

CSC 4792 – DATA MINING

PROGRAMMING EXERCISE 3//2019

You work as a computer engineer at CSC company where they manufacture electric power generators. As part of a process before rolling out the generators on the market, you test the generators to determine whether they are good or faulty. To do this, you take revolutions per minute (rpm) that the generator is running at and the amount of vibrations in the generator. Your boss, wants an IT solution that will help to predict the status (good or faulty) based on the two aforementioned measurements (rpm and vibrations). You are provided with a list of measurements from past days of generators that were determined to be good and faulty.

Complete the following tasks:

- i. Plot the data points to visualize its distributions on Cartesian coordinate.
- ii. Develop and train the logistic regression model. Instead of using the gradient descent, in this exercise you shall use the “fmin” advanced optimization technique.
(Link: <https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.fmin.html>)
- iii. What is the cost found by `fmin()` at minimum point.
- iv. What is minimum point found by `fmin()`.
- v. Write the model in form: $h(x) = w_0x_0 + w_1x_1 + w_2x_2$
- vi. What is the status of the generator with the following measurements:
RPM=810 and Vibrations=495?
- vii. Plot the decision boundary separating the good and the faulty generator.

Link to the dataset:

<https://drive.google.com/open?id=1bNNgJ8Y4Qz5stzFgISHPTerWKmM1QMqM>