ASSIGNMENT # 2
22i-2336
REPORT

Stochastic Gradient Descent :

For every row we calculate parameters/weights (via that derivation formula) and then we update weights using the weight update equation and then we calculate error ,classic ! more variance (more changes alongside) less convergence (it keeps on recalculating the best fit , so does not approaches perfection smoothly)

Batch Gradient Descent :

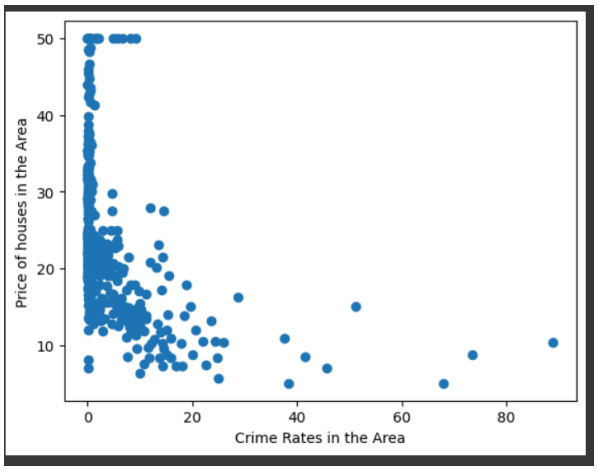
All our formulas imagine a summation of all rows along standard formulas , so we find weights for all of data , update , find error of whole data , all of data is taken as 1 batch less variance obv but a lot of data processed at every epochs so resource utilization high better smoother convergence

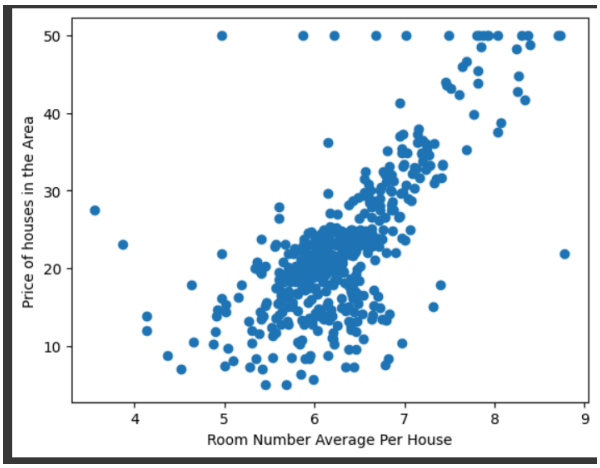
Mini-Batch Gradient Descent :

The best of both worlds , combines the updations/feedback of stochastic with data convergence and low variance of batch divides data into batches , in context of model batch happening in context of batches batch gd happening

DataSet Explanation :

I chose Boston Housing Dataset , which can easily be accessed via their url , i loaded the boston housing dataset into a df , from the dataframe i analysed 2 of its features by making a graph to see their distributions



