# PGP in AI/ML

Regression Project Part C

Submission Deadline - 23:59 07 July 2019

**Total Marks: 14**

## Objective

The goal of this part of the project is to develop a linear regression model for multiple featured dataset. The regression models should be implemented using the libraries **sklearn** and **numpy**.

## Problem Description

Compressive strength is the capacity of a material to withstand loads tending to reduce size. Given the components of concrete we attempt to predict the compressive strength using linear regression.

In the dataset file, there are 1030 data points with components of concrete as features and compressive strength as the target variable.

Train the linear regression model using the following methods

- Stochastic Gradient Descent (**sklearn.linear_model.SGDRegressor**)
- Normal Equations (**sklearn.linear_model.LinearRegression**)

## Instructions

1. Target variable for this part of the project is **Concrete compressive strength**.
2. Take all the remaining columns as features.
3. Following parameters of **SGDRegressor** need to be tuned for satisfactory results

- max_iter
- tol
- eta0

4. Every model will be evaluated using Residual Standard Error (RSE) and $R^2$ score as discussed in "**Performance measure for multiple linear regression models**" video.

5. Implement forward feature selection. Algorithm is explained in "**Forward and backward feature selection algorithms for multiple linear regression models**" video.

6. Implementation of forward feature selection is to be done using Normal Equations only and **not** SGDRegressor.

# Submission Details

1. SGD Code - **id_sgd_part_c.py**. Print the obtained weights for each feature and the RSE and $R^2$ of the model. (2+1+0.5+0.5 = 4M)

2. Normal Equations Code - **id_ne_part_c.py**. Print the obtained weights for each feature and the RSE and $R^2$ of the model. (2+1+0.5+0.5 = 4M)

3. Forward Feature Selection

   - Code - **id_ffs_part_c.py**. (4M)
   - Document - **id_ffs_analysis.docx**. This document should contain the results of your feature selection at each progressive stage, i.e. the name of the features selected and the RSE at each step. (2M)

# Contacts

You should put up questions in discussion forum of the corresponding assignment folder only.

Please put all your queries to the following TAs in Canvas but not to the instructor.

- Kartheek f20160015@hyderabad.bits-pilani.ac.in
- Agam f20160484@hyderabad.bits-pilani.ac.in

- Shristy f20160688@hyderabad.bits-pilani.ac.in
- Himanshu himanshuchawla437@gmail.com
- Mohit mohitrathore15@gmail.com