Programming Assignment 01

Usama Munir Sheikh

February 9, 2017

1 Analytical Solution

The equation used: $\Theta = (X^TX)^{-1}X^TY$

where, X= $\begin{bmatrix} X^T[1] & 1 \\ X^T[2] & 1 \\ \vdots & \vdots \\ X^T[N] & 1 \end{bmatrix}$, Y= $\begin{bmatrix} Y[1] \\ Y[2] \\ \vdots \\ Y[N] \end{bmatrix}$ and $N = 10000$, $X \in^{10\times1}$, $Y \in^{1\times1}$

Solution:

$$\Theta = \begin{bmatrix} 0.43672729 \\ 0.73829186 \\ 0.2462168 \\ 0.78105533 \\ 0.01109783 \\ 0.96073395 \\ 0.17140692 \\ 0.44954371 \\ 0.81574237 \\ 0.92059869 \\ 0.08529076 \end{bmatrix}$$

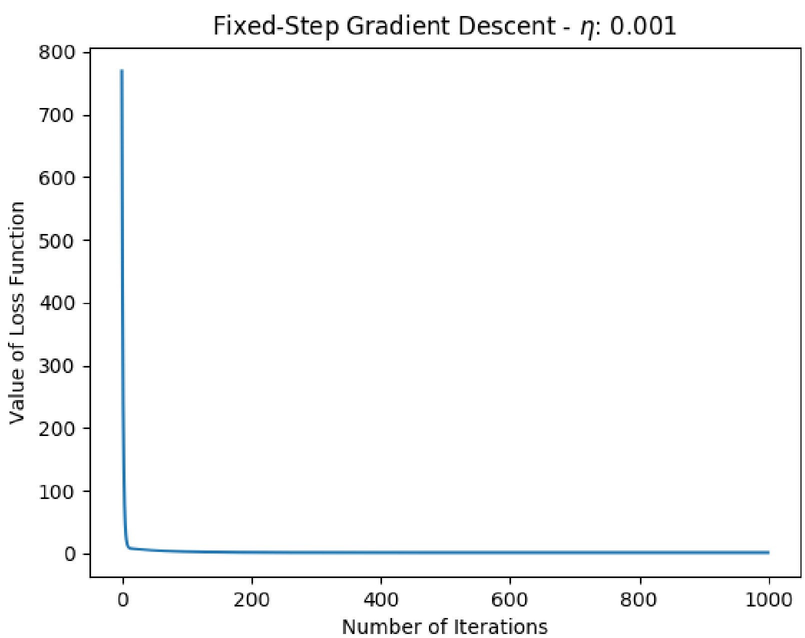
Final LMS Error Value = 0.98522103

2 Fixed Step Gradient Descent

Using $\eta = 0.001$ and 1000 iterations
Solution:

$$\Theta = \begin{bmatrix} 0.43617994 \\ 0.73767656 \\ 0.24574108 \\ 0.78040385 \\ 0.01068305 \\ 0.96006054 \\ 0.17093267 \\ 0.44900113 \\ 0.81508505 \\ 0.91994727 \\ 0.11468174 \end{bmatrix}$$