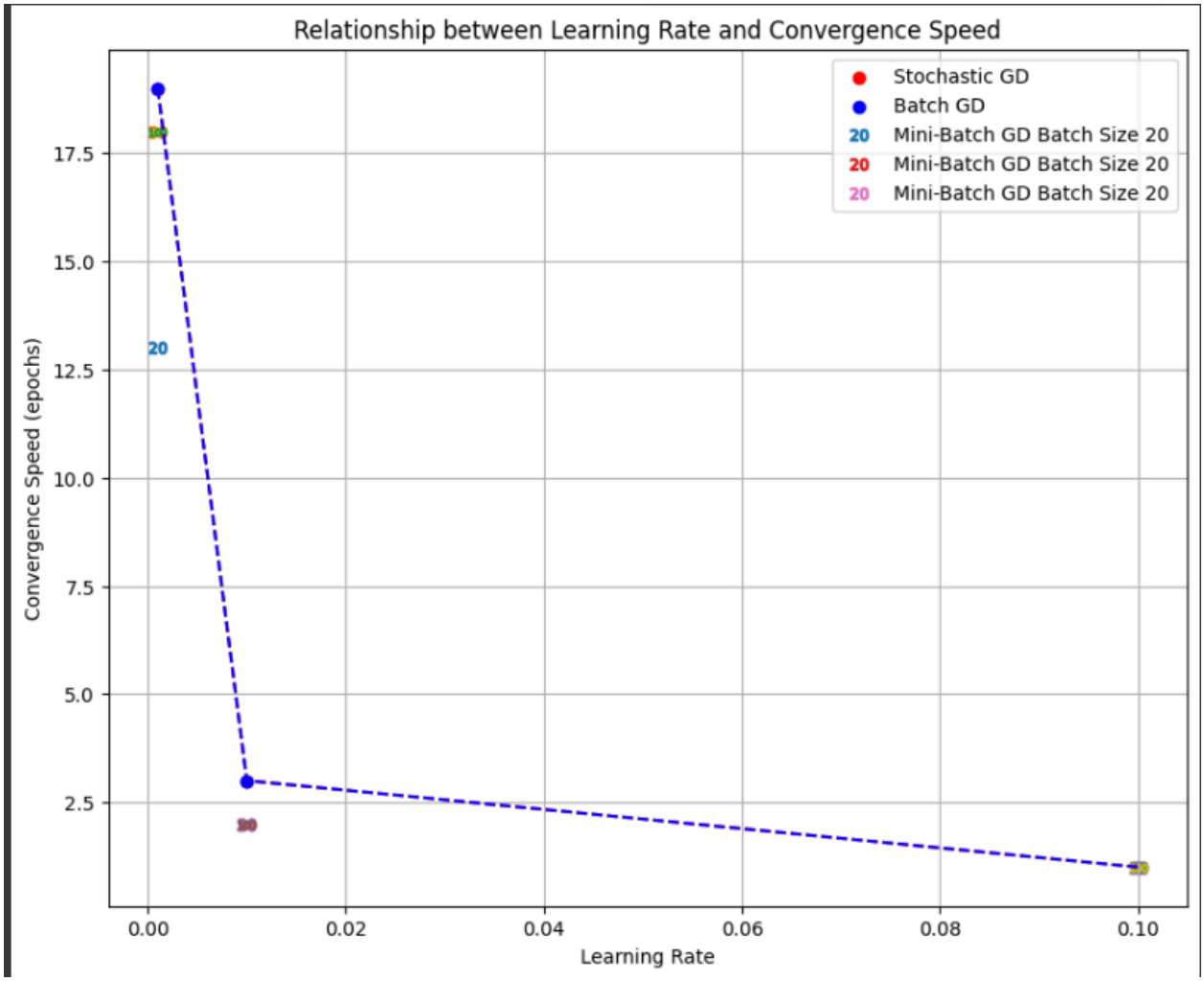


By analyzing these graphs i decided my input feature (X) will be room number average per house and my y , predict feature will be Median Value of House (Price Of house) because the room graph had more variance and more spread out distribution for it to be represented by a best fit line .

Training Evaluation

GD Type	Learning Rate	Batch Size	MSE	Convergence Speed /epochs
Stochastic GD	0.001	none	46.18	19
Stochastic GD	0.01	none	46.83	3
Stochastic GD	0.1	none	46.93	1
Batch GD	0.001	none	46.18	19
Batch GD	0.01	none	46.83	3
Batch GD	0.1	none	46.93	1
Mini-Batch GD	0.001	20	46.12	13
Mini-Batch GD	0.001	50	46.15	18
Mini-Batch GD	0.001	100	46.13	18
Mini-Batch GD	0.01	20	46.23	2
Mini-Batch GD	0.01	50	46.41	2
Mini-Batch GD	0.01	100	46.35	2
Mini-Batch GD	0.1	20	46.92	1
Mini-Batch GD	0.1	50	45.75	1
Mini-Batch GD	0.1	100	51.08	1

