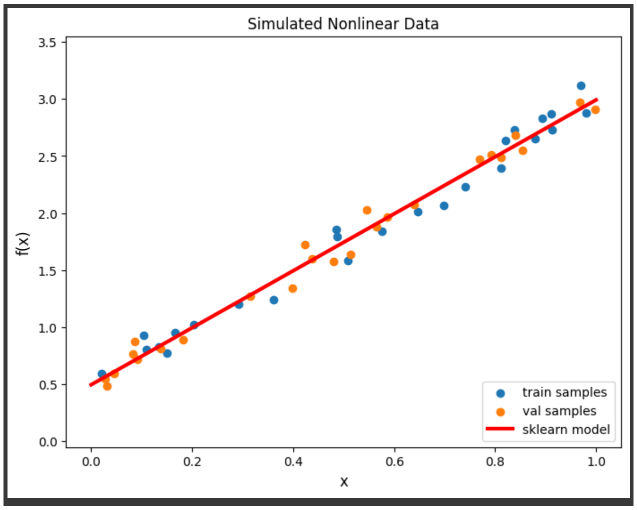
https://colab.research.google.com/drive/18e1JgI-BL_tiDt1f7mypSOCyHjEAMj_q?usp=sharing

Part 1.a Results

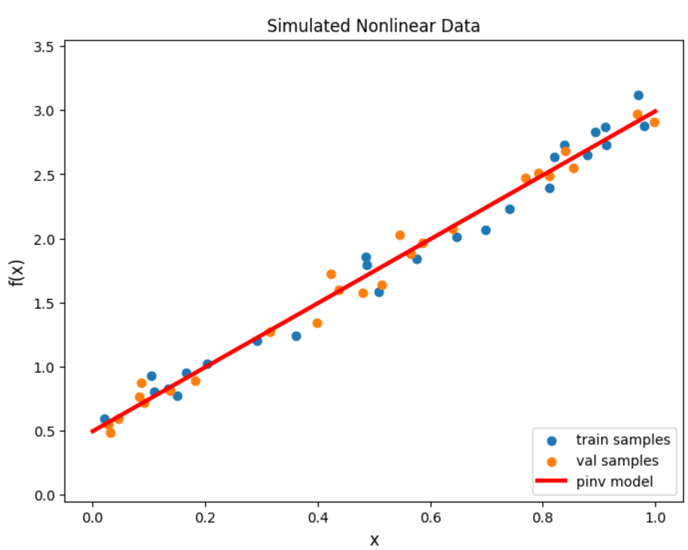
MSE of sklearn model: 0.00763903766246723



Part 1.b Results

MSE of manual model: 0.007639037662467223

W = [[2.49513012]
 [0.49648597]]

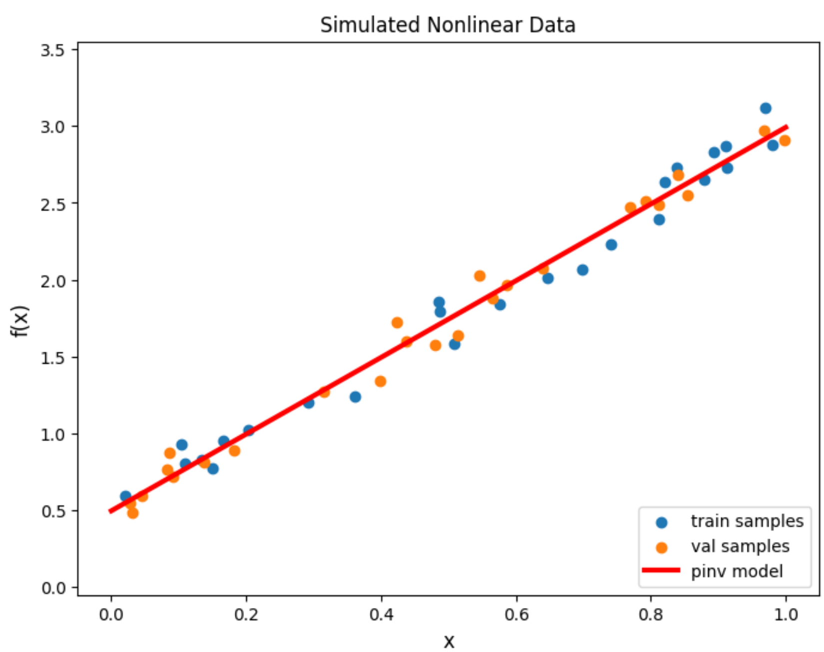


As it can be seen MSE errors in part 1.a and 1.b are very close, thus obtained plots are very similar as well.

