**Name: Taiman Arham Siddiqui**

**CS 5900 HW Assignment 2**

**2.** Simple Regression:

Results:

Params GD: [254449.99982048 93308.92004027]

Params SGD: [254550.06767351 92946.32643744]

Params Normal eq: [254450. 93308.9201061]

The parameters obtained from the Gradient Descent is similar to that obtained from the Normal equations.

Plot of J(w) vs. the number of epochs in increments of 10 for batch and stochastic gradient descents:

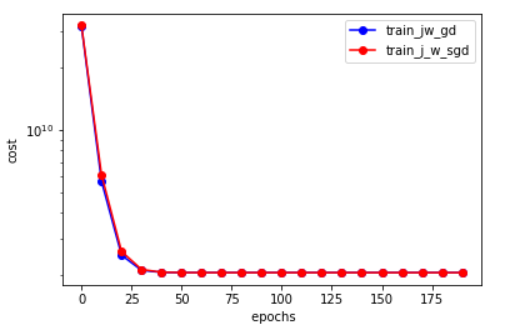


Fig: gd\_cost\_simple.png

Both plots are similar, although batch gradient descent seems to be minimizing the cost faster (in terms of epochs). The overlap of the two plots is why the blue curve is not visible at most points.

Plot of linear approximation of each training method as well as test and training data:

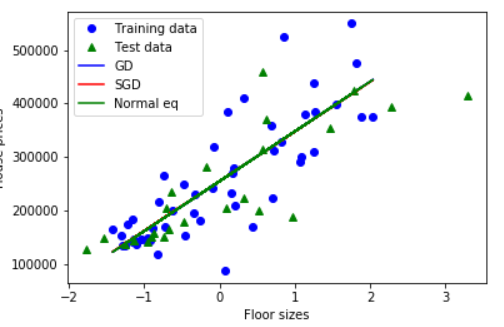


Fig: train-test-line.png

The linear approximation for all methods are similar, causing overlap of the lines in the figure above , which is why the different line colors for each method are not visible.

**3.** Multiple Regression:

Results:

Params GD: [254449.99982048 78079.18106675 24442.5758378 2075.95636731]

Params SGD: [254580.71771403 77853.15388901 24170.61766679 1508.27966708]

Params Normal eq: [254450. 78097.1125582 24424.59918938 2079.70848716]

The parameters obtained from the Normal equations is slightly bigger than that obtained from Gradient Descent.

Plot of J(w) vs. the number of epochs in increments of 10 for batch and stochastic gradient descents:

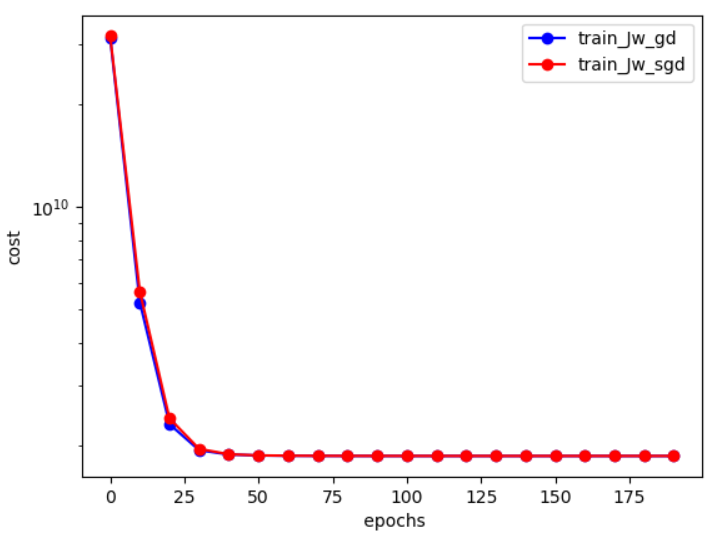


Fig: gd\_cost\_multiple.png

This graph is similar to the corresponding one obtained for Simple Regression.

