

PGP in AI/ML

Regression – Mini Project

Total Marks: 24

Submission Date: 2359hrs on 10-12-2019

"No Plagiarism is accepted; if detected directly 0 marks will be given without any discussion".

Objective: -

Understand dataset that contains information from Air Quality data regarding their chemical's compositions.

Question1 [8 marks]

1. Describe why a linear regression model may be appropriate to describe the relationship between **Column C “CO (GT)” & Column H “NO_x (GT)”**, by code & graphs. [4 marks]
2. Try to understand which functions from **Pandas, Numpy V/s sklearn library** which perform better than other with explanations for question 1. [4 marks]

Question2: - [16 marks]

1. Implement regression model using gradient descent algorithm to predict which days are good for living, with justifications.
 - Solve this problem using gradient descent algorithm [3 Marks]
 - Solve this problem using stochastic gradient descent algorithm [3 Marks]
 - Solve this problem using mini-batch gradient descent algorithm [3 Marks]

Hint: - PM (Fine particulate matter) level calculated based on CO, NO, NO_x and other chemicals values in air. So, lower the values of CO, NO, NO_x and other chemicals will be good fit for living.

2. Name combination of attributes which suggest good air quality. [3 Marks]
3. Give logical explanation for selection of any one approach. [4 marks]

Submissions:

Solutions or answers to all questions (excepting programming problems) should be submitted in a word document named ‘Roll_No.doc’ (Roll_No is your identity number of this programme).

All python code should be submitted in a Jupyter notebook named "Roll_No.ipynb".