ELL409 Assignment 1 Report

Group 18

Group members

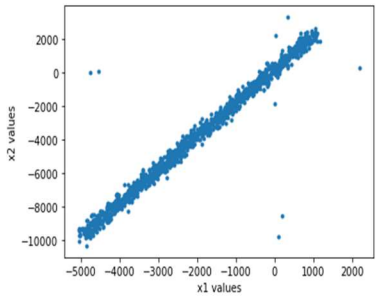
Perugu Krishna Chaitanya Yadav(2020EE10521)

Muvva Srija(2020EE30605)

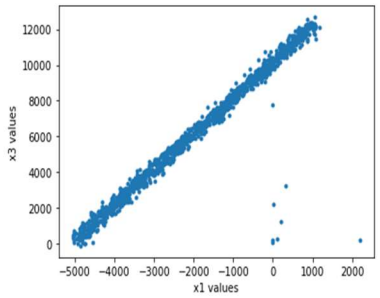
Devarakonda Ronith Kumar(2020EE10486)

Part 1: Multivariate Linear Regression

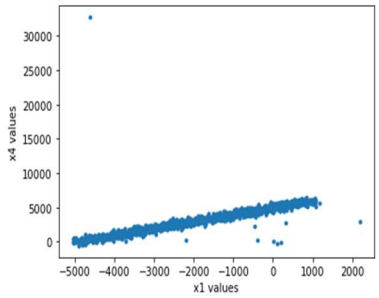
1) Understanding Data: Plotting the underlying pairwise-correlation between the variables



