

# PGP in AI/ML

## Regression – Project – Part A

**Submission Date: 2359hrs on 27-06-2019**

**Total Marks: 8**

The goal of this part of the project is to generate a linear regression model for the given dataset by solving normal equations obtained by equating gradient to zero vector. The regression model should be built by making using of libraries such as sklearn and numpy.

### Problem Description:

Karl Pearson conducted a survey around 1903 on the height of heights of fathers and their sons. The number of data points in the survey are 1078 samples. It is interesting to see whether son's height can be predicted given father's height. Hence a simple linear regression model can be built with the given data set.

The objective of this project is to build the simple linear regression model using of libraries such as sklearn, numpy and matplotlib.

The data set is given in the attached csv file named "father\_son\_heights.csv"

An incomplete python notebook has been attached to help you to get started with the code in a structured way.

The following outcomes are expected from this part of the project.

- 1) A scatter plot of the dataset as a png file. Name this as 'scatterplot.png'.
- 2) The code that you write with the help of the attached notebook as a xxxx.py file. Name this as 'ols.py'.
- 3) A plot of the line you obtained along with the scatterplot as a png file. Name this as 'line\_and\_scatter.png'.
- 4) A text file in the following format:
  - a) Coefficient value obtained:
  - b) Intercept value obtained:

Upload this folder as a zip file with the name '2018AIML<your id>\_assignment\_partA.zip'.

You can refer to the following for additional information:

- 1) Matplotlib documentation: <https://matplotlib.org/tutorials/index.html>
- 2) Numpy documentation: <https://www.numpy.org/devdocs/user/quickstart.html>
- 3) Sklearn library documentation: [https://scikit-learn.org/stable/modules/generated/sklearn.linear\\_model.LinearRegression.html](https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html)