Part 1.c Results

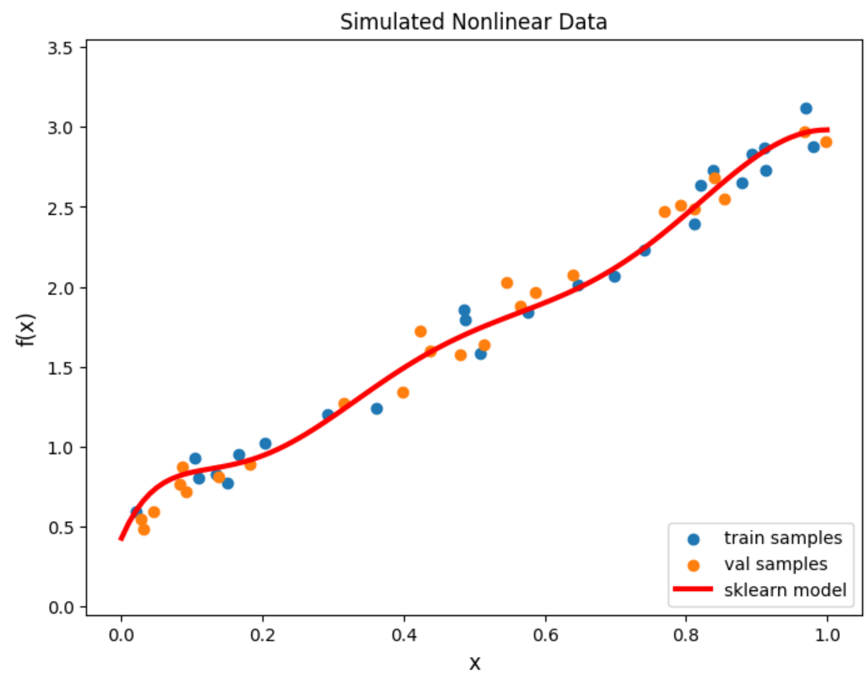
MSE error at step 1: 3.8596
MSE error at step 100: 0.0733
MSE error at step 200: 0.0254
MSE error at step 300: 0.0151
MSE error at step 400: 0.0128
MSE error at step 500: 0.0124
MSE error at step 600: 0.0123
MSE error at step 700: 0.0122
MSE error at step 800: 0.0122
MSE error at step 900: 0.0122
MSE error at step 1000: 0.0122

W = [[2.49471105]
 [0.49673457]]



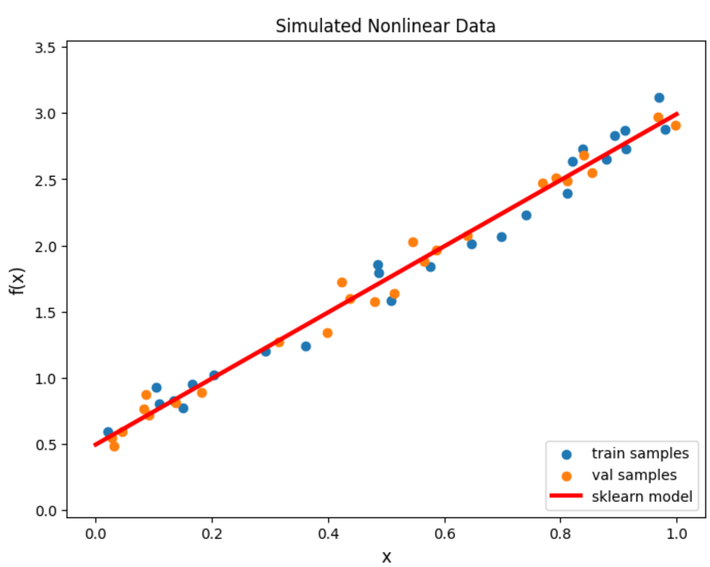
Part 2.a Results

MSE of sklearn model: 0.8162778807826367
MSE of sklearn model: 0.8174454285285526
MSE of sklearn model: 0.8133335050058378
MSE of sklearn model: 0.7884288757156074



Part 2.b Results

MSE of sklearn model: 1.06477782051187
W = [[0.24824298]
[0.24824298]
[2.49513012]]



The degree parameter controls the complexity of the model. When it is taken too large, we might have a very flexible model which would cause overfitting. When it is taken too small, it underfits the data, making the model simplistic to capture the underlying patterns and relationships between the input and output variables. The optimal degree is the one that minimizes the mean squared error on the validation set.

As it can be seen in 2.b when polynomial degree was set as 3 the model underfit. The model was not flexible enough to represent the relationship between the features and the target variable.

In our case degree 5 is the optimal degree. Because when the degree is smaller it underfits the data, when the degree is larger than 5, it makes the model too flexible which results in lower training error but higher validation error compared to degree = 5.