

Regression Project – 3

Gradient Descent - Combined Cycle Power Plant

Problem Statement:

Gradient Descent - Combined Cycle Power Plant

[Send Feedback](#)

Combined Cycle Power Plant dataset contains 9568 data points collected from a Combined Cycle Power Plant over 6 years (2006-2011), when the power plant was set to work with full load. Features consist of hourly average ambient variables Temperature (T), Ambient Pressure (AP), Relative Humidity (RH) and Exhaust Vacuum (V) to predict the net hourly electrical energy output (EP) of the plant.

You are given:

1. A Readme file for more details on dataset.
2. A Training dataset csv file with X train and Y train data
3. A X test File and you have to predict and submit predictions for this file.

Data

test.csv 
train.csv 
readme.txt 

 Upload Predictions Predictions file should be in .csv format

 Upload Notebook Notebook file should be in .ipynb format

Note :

1. You need to upload both predictions and notebook file to submit.
2. Please make sure that your submissions are genuine and you haven't cheated/copied from anywhere. Also only use techniques taught till this stage for submissions.
3. Your batchmates and TAs can view your submissions, and report them for use of invalid means if that happens. Your score will be reduced to 0 in that case.

Your task is to:

1. Code Gradient Descent for N features and come with predictions.
2. Try and test with various combinations of learning rates and number of iterations.
3. Try using Feature Scaling, and see if it helps you in getting better results.

Read Instructions carefully -

1. Use Gradient Descent as a training algorithm and submit results predicted.
2. Files are in csv format, you can use `genfromtxt` function in numpy to load data from csv file. Similarly you can use `savetxt` function to save data into a file.
3. Submit a csv file with only predictions for X test data. File should not have any headers and should only have one column i.e. predictions. Also predictions shouldn't be in exponential form.
4. Your score is based on coefficient of determination. So it can be possible that nobody gets full score.

