HW4 CS289A

Antonio Rohit De Lima Fernandes

Question 1:

4. Using a learning rate of 0.5

a. $\mu^{(0)}$ =

0.9526

0.7311

0.7311

0.2689

b. $\beta^{(1)}$ =

-1.0008

1.8101

-2.6092

c. $\mu^{(1)}$ =

0.9438

0.8606

0.3102

0.1419

d. $\beta^{(2)}$ =

-0.8722

2.2867

-3.6905

Question 2:

1. Code (bCourses)
2. Min Predicted Value: 1953.85

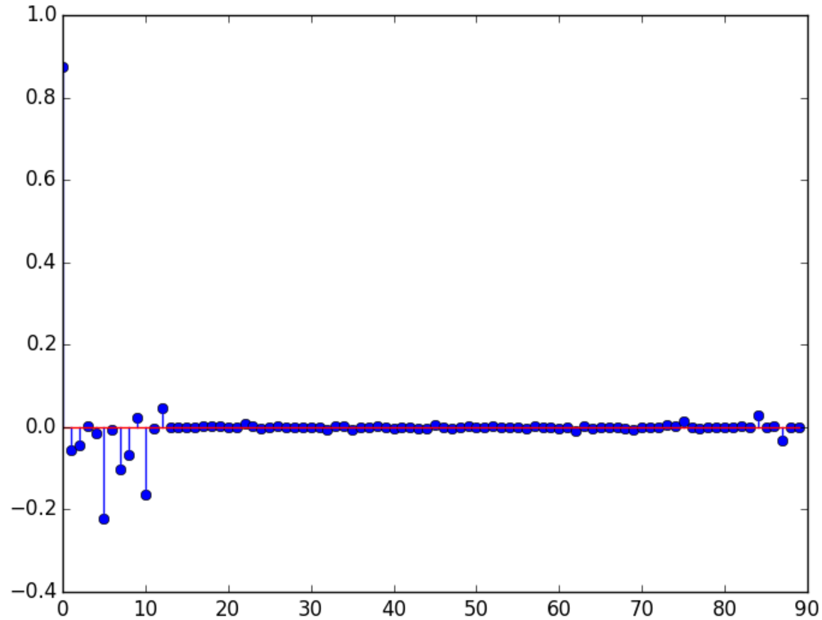
   Max Predicted Value: 2045.55

   They are close to what the values should be, but not exactly.

RSS: 467M

3. β_0: 1951.12

   Rest of β values:

