dataset. Our target will be to train a model which can predict the number of crimes in each community center based on different input features. Hence, you are expected to build a linear regression model with multiple input features. A reduced version of this dataset is available for this assignment.

Output Variable: Crime Count

The performance of your model is evaluated based on mean-squared-error cost function.

The performance of your model will be evaluated based on some unseen data from the same distribution.

You are expected to separate the data to different datasets; 70 % of the data for training, and 30 % of it for evaluating your work.

You should upload your codes (in *.py and .ipynb* format) as well as a description about your algorithm.

The description should contain a screenshot from the results.

Some Hints:

You can consider 'Community name' as the index column of the dataset.

You may want to use *get_dummies()* from pandas library to convert categorical variable into dummy/indicator variables.

Feel free to apply feature engineering, if you think it can be useful to achieve better results.