Final LMS Error Value = 0.98525065



2.1 Let's try a different stopping criteria:

Stop when,

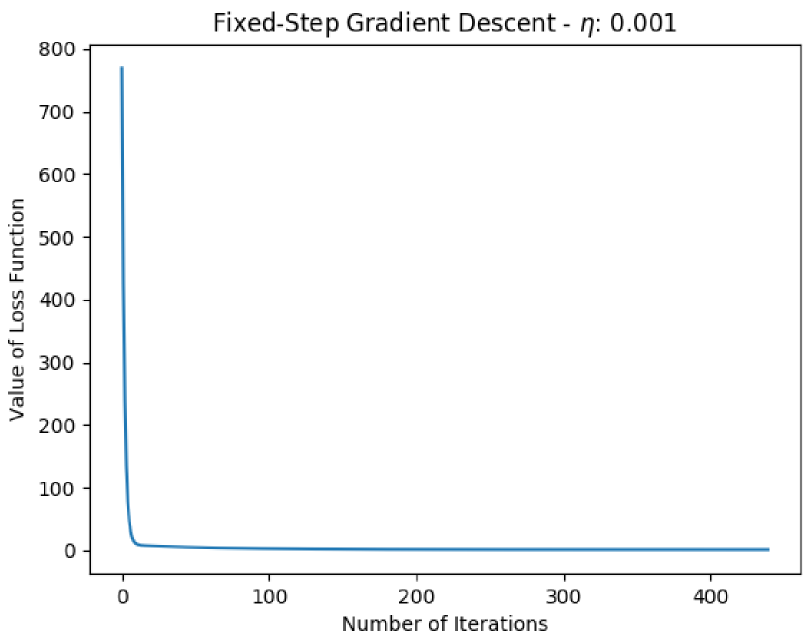
$$LossValueatCurrentIteration - LossValueatPreviousIteration < Threshold$$

where, Threshold = 0.0001
Now,

For $\eta = 0.001$
$Numberofiterationstostop = 440$

$$\Theta = \begin{bmatrix} 0.43919888 \\ 0.73260444 \\ 0.25350171 \\ 0.77351481 \\ 0.02430954 \\ 0.94902444 \\ 0.18189262 \\ 0.45190489 \\ 0.80699557 \\ 0.9122777 \\ 0.11519737 \end{bmatrix}$$

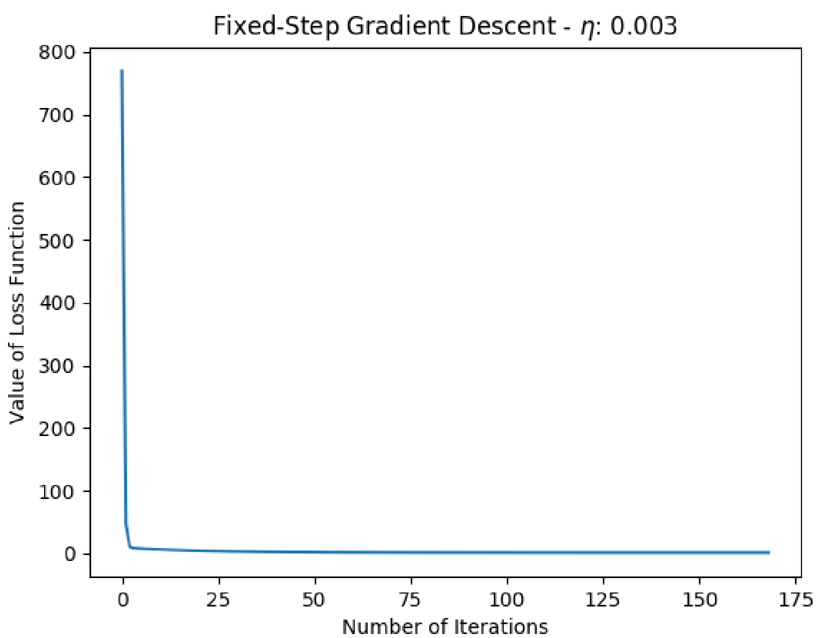
Final LMS Error Value $= 0.99122888$



For $\eta = 0.003$
Number of iterations to stop $= 169$

$$\Theta = \begin{bmatrix} 0.43783566 \\ 0.73488098 \\ 0.24999776 \\ 0.77652889 \\ 0.01813937 \\ 0.95394903 \\ 0.17704488 \\ 0.45062625 \\ 0.81051683 \\ 0.91588986 \\ 0.11514068 \end{bmatrix}$$