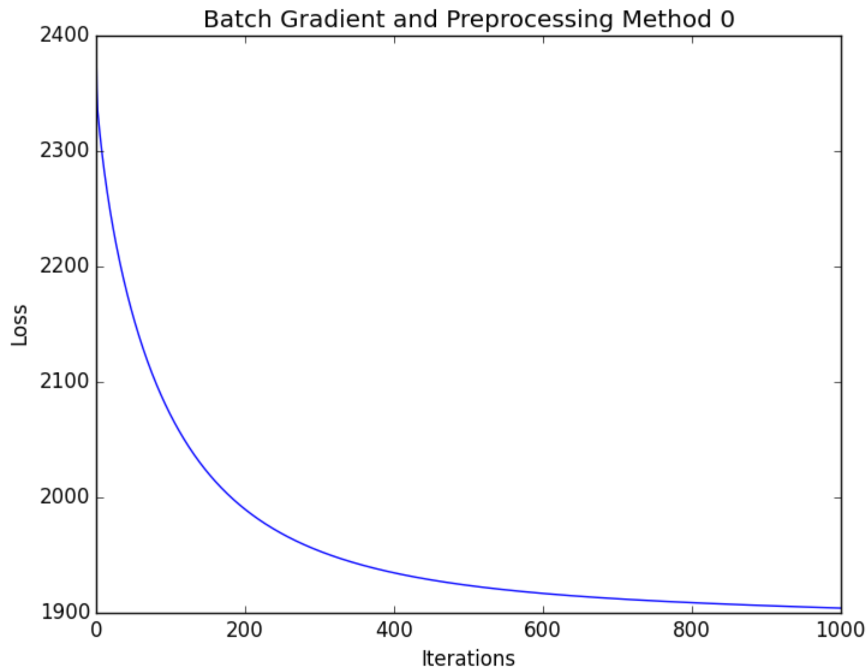


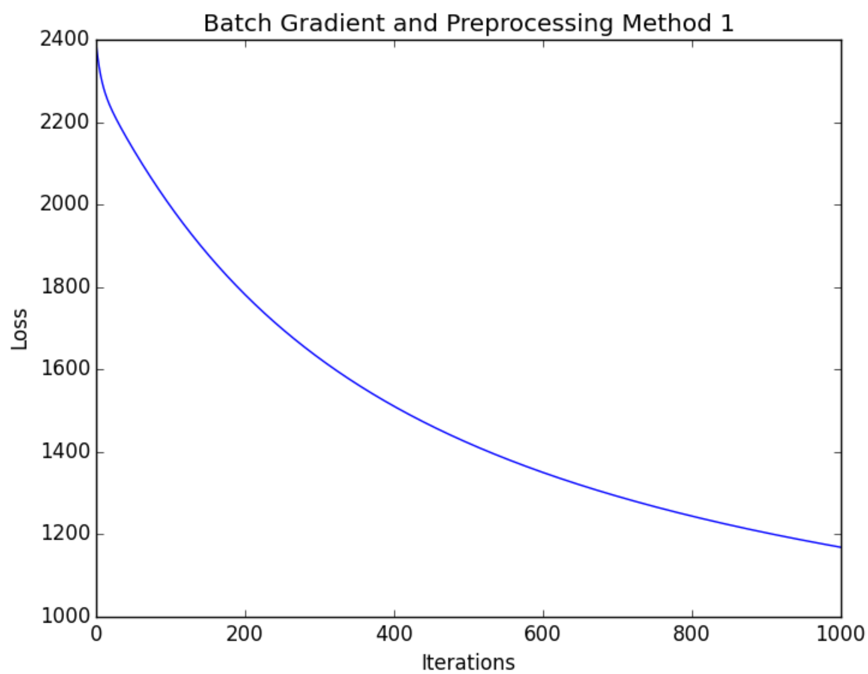
Question 3:

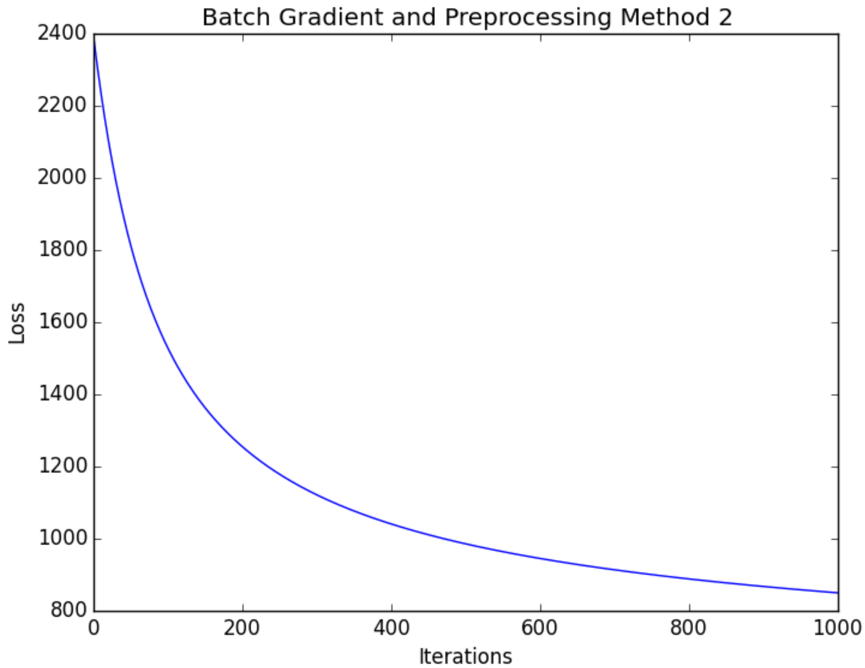
1. **Batch Gradient Descent**

$$\beta^{(t+1)} = \beta^{(t)} + \eta \nabla_\beta \ell\left(\beta^{(t)}\right)$$

Using the gradient derived in question 1:


Batch Gradient and Preprocessing Method 0
Loss
Iterations
2400
2300
2200
2100
2000
1900
0
200
400
600
800
1000

