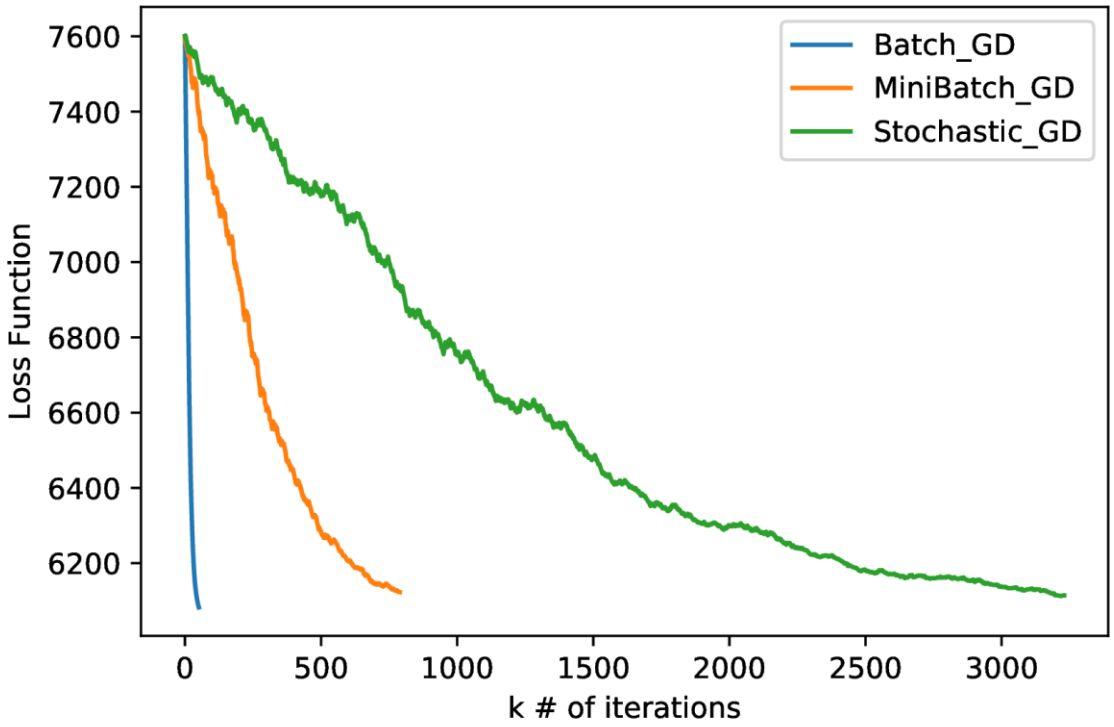
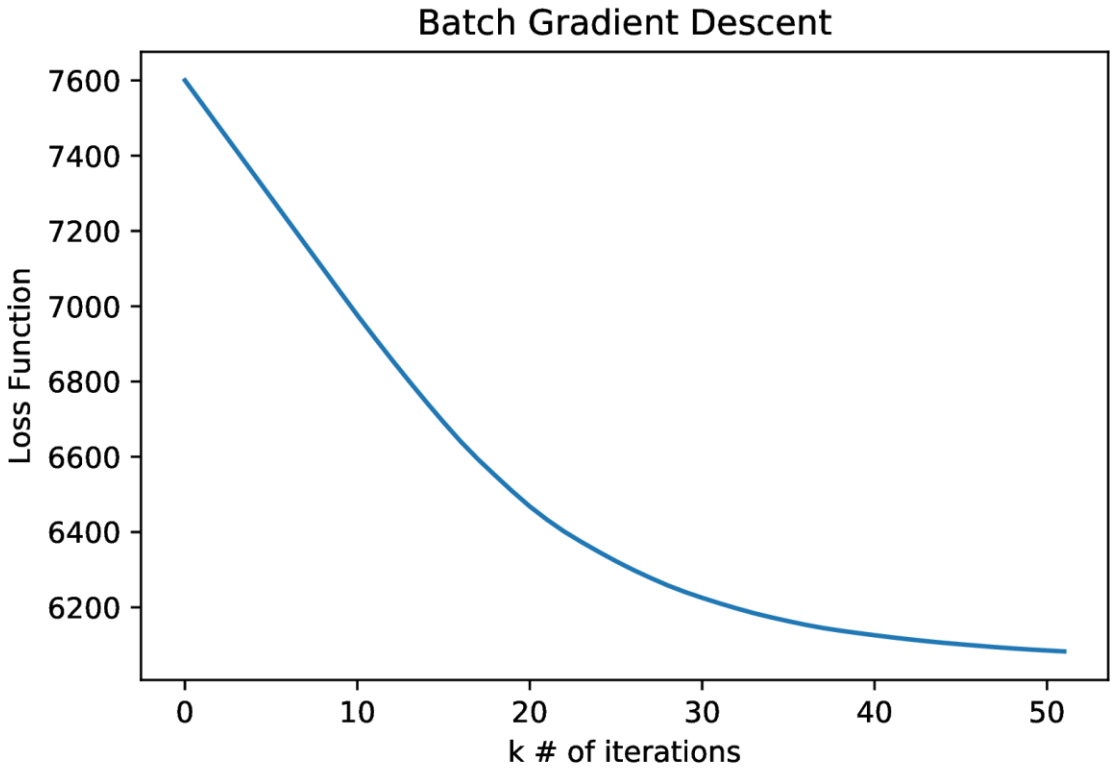
1. Gradient of the cost function respect to b for the Batch Gradient Descent is equal to