**4.** Polynomial Curve Fitting:

Plot of dataset.txt: Plot of train.txt:

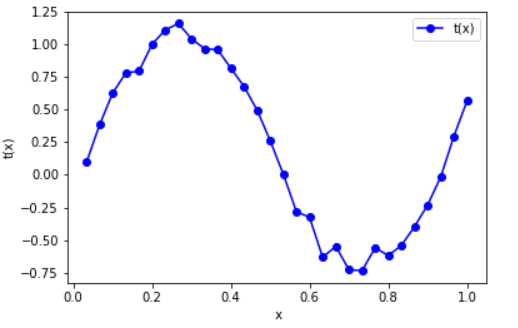
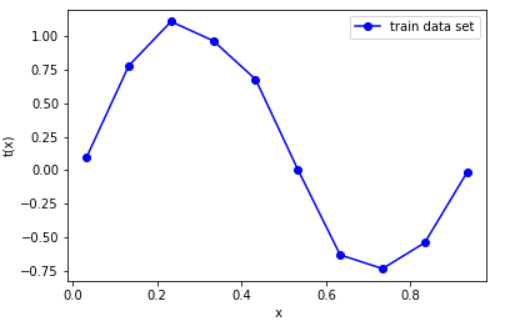
 

Fig: dataset.png Fig: traindataset.png

Plot of devel.txt: Plot of test.txt:

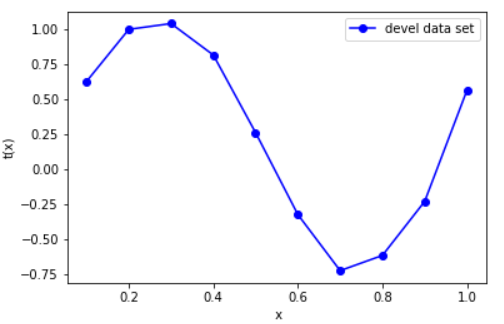
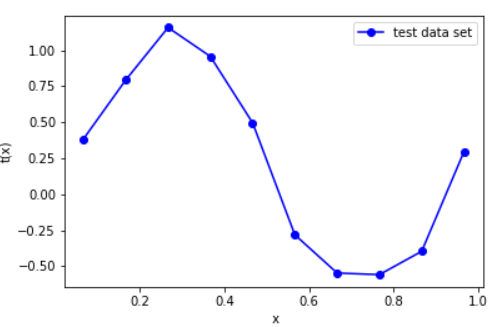
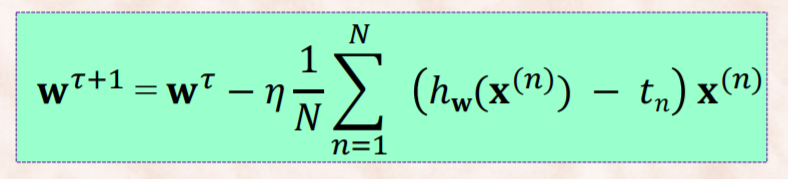
 

Fig: develdataset.png Fig: testdataset.png

**c)** The gradient update:

 ****

**d)**

**1.** Plot of J(w) vs. epochs for given range of learning rates for M=5:

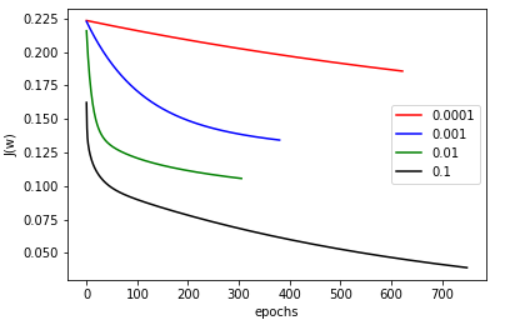
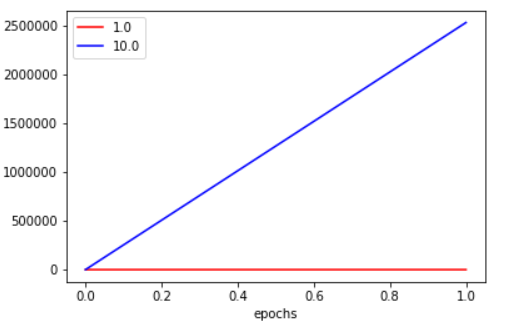
 