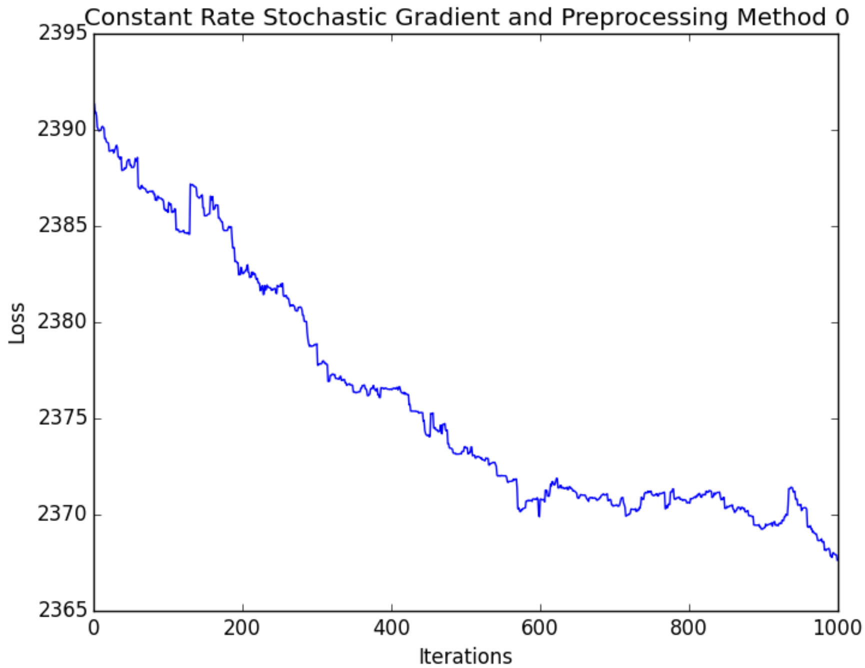

Batch Gradient and Preprocessing Method 1
Loss
Iterations
2400
2200
2000
1800
1600
1400
1200
1000
0
200
400
600
800
1000

