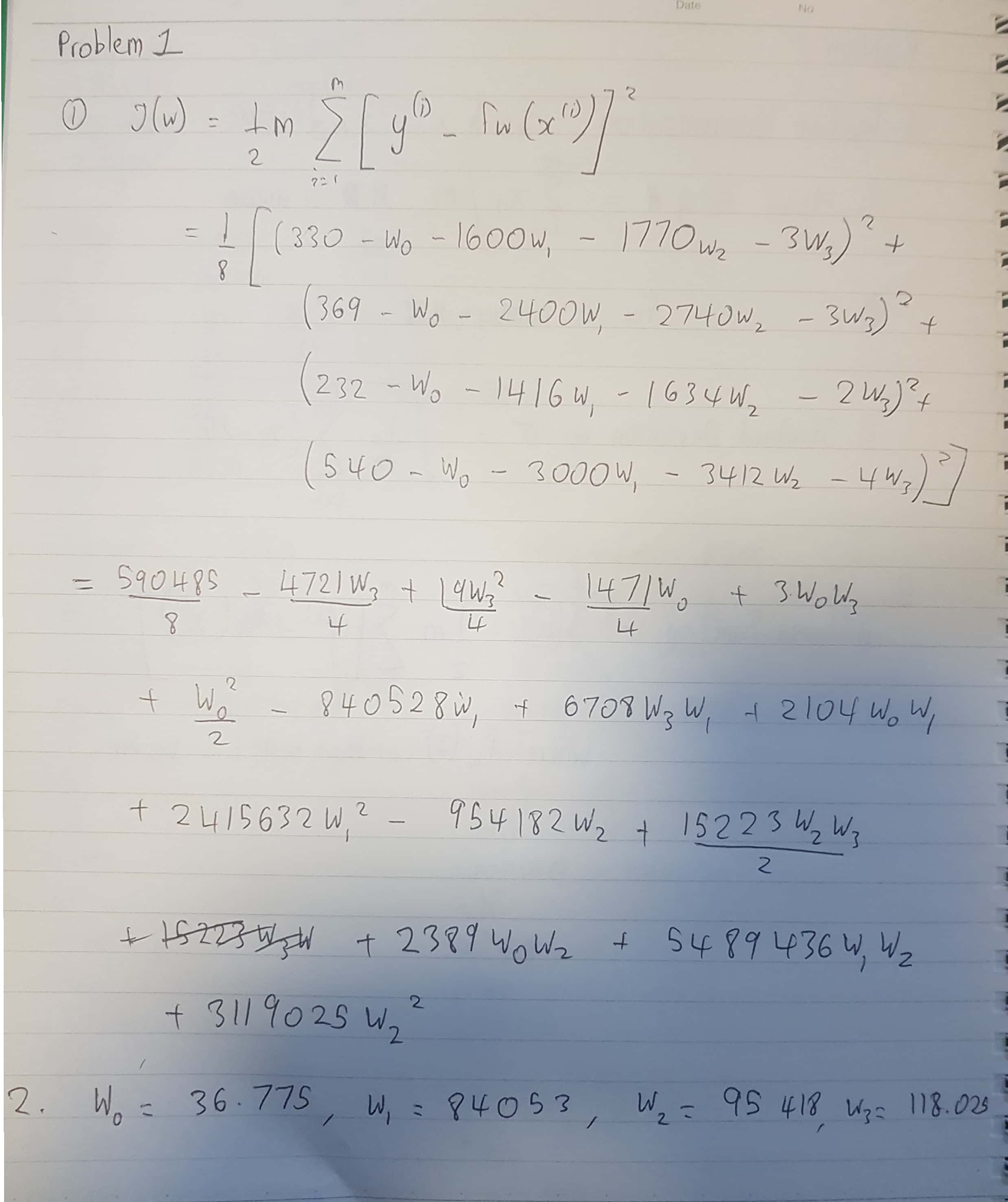
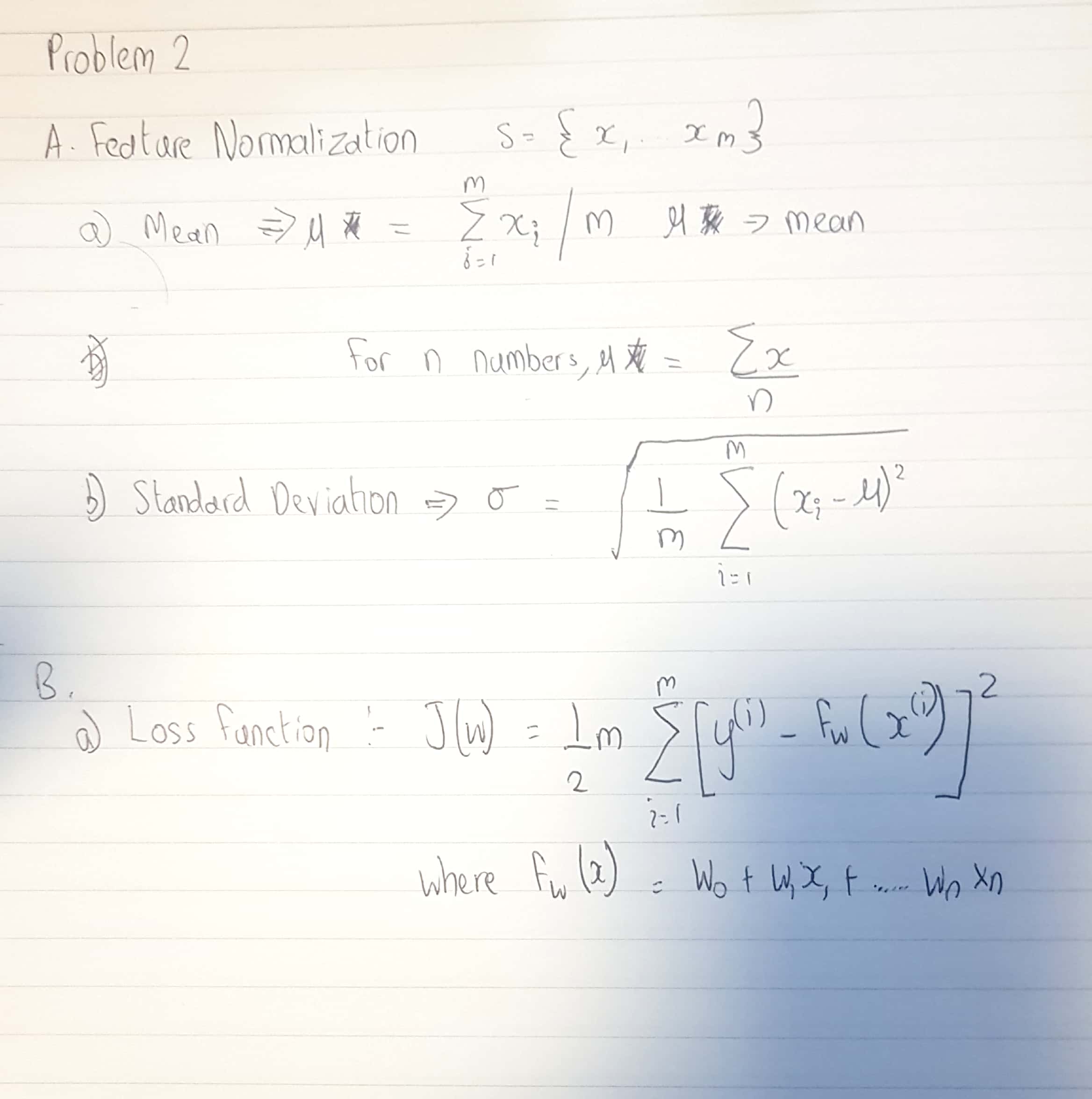
Gradient Descent

Abdullah Zameek (arz268)

Q1.   
  
Q2.



B.

The following graph was obtained for learning rates of 0.01, 0.1, and 0.2 iterations.

|  |  |  |  |
| --- | --- | --- | --- |
|  | 0.01 | 0.1 | 0.2 |
| 10 | 1.705103 | 0.75306 | 0.594852 |
| 20 | 1.455809 | 0.598991 | 0.542142 |
| 30 | 1.268001 | 0.55879 | 0.535238 |
| 40 | 1.12561 | 0.543818 | 0.534269 |
| 50 | 1.016851 | 0.537947 | 0.534132 |
| 60 | 0.933075 | 0.535628 | 0.534113 |
| 70 | 0.867924 | 0.534711 | 0.53411 |
| 80 | 0.81672 | 0.534348 | 0.53411 |
| 90 | 0.776016 | 0.534204 | 0.53411 |
| 100 | 0.743264 | 0.534147 | 0.53411 |

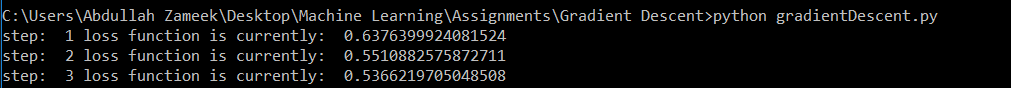
The following graph was obtained for learning rates of 0.03 and 0.5

|  |  |  |
| --- | --- | --- |
|  | 0.03 | 0.5 |
| 10 | 1.296449 | 0.535202 |
| 20 | 0.942947 | 0.534113 |
| 30 | 0.779125 | 0.53411 |
| 40 | 0.695167 | 0.53411 |
| 50 | 0.647016 | 0.53411 |
| 60 | 0.616441 | 0.53411 |
| 70 | 0.595498 | 0.53411 |
| 80 | 0.580435 | 0.53411 |
| 90 | 0.569291 | 0.53411 |
| 100 | 0.560917 | 0.53411 |

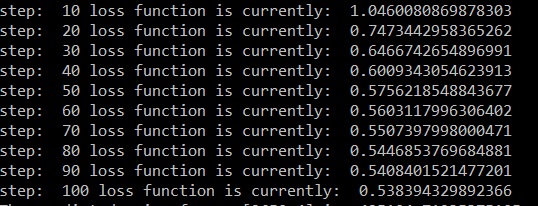
Since it can be clearly seen that the graph for a learning rate of 0.5 converges faster than the rest, a learning rate of 0.5 is optimal for this setting.

C.   


The predicted price for the given value is approx. 423,554

D.  


The output of the loss function seems to converge faster in the stochastic gradient descent algorithm, compared to the regular gradient descent algorithm in much fewer steps for the same learning rate of 0.05. (3 versus 100 steps)



The output for gradient decent with upto 100 steps seems to be similar to the output for the SGD, except for the fact that the SGD algorithm reached that value much faster.

Problem 3.