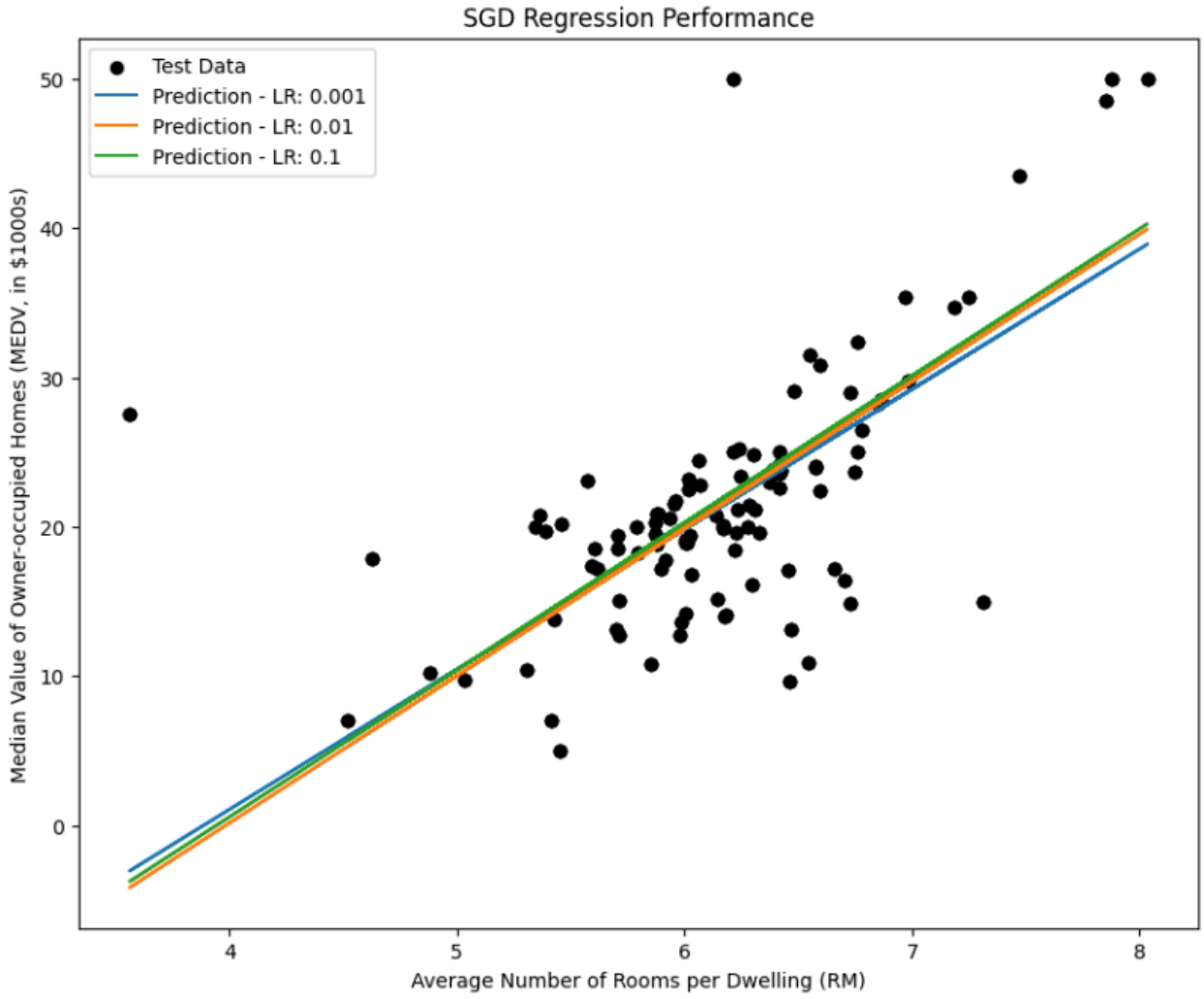
Below is a simple graph identifying the relationship between convergence speed and learning rates to better understand the results