Fig: Epoch-vs-Jw-without-reg.png Fig: Epoch-vs-Jw-without-reg2.png

The plots for learning rates of 1.0 and 10.0 are shown separately as they are anomalous as compared to the other plots.

Plot of J(w) vs epochs for learning rate=0.1 for Gradient Descent methods:

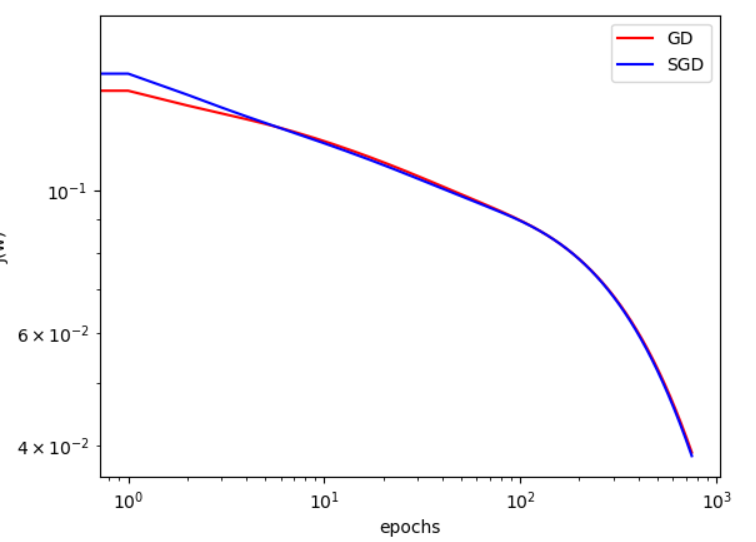


Fig: Epoch-vs-J-without-reg-eta-01.png

Results:

Params GD: [ 0.17105746 0.63102635 -1.41468986 -0.95661773 0.14412845 1.26837826]

Params SGD: [ 0.18119596 0.63138496 -1.43948595 -0.96174437 0.15650972 1.29034043]

Params Normal eq: [ 0.17105746 1.60864309 5.83533977 -35.12355592 43.55862865

-16.11708996]

The parameters for the Normal equations are larger in magnitude than that for Gradient Descent.