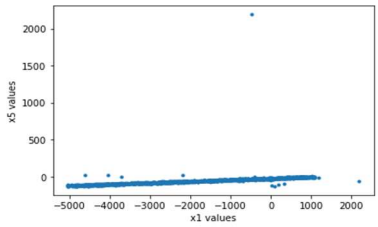
Co-relation between x_1 and x_2



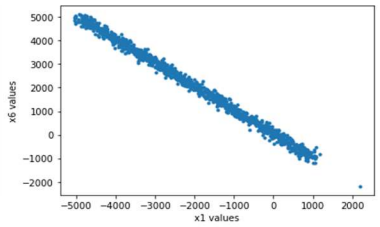
Co-relation between x_1 and x_3



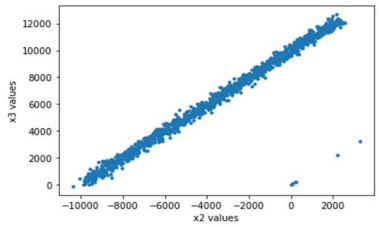
Co-relation between x_1 and x_4



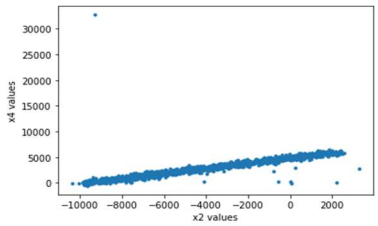
Co-relation between x_1 and x_5



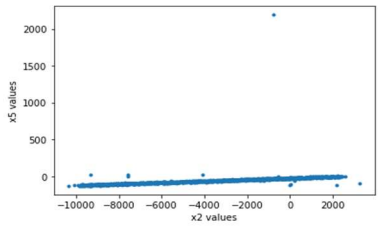
Co-relation between x_1 and x_6



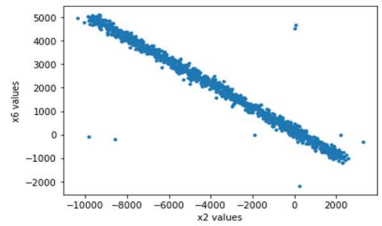
Co-relation between x_2 and x_3



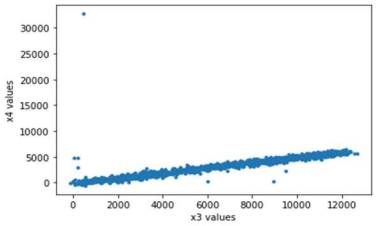
Co-relation between x_2 and x_4



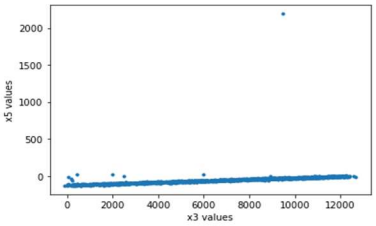
Co-relation between x_2 and x_5



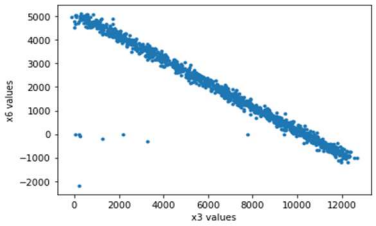
Co-relation between x_2 and x_6



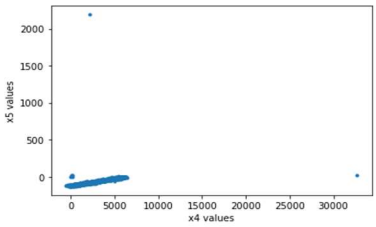
Co-relation between x_3 and x_4



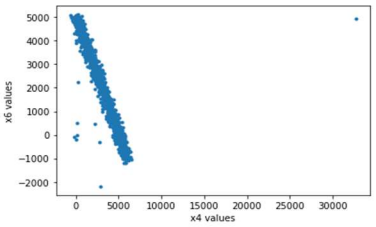
Co-relation between x_3 and x_5



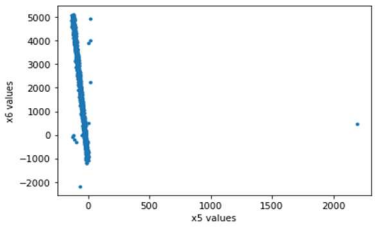
Co-relation between x_3 and x_6



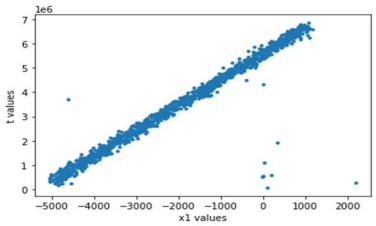
Co-relation between x_4 and x_5



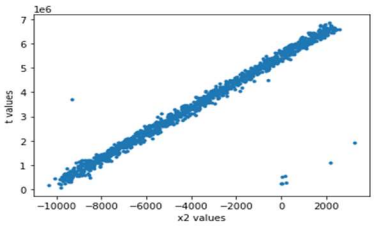
Co-relation between x_4 and x_6



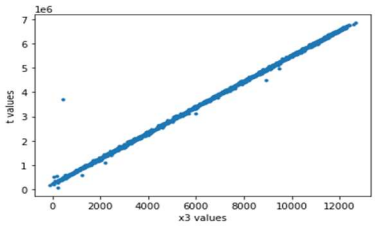
Co-relation between x_5 and x_6



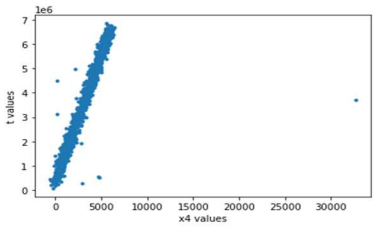
Co-relation between x_1 and t



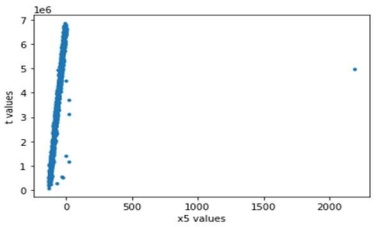
Co-relation between x_2 and t



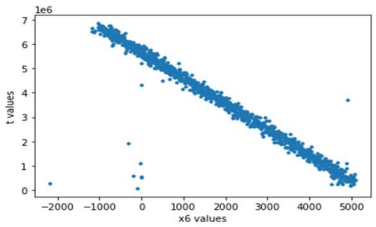
Co-relation between x_3 and t



Co-relation between x_4 and t



Co-relation between x_5 and t



Co-relation between x_6 and t

	x1	x2	x3	x4	x5	x6	t
x1	1.000000	0.983218	0.971908	0.847482	0.475974	-0.996590	0.970183
x2	0.983218	1.000000	0.973034	0.847281	0.476185	-0.983578	0.971919
x3	0.971908	0.973034	1.000000	0.851716	0.480625	-0.971250	0.998455
x4	0.847482	0.847281	0.851716	1.000000	0.409703	-0.845230	0.878992
x5	0.475974	0.476185	0.480625	0.409703	1.000000	-0.476806	0.479483
x6	-0.996950	-0.983578	-0.971250	-0.845230	-0.476806	1.000000	-0.969157
t	0.970183	0.971919	0.998455	0.878992	0.479483	-0.969157	1.000000