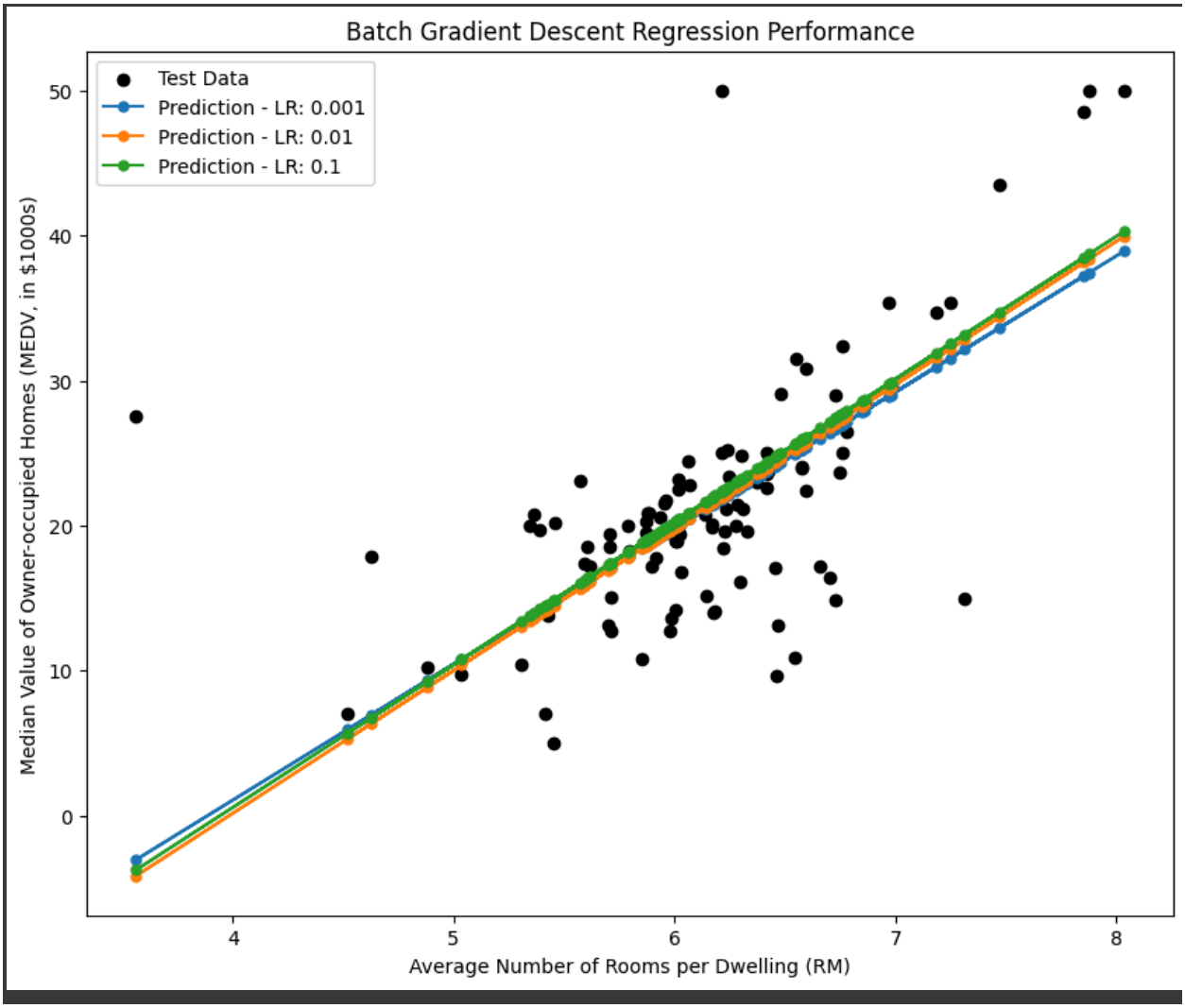
Analysis :

- For all type of gradient descents a higher learning rate ie 0.1 always gave us better results in context of convergence speed , speed is best
- Learning rate $increase$ convergence speed $increase$
- But if you want a better accuracy / precision of model a learning rate if 0.01 is good enough it,s a tradeoff between speed and accuracy
- Having a smaller batch size in mini batch gradient descent is better for speed and mse both but batch size is secondary to learning rate
- Having a very high learning rate in mini batch with a high batch number gave us a relatively high mse so accuracy of model is affected

Stochastic Gradient Descent



Batch Gradient Descent



Mini Batch Gradient Descent