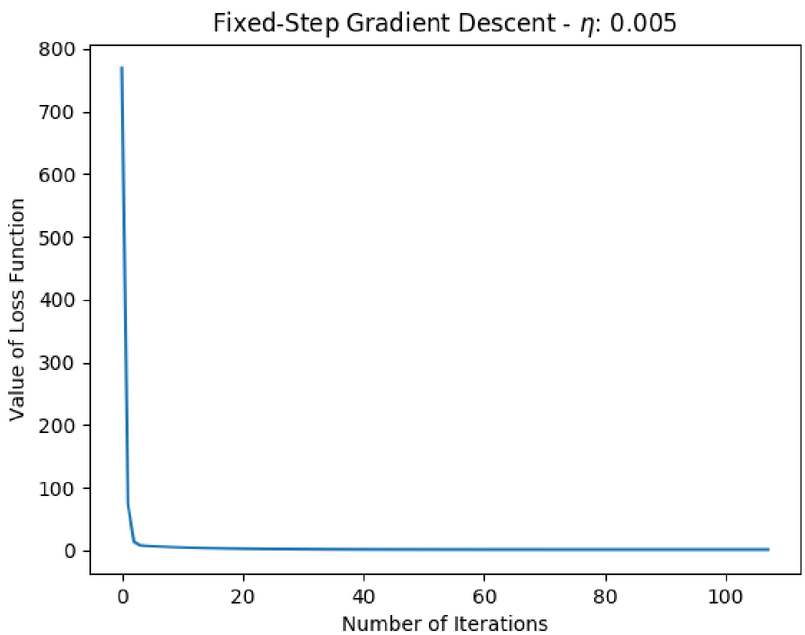
Final LMS Error Value = 0.98716229



For $\eta = 0.005$
Number of iterations to stop = 108

$$\Theta = \begin{bmatrix} 0.43737888 \\ 0.73564494 \\ 0.24882592 \\ 0.77755541 \\ 0.01607961 \\ 0.95560712 \\ 0.17540315 \\ 0.45019144 \\ 0.81172019 \\ 0.91706663 \\ 0.11511151 \end{bmatrix}$$

Final LMS Error Value = 0.98631072



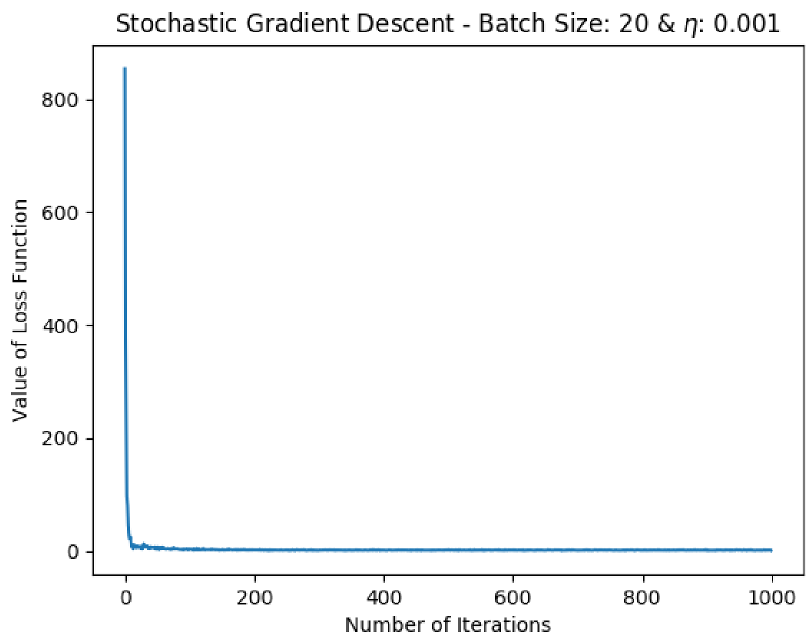
From our results we can see that the number of iterations taken to converge to a final solution has decreased dramatically with a small increase in the value of the learning rate η

