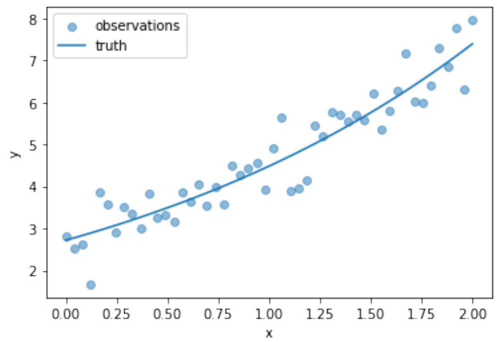
ACTU PS5841 Data Science in Finance & Insurance - Spring 2022 (Y. Wang)
Assignment - 5
Assigned 2/17/22, Due 2/27/22 (Sun)

Problem 1. Method of Least Squares, Newton Raphson
The date file, *data.csv*, contains observations generated by

$$y_i = \exp(1.0 + 0.5x_i) + \epsilon_i$$

as indicated by the following graph



You decide to adopt the following model:

$$E(y_i|x_i) = \exp(\beta_0 + \beta_1 x_i)$$

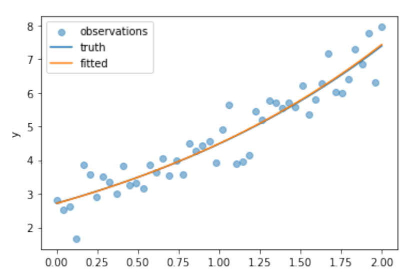
Please fit the model by minimizing the sum of squared losses using the Newton Raphson method. Please use the initial value $\boldsymbol{\beta}^{(0)} = (0.5, 0.75)^T$.

[1] Please let your code report the following
SSE at $\boldsymbol{\beta_{truth}} = (1.0, 0.5)^T$
The smallest loss the algo has achieved (to the 16th decimal place),
the corresponding fitted parameters (to the 5th decimal place),
the number of iteration needed to reach this loss.

Note: If $\boldsymbol{\beta}^{(1)}$ produces the smallest loss, the number of iteration needed is 1.

[2] Let your code print out the following table, up to and including the smallest loss achieved.

iteration	$\hat{\boldsymbol{\beta}}$	loss
initial guess	0.50000, 0.75000	x.xxxxxxxxxxxxxxxx
1	x.xxxxx, x.xxxxx	x.xxxxxxxxxxxxxxxx
⋮	⋮	⋮



[3] Let your code produce a plot of the fitted regression function

[4] Please estimate the prediction error of your model, defined as $RMSE = \sqrt{E(\text{loss per test observation})}$, using LOOCV.

[5] Please estimate the prediction error of your model, defined as $RMSE = \sqrt{E(\text{loss per test observation})}$, using the 10-fold cross validation approach. For grading purposes, please use the same random_state as follows,

KFold(n_splits = 10, shuffle = True, random_state = 10)

You may find the following resources useful.

K-Fold cross-validator
https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.KFold.html

Validation
At the initial value $\boldsymbol{\beta}^{(0)} = (0.5, 0.75)^T$,

$$H(\boldsymbol{\beta}^{(0)}, R) = J(\boldsymbol{\beta}^{(0)}, \nabla R) = \begin{bmatrix} 1474.68839977 & 2265.88996475 \\ 2265.88996475 & 3746.90298484 \end{bmatrix}$$