Project

Use a dataset from here: http://www.cs.utah.edu/~jeffp/teaching/cs4964/D3.csv and import it in Python.

1- let the first three columns of the data set be separate explanatory variables x1, x2, x3. Again let the fourth column be the dependent variable y.

a) Run linear regression simultaneously using all three explanatory variables. Report the linear model you found. Predict the value of y for new (x1,x2,x3) values (1, 1, 1), for (2, 0, 4), and for (3, 2, 1).

b) Use cross-validation to predict generalization error, with error of a single data point (x1, x2, x3, y) from a model **h** as (**h**(x1, x2, x3) − y)^ 2 . Describe how you did this.

c) Now follow the same procedure (a & b) using scipy.stats.linregress(*x*, *y=None*) and compare the results with your version of the GD algorithm.

2- Let the first column of the data set be the explanatory variable x, and let the fourth column be the dependent variable y.

a) Run simple linear regression to predict y from x. Report the linear model you found. Predict the value of y for new x values 0.5, 1, 1.5, 2, 2.5, 3.
b) Now use Locally weighted regression and predict the value of y for new x value 0.5, 1, 1.5, 2, 2.5, 3. Compare the result with your experiment in 2.a).
c) Use the following weighting schema:

$$w^{(i)} = \exp\left(-\frac{(x^{(i)} - x)^2}{2\tau^2}\right)$$

You can adjust meta-parameter **τ** and watch in real time its influence on the model.