Co-relation Matrix

2) Effect of Batch size:

a) Stochastic Gradient Descent (SGD):

Weights for Batch size 1:

[-28.681681317492323, -49.832141013456805, 361.91366778344826, 177.709567919037, -1.6143406791566635, 30.107381852429047, 0.04112609695211394]

W_0 = 0.04112609695211394

W_1 = -28.681681317492323 W_2 = -49.832141013456805 W_3 = 361.91366778344826

W_4 = 177.709567919037 W_5 = -1.6143406791566635 W_6 = 30.107381852429047

Weights for Batch size 2:

[-35.094769319229435, -63.36462024440457, 357.0711074626355, 171.27594409457745, -1.803659049278778, 37.103410751609346, 0.04213745887404928]

W_0 = 0.04213745887404928

W_1 = -35.094769319229435 W_2 = -63.36462024440457 W_3 = 357.0711074626355

W_4 =171.27594409457745 W_5 = -1.803659049278778 W_6 = 37.103410751609346

Weights for Batch size 5:

[-11.191583437921787, -12.273445855767884, 447.3745927337001, 210.4955677512204, -1.014501661961397, 13.151453250592361, 0.04532507242618251]

W_0 = 0.04532507242618251

W_1 = -11.191583437921787 W_2 = -12.273445855767884 W_3 = 447.3745927337001

W_4 = 210.4955677512204 W_5 = -1.014501661961397 W_6 = 13.151453250592361

Weights for Batch size 30:

[-28.50296320458481, -47.474753219403574, 374.7653785240021, 179.77880229334787, -1.4788287364978236, 30.316149887794218, 0.04226510346657968]

W_0 = 0.04226510346657968

W_1 = -28.50296320458481 W_2 =-47.474753219403574 W_3 = 374.7653785240021

W_4 = 179.77880229334787 W_5 = -1.4788287364978236 W_6 = 30.316149887794218

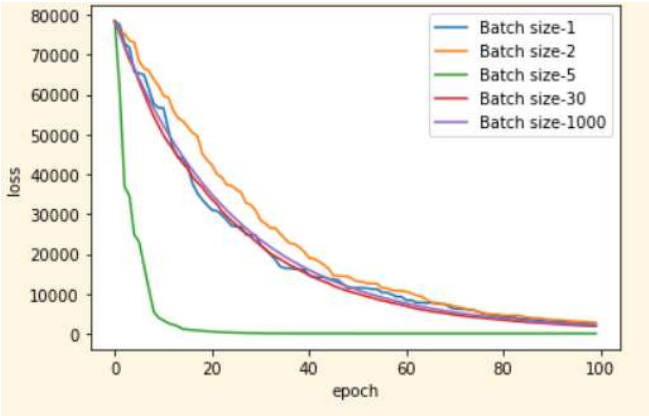
Weights for Batch size 1000:

[-30.323105361649706, -51.36637722629294, 369.58419380209455, 178.43458217131285, -1.555096561374349, 32.20598653308512, 0.04213036716240733]

W_0 = 0.04213036716240733

W_1 = -30.323105361649706 W_2 = -51.36637722629294 W_3 = 369.58419380209455

W_4 = 178.43458217131285 W_5 = -1.555096561374349 W_6 = 32.20598653308512



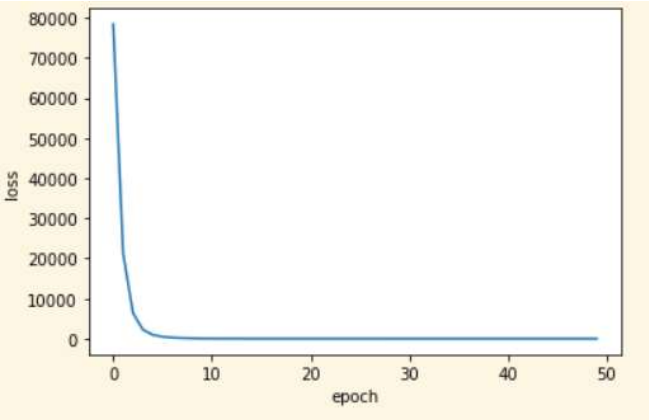
b) Batch Gradient Descent:

Weights (w vector) = [-18.274086347037695, -20.07445820050293, 486.7743038881445, 128.27650250478814, -0.2199774791725567, -11.42617980308219, 0.023187241721944907]

W_0 = 0.023187241721944907

W_1 =-18.274086347037695 W_2 = -20.07445820050293 W_3 = 486.7743038881445

W_4 = 128.27650250478814 W_5 = -0.2199774791725567 W_6 = 24.179590692249135



3) Effect of Regularization:

a) Lasso Regression:

Lasso Regression is a type of linear regression that uses shrinkage.

$$\mathbf{J(W)} = \frac{1}{2N}\sum_{i=1}^{N}((W_0 + W_1X_1^{(i)} + \ldots + W_PX_P^{(i)}) - Y_i)^2 + \frac{\lambda}{2N}\sum_{j=1}^{P}|W_j|$$

Lambda for Lasso Regression: 0.0879270583925551

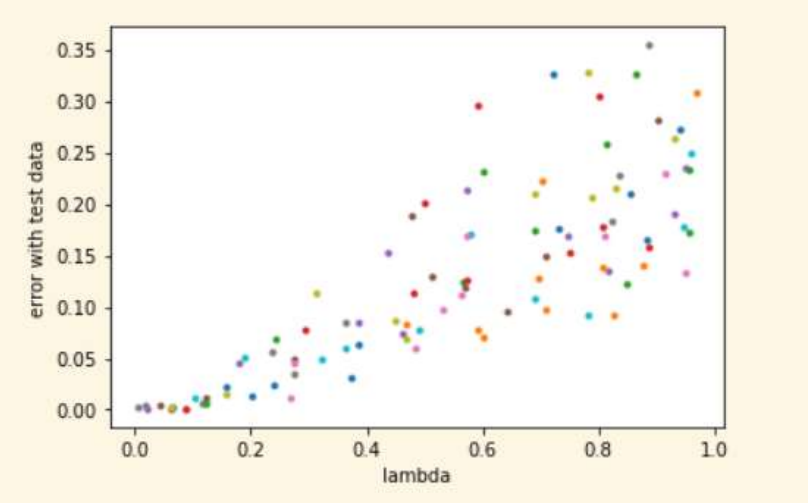
Weights (w vector): [0.02228199812619341, 0.030076133972875015, 0.27974356953369656, 0.05258742847465087, -0.0031931194977181552, -0.004447879233911833, 0.3330748952714911]

Error for test data with respect to above weights: 0.0010414856308377228

W_0 = 0.3330748952714911

W_1 = 0.02228199812619341 W_2 = 0.030076133972875015 W_3 = 0.27974356953369656

W_4 = 0.05258742847465087 W_5 = -0.0031931194977181552 W_6 = -0.004447879233911833



b) Ridge Regression:

$$\mathbf{J(W)} = \frac{1}{2N}\sum_{i=1}^{N}((W_0 + W_1X_1^{(i)} + \ldots + W_PX_P^{(i)}) - Y_i)^2 + \frac{\lambda}{2N}\sum_{j=1}^{P}W_j^2$$