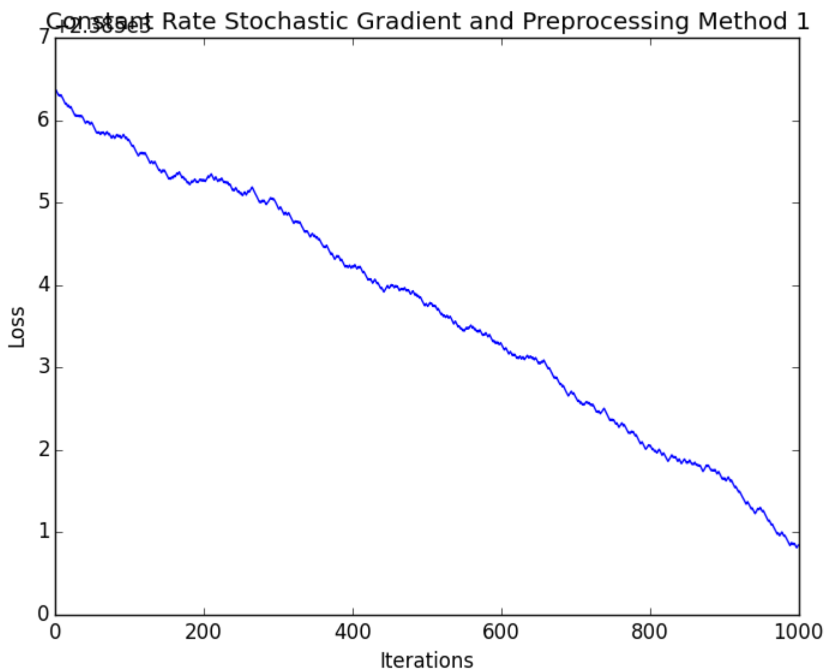


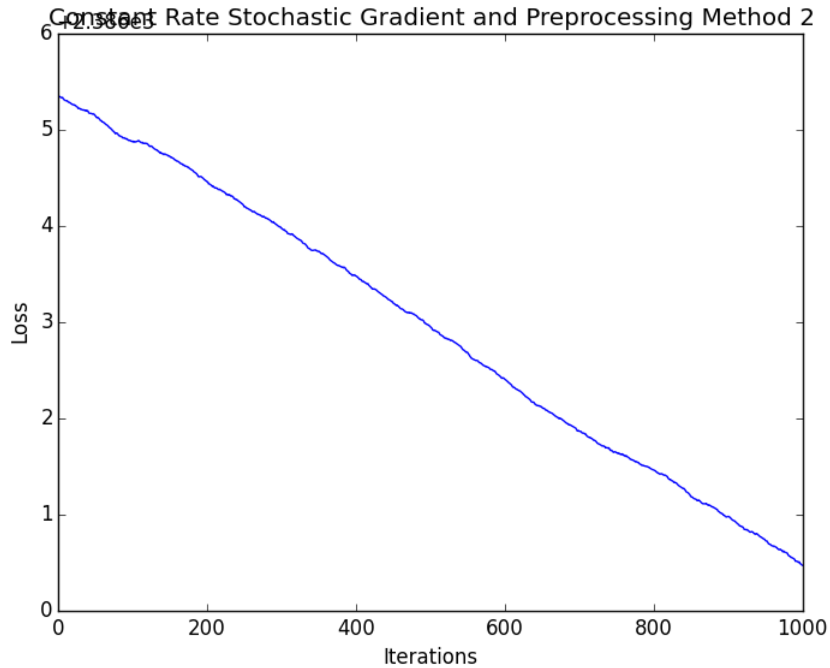
2. **Stochastic Gradient with constant learning rate**

$$\beta^{(t+1)} = \beta^{(t)} + \eta \left(y_{i_t} - \mu_{i_t}(\beta^{(t)}) \right) x_{i_t}$$

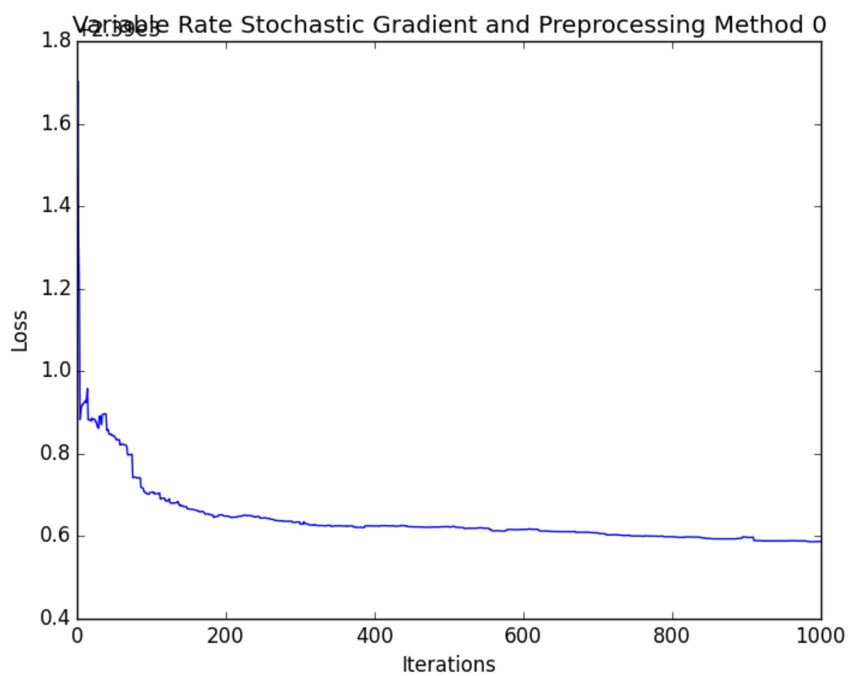
The graphs clearly look different from the batch gradient method above: The curve is not smooth, and does not decrease monotonically, though it does decrease over many iterations. This is because the gradient taking a single a single point may not point in the right direction, but over many samples, averages out in the right direction.

