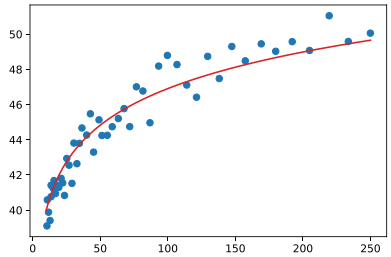
The first part of the project consists of two small tasks. You can find a `.csv` file for each one in the `warm-up.zip` archive from the resource section. Write the solution for all the tasks in a `warm-up-solution.ipynb` notebook that should run with the `exts-ml` course environment.

Task 1

The first task consists of fitting the following equation to a set of 50 x/y data points.

$$y = a * \log(x) + b$$

Here is a plot of the data points with the desired model curve.



Your curve should be optimal with respect to the residual sum of squares metric (RSS). Perform the following steps

1. Fit the curve, plot it
2. Compute the RSS and R^2 measures
3. Discuss the results, is it a good R^2 score?

Task 2

The second task consists of fitting a model to a set of data points that contains outliers. Here are the first five entries

	x1	x2	x3	y
0	7.892287	318.817999	162.969896	2112.420441
1	8.829627	303.180318	181.398715	2096.231124
2	13.810566	296.230913	145.848743	2067.044905
3	12.863271	325.830097	167.996165	2269.262403
4	13.697517	254.035329	171.892006	1966.604422

The goal is to try and compare different approaches to handle outliers. Use the train/test split methodology (ex. 80-20 splits) and compare the test MAE score for each one of the following approaches

1. Fit a linear regression with the outliers
2. Fit a huber regression with the outliers
3. Fit a linear regression without the outliers

The linear regression without outliers and huber regression models should both perform better than the linear regression with outliers.

RESOURCES

warm-up.zip
14.6 KB