**2.** Plot of J(w) vs. epochs for given range of learning rates:

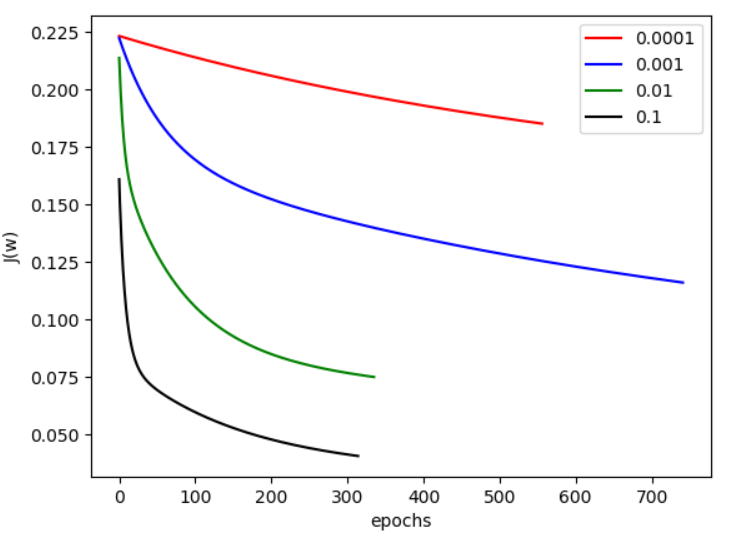
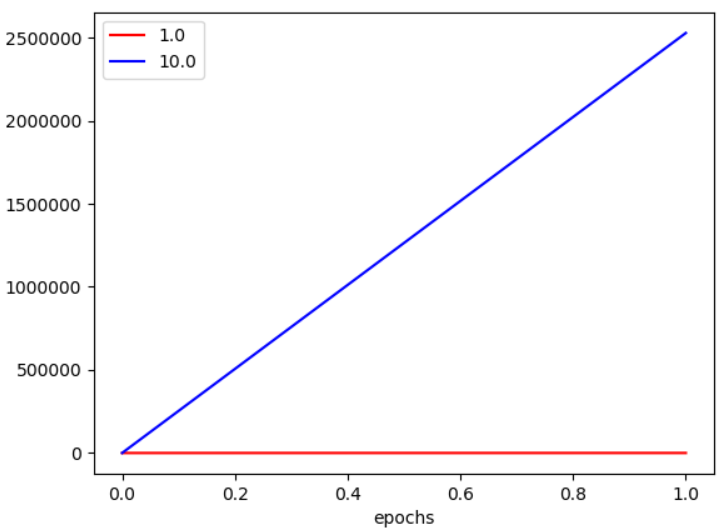
 

Fig: Epoch-vs-Jw-with-reg.png Fig: Epoch-vs-Jw-with-reg.png

Plot of J(w) vs epochs for learning rate=0.1 for Gradient Descent methods:

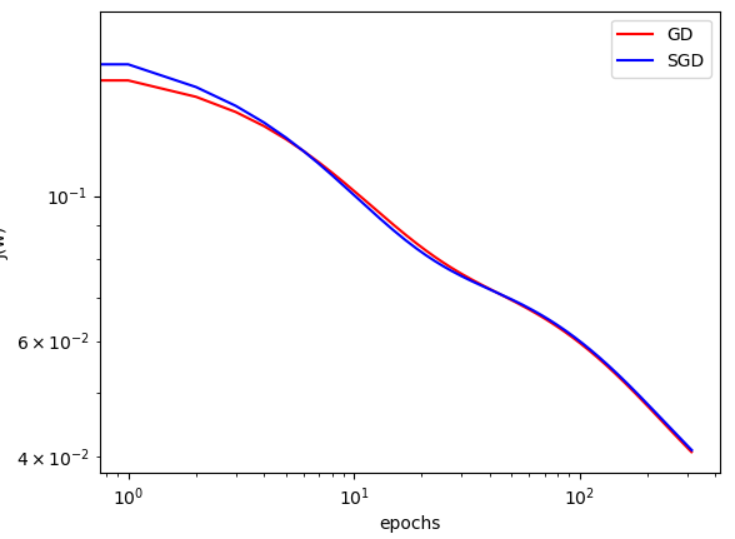


Fig: Epoch-vs-J-with-reg-eta-01.png

Results:

Params GD: [ 0.17104969 0.34994191 -0.63230616 -0.63528569 -0.389097 -0.12687191

0.08924614 0.25105582 0.36598271 0.4444359 ]

Params SGD: [ 0.17517104 0.34097736 -0.64248211 -0.63224561 -0.37921524 -0.11688335

0.09512865 0.25066153 0.3584307 0.42959084]

Params Normal eq: [ 0.17104969 2.72084711 -4.40680131 -2.37752723 0.52755336 2.1528053

2.29098436 1.31416155 -0.30298511 -2.1576861 ]

The parameters for the Normal equations have slightly larger magnitudes than that for Gradient Descent.

As observed from the plots above, SGD seems to take fewer epochs to arrive at the same parameters as GD. How the difference between the number of epochs is relatively small between GD and SGD to arrive at the same parameters.