3 Stochastic Gradient Descent

Using $\eta = 0.001$ and 1000 iterations and batch size of 20

Solution:

$$\Theta = \begin{bmatrix} 0.431741 \\ 0.74421006 \\ 0.25169167 \\ 0.78037113 \\ 0.00584533 \\ 0.95606256 \\ 0.1757403 \\ 0.44549447 \\ 0.82539666 \\ 0.92212236 \\ 0.11725251 \end{bmatrix}$$



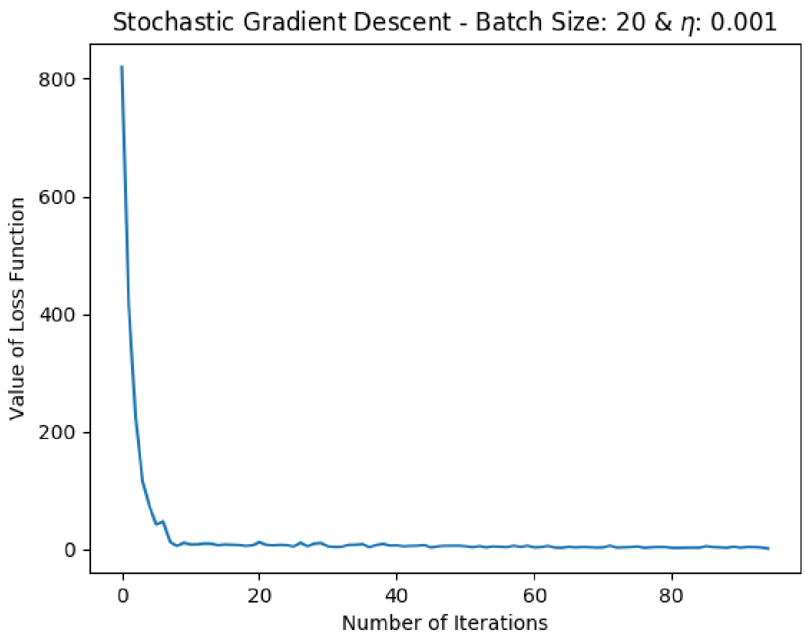
3.1 Let's try a different stopping criteria:

Stop when,

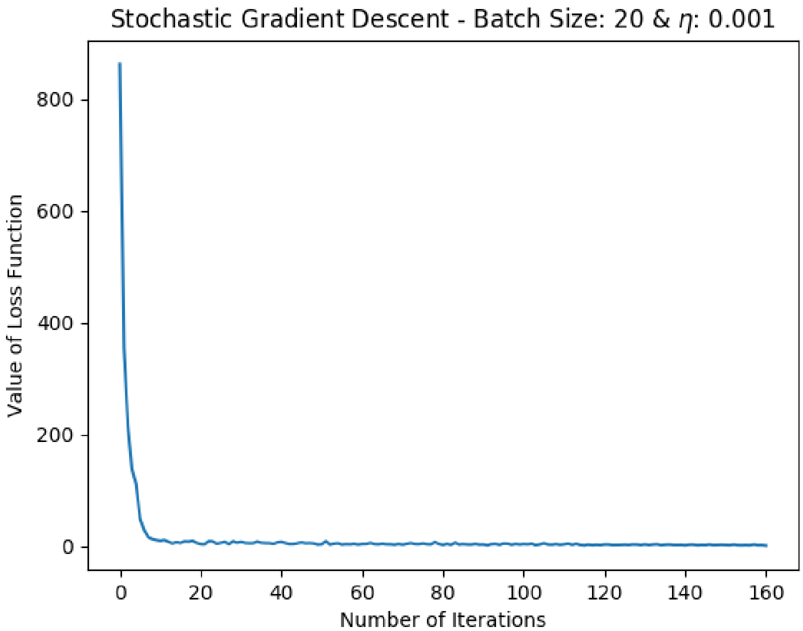
$$LossValueatCurrentIteration < Threshold$$

where, Threshold = 0.985
Choice of Threshold is taken from part (a), the analytical solution
Now, the effect of different batch sizes is studied. To do this, we must make sure we get the same batches for each run. If we pick points for each batch randomly, we won't be able to compare the effect of batch size. Even still, since we randomly pick which batch to use at every iteration, the results will be varied

