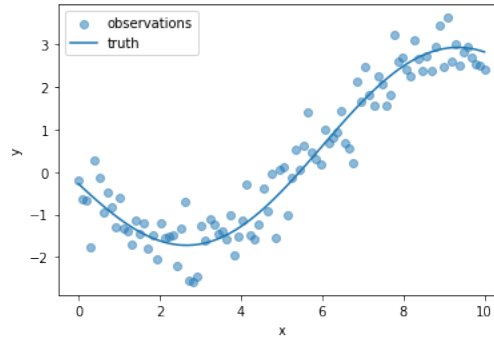**Problem 1.** Method of Least Squares, Gradient Descent

The date file, *data.csv*, contains observations generated by *truth + noise* as indicated by the following graph


You decide to adopt the following regression function:

$$E(y_i|x_i) = \beta_0 \sin(\beta_1 x_i - \beta_2) + \beta_3 x_i$$

Please fit the model by minimizing the squared loss using the gradient descent method. Please use the initial value $\boldsymbol{\beta}^{(0)} = (0.1, 0.1, 0.1, 0.1)^T$ and a fixed learning rate of your choice. A learning rate in the neighborhood of 0.00005 appear to represent a sweetspot.

Please let your code report the following
[1] The smallest loss you have achieved.
[2] The corresponding learning rate.
[3] The corresponding "best" estimated $\hat{\boldsymbol{\beta}}$
[4] The corresponding number of updates at the "best" estimated $\hat{\boldsymbol{\beta}}$
[5] A plot of updates of the loss values vs the update number, in the following format
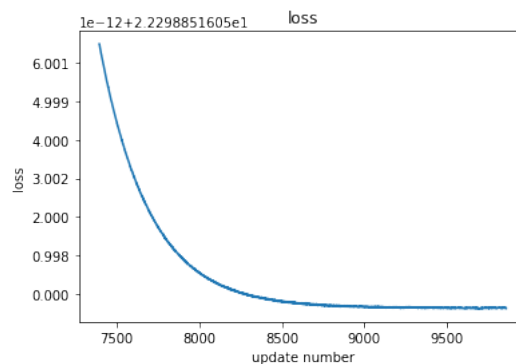
[6] A plot of the fitted regression function, in the following format. For the blue line, please use the truth $y_i = 2\sin(0.5x_i - 3) + 0.1x_i$
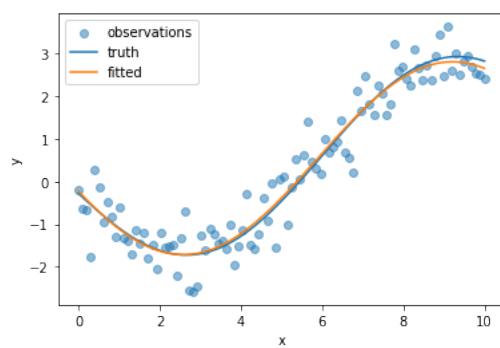
You may find the following resources useful.