Lambda for Ridge Regression: 0.00047586547375610966

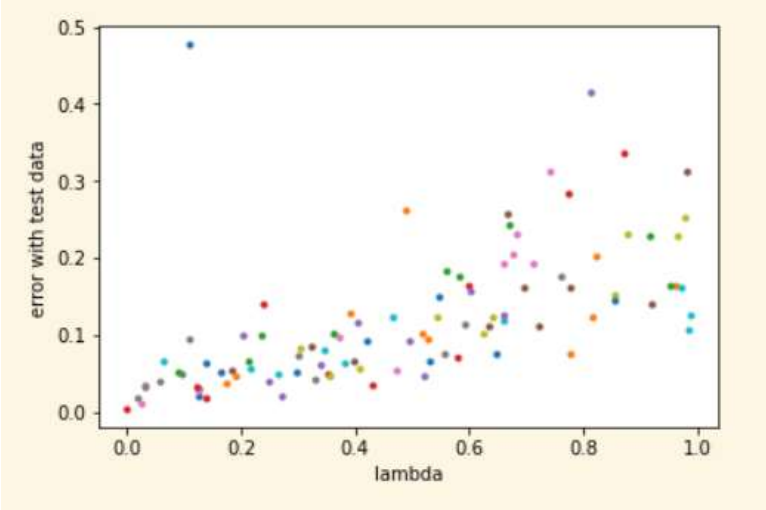
Weight vector: [-0.02260609758164117, 0.03296232542616735, 0.2646004776706431, 0.04988591014896376, -0.0034452628494108114, 0.014773621259699759, 0.32609667095226635]

Error for test data with respect to above weights: 0.00392510036949917

W_0 = 0.32609667095226635

W_1 = -0.02260609758164117 W_2 = 0.03296232542616735 W_3 = 0.2646004776706431

W_4 = 0.04988591014896376 W_5 = -0.0034452628494108114 W_6 = 0.014773621259699759



4) Feature Engineering:

Yes. We can eliminate some input variables by looking into correlation matrix there we observe x_1 & x_2 & x_6 columns are almost same so we can discard x_2 & x_6 columns as we can replace x_2 & x_6 with x_1 as they were approximately same. And then we used Lasso regression cost function to Now our input number of variables converted to four.

We got error value as 0.013182526695405983

And weight vector: [-0.10862917802789644, 0.015652476747846214, 0.0071514066086787, -0.002720034204480699, 2.2706754197404523e-05]

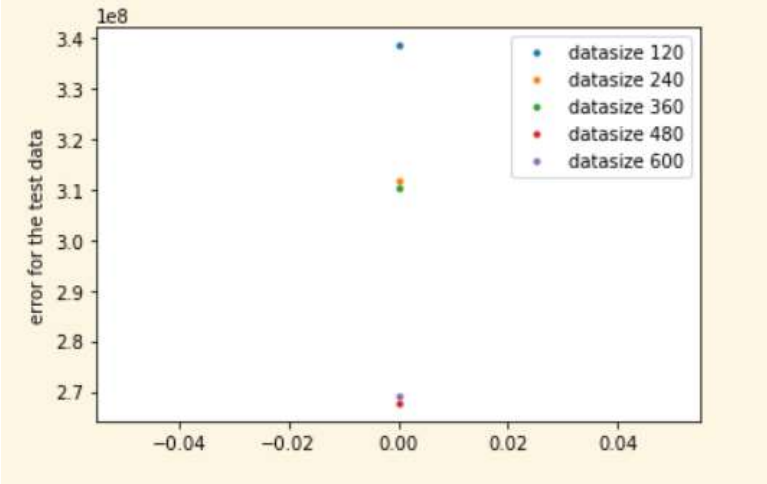
W_0 = 2.2706754197404523e-05

W_1 = -0.10862917802789644 W_2 = 0 W_3 = 0.015652476747846214

W_4 = 0.0071514066086787 W_5 = 0.002720034204480699 W_6 = 0

5) Effect on sample distribution:

As we see that with increase in sample size the error decreases due to increase in accuracy as the no of training data increases. The graph below clearly shows the above explanation



6) Estimate the variance:

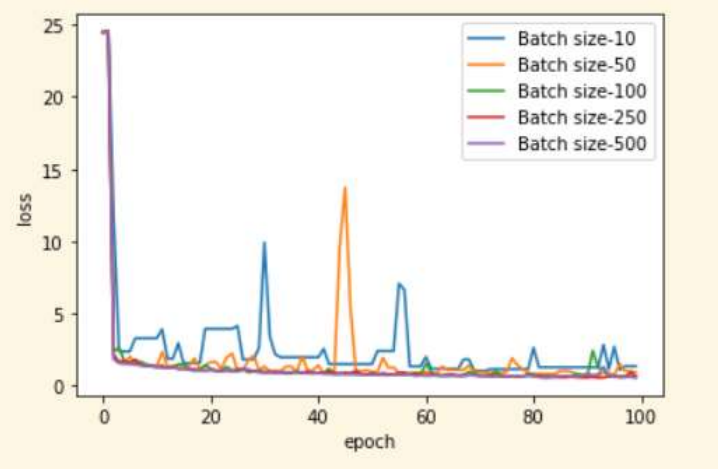
By the above observations we see that ridge regression gives good results so we found out the error in that case and calculated variance by applying statistical variance formula on the data