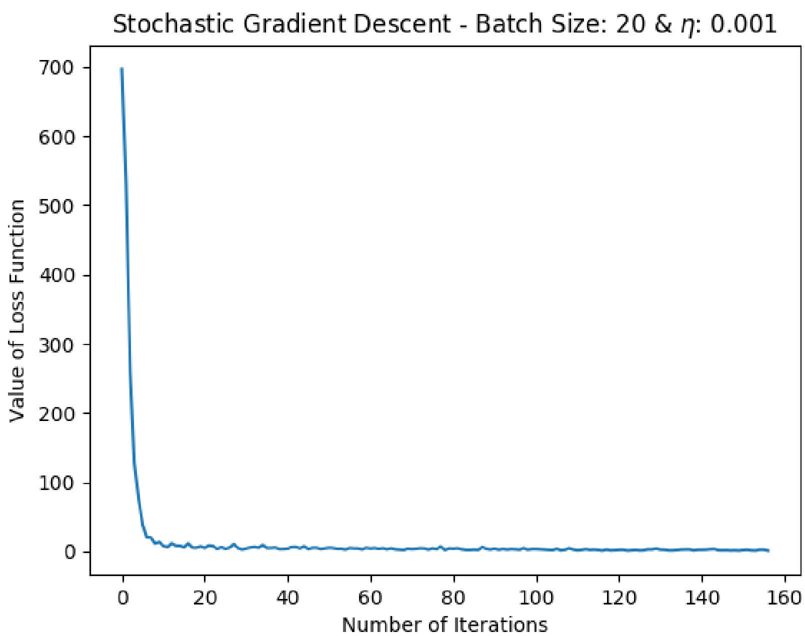
and it is hard to compare. In summary, there are two areas of randomness here. First we randomly assign data points to each batch and then we randomly select which batch to use.
For Batch Size = 20
Number of iterations to stop on first run= 95



