

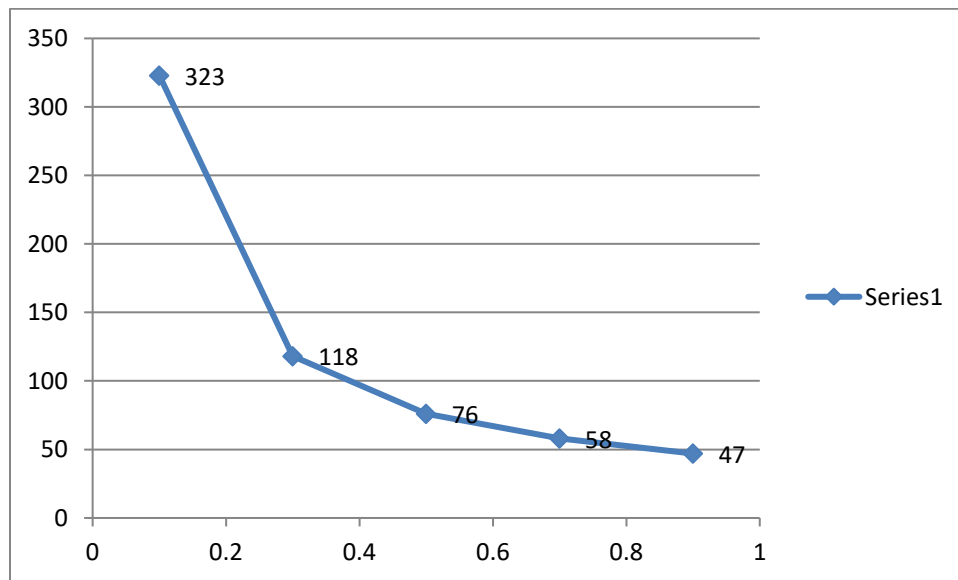
## Two Problems of Regression

Mohammad Mirzanejad

- **Banknote Authentication:**

Data were extracted from images that were taken for the evaluation of an authentication procedure for banknotes produced by UCI.

Since the evaluation of the algorithm is in 5 folds, the number of repetitions is reported for the highest amount of repetitions in folds.



Changes in the number of repetitions according to changes in learning rates

For eachFold, statistics are presented in detail in the program. Here we refer to the average accuracy criteria in 5 folds.

Learning rate	accuracy	precision	recall	F-Score	RMSE	Max Epochs
0.1	98.18	5.6	1.4	2.2	0.099	323
0.3	98.18	5.6	1.4	2.2	0.101	118
0.5	98.25	5.6	1.2	1.95	0.103	76
0.7	98.25	5.6	1.2	1.95	0.105	58
0.9	98.25	5.6	1.2	1.95	0.107	47

### Program Steps:

- **loadData** reads the dataset
- **findDatasetMinMax** finds the maximum and minimum values of each feature for normalization
- **normalizeDataset** normalizes the dataset values
- **crossValidationSplit** splits the dataset according to the number of folds
- **crossValidationEvaluate** performs train and test with related calls and according to the division of folds. In this function, for each fold, the values of accuracy criteria as well as RMSE are performed by calling related functions
- **calculateEvaluationScores** calculates the values of accuracy criteria such as precision, recall, F-score, etc.
- **gradientDescent** executes the gradient descent algorithm according to the set learning rate on the training data, and the stop condition is that the error (delta) is less than 0.002
- **logisticRegression** is the main function of the program, which obtains the values of the coefficients of the linear equation by calling the gradientDescent function, and then calls the predict function and returns the prediction results.
- **predict** multiplies the coefficients obtained from the training by the test data value and gives the result to the expit function which calculates the value of the sigmoid function for it and outputs a number between 0 and 1, then this output is used for The classify function is sent and there the output value is attributed to the related class depending on how close it is to any of the values 0 or 1.

- **Face Completion:**

Using the normal form equation in closed form and with the respective formula the pixels of the lower half for the input data of the test is predicted. The upper half matrix is multiplied by the  $W$  matrix, and the resulting output (bottom half) is saved as image file. Implementation was also done with scikit .

It should be noted that the input and output data are normalized

RMSE criterion for the test data is printed in the program, due to its high numbers, here we refer to its lowest value for the program as well as the output of Scikit :

ProgramRMSE	Scikit RMSE
45	22

Also, the best example of the test, which is more similar to the original example, as well as the output generated by the scikit library is given below.



Scikit



program output