7)Optimal weights:

[-11.564882645841365, -11.237229832595743, 451.9216634867763, 197.9508092856194, -1.1418201482490027, 13.442564944426275, 0.046158108681268] (400,) (7, 400)
55.689313692689915

Part 2: Multivariate Logistic Regression

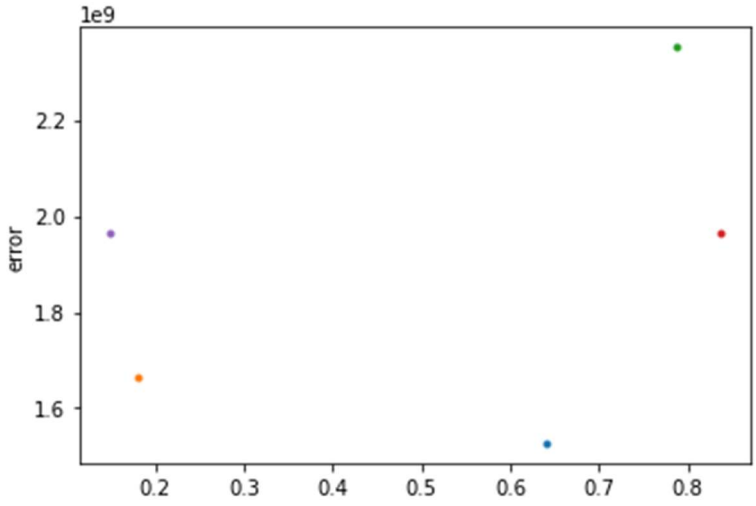
1) Effect of batch size:



2) Effect of Regularization:

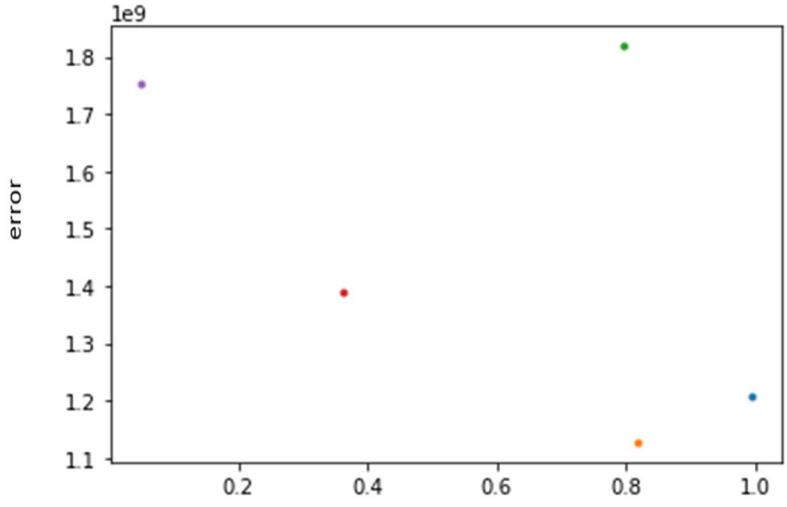
a) Lasso Regression:

Lambda for Lasso Regression: 0.6412112909078909



b) **Lasso Regression:**

Lambda for Ridge Regression: 0.8168802270524591



3)

With Stochastic Gradient descent we got this with highest accuracy

Error: 0.633

Rate : 0.1

Iterations: 100

batch size: 200

Part 3:(Bonus part)

Same as part1 we classify 10 lines.. Firstly we make first line as 1 and all other as 0 repeating this the second line takes all 1s and all other are zero.

We find w0x,w1x.. the highest value among these will get us know whether it is 0, 1, 2..