Number of iterations to stop on second run= 161

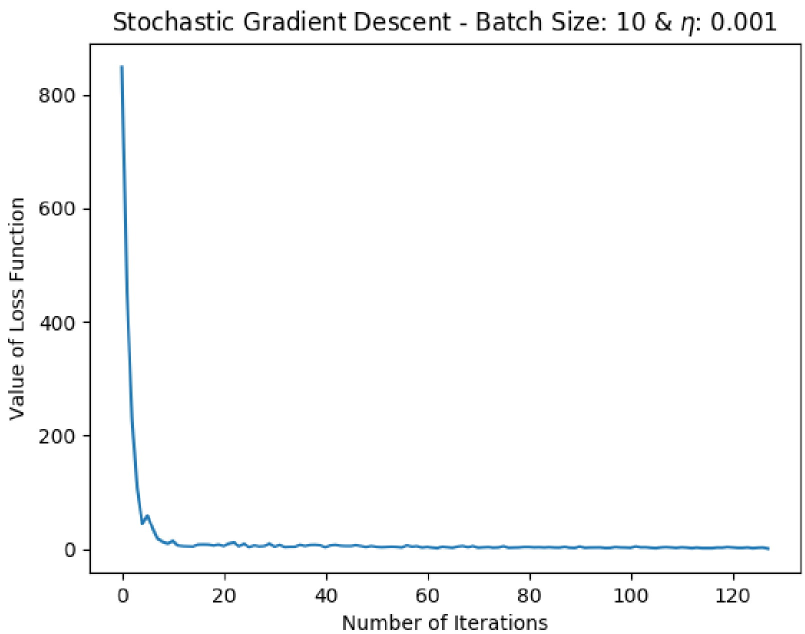


Number of iterations to stop on third run= 157

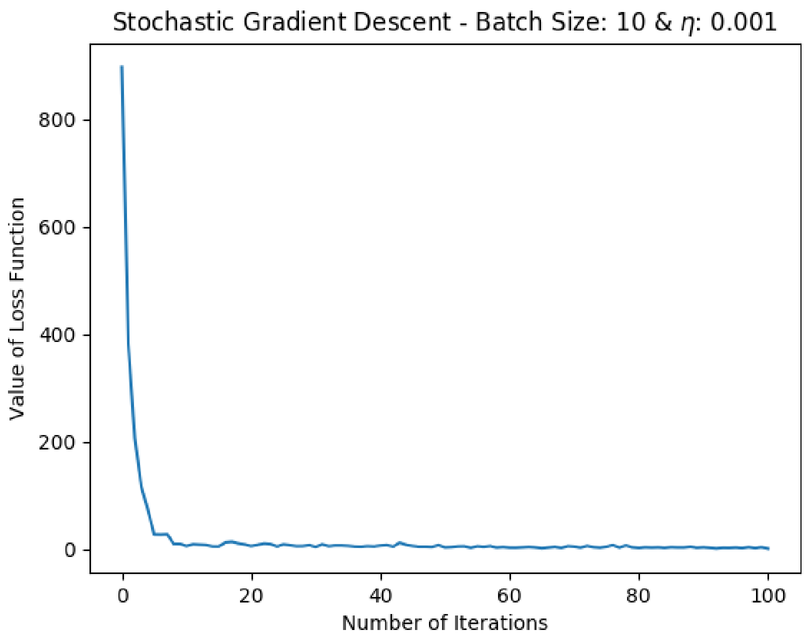


For Batch Size = 10

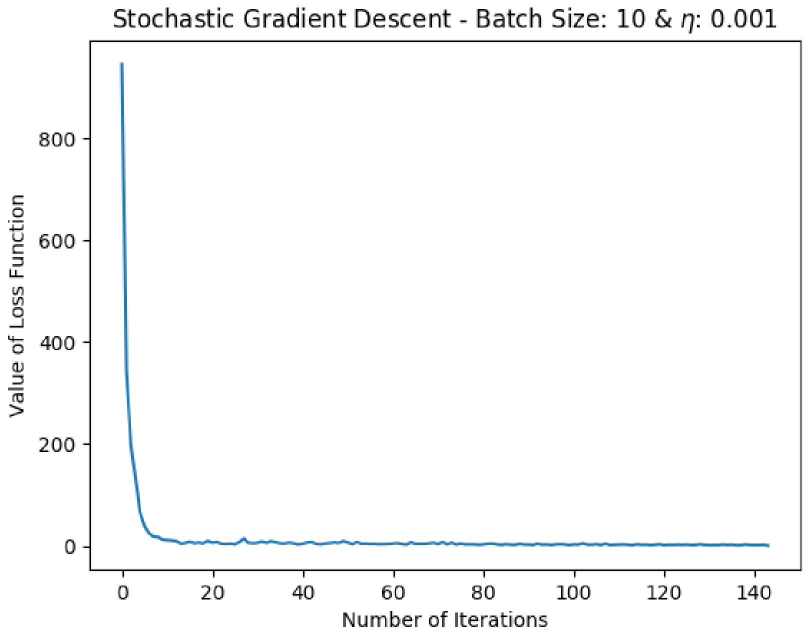
Number of iterations to stop on first run= 128



Number of iterations to stop on second run= 101



Number of iterations to stop on third run= 144



From these results, we cannot see the effect of the batch size. There is not much difference between the number of iterations taken to reach a solution within the threshold. Therefore, we can conclude that for our data set, batch sizes between 10 and 20, do not yield results that are too different from each other, for stochastic descent.