Hi Idit,

Following our interview, below is the assignment:

In the attached csv file there are measurements (features) were done on malignant tissues of patients with lung cancer.

The first column (measurement0) is the severity of the illness (the higher value, the illness is more advanced). measurement1 ... measurementsN are different measurements were done on the tissues.

Each raw represents a patient, from patient0 ... patientN.

1. Develop a model that can predict the severity of the illness (measurement0) for patients from other measurements in the data.
2. The model should be generalized to a test set.
3. Print the selected features of the model to a file.
4. Partition the patient to two groups with respect to the severity of the illness: low and high severity.
5. Repeat 1-3 again, except that now you should build a classifier.
6. How many features appear both in the regression and classification model?

Do the assignment in Python 3.

Good luck,

Ido