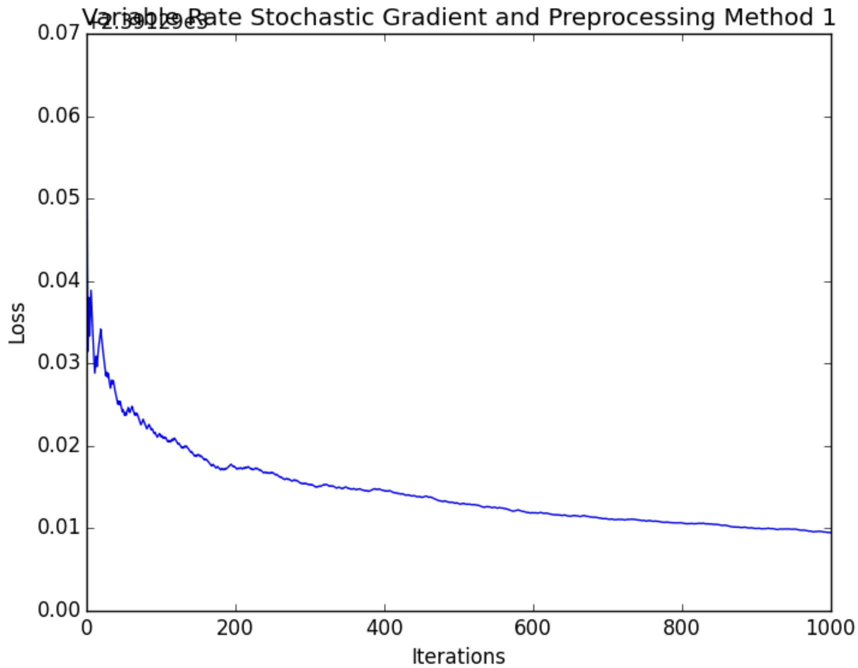

Constant Rate Stochastic Gradient and Preprocessing Method 0
Loss
Iterations
2395
2390
2385
2380
2375
2370
2365
0
200
400
600
800
1000

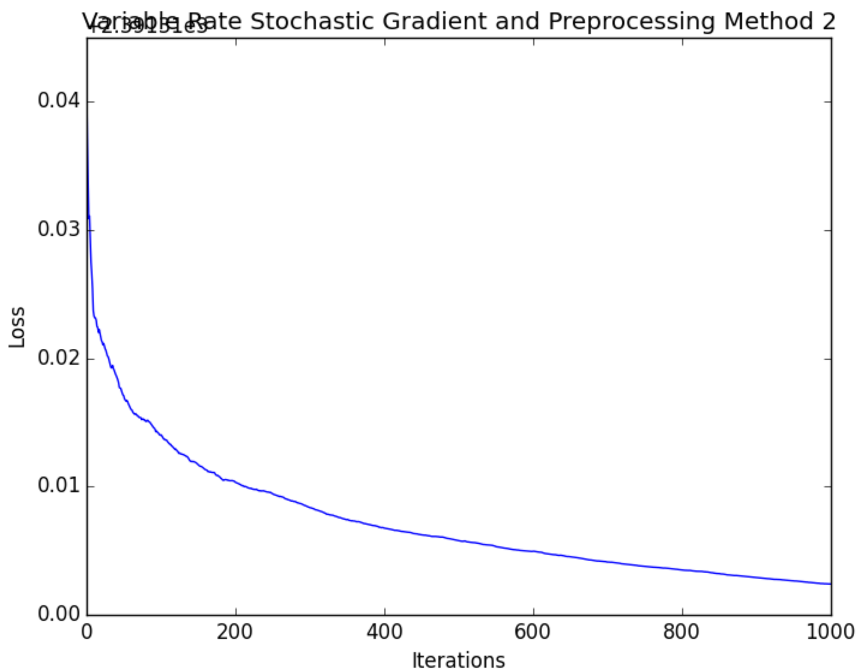

Constant Rate Stochastic Gradient and Preprocessing Method 1
+2.389e5
Loss
Iterations
7
6
5
4
3
2
1
0
0
200
400
600
800
1000



3. **Stochastic Gradient with Variable Learning Rate**
   Using a variable learning rate seems to smoothen the graphs somewhat. I believe this to be a good idea so that once we are close to the minimum, the steps are smaller, and the code converges to this minimum.




Variable Rate Stochastic Gradient and Preprocessing Method 1
+2.39129e3
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
Loss
0
200
400
600
800
1000
Iterations


Variable Rate Stochastic Gradient and Preprocessing Method 2
+2.39151e3
0.04
0.03
0.02
0.01
0.00
Loss
0
200
400
600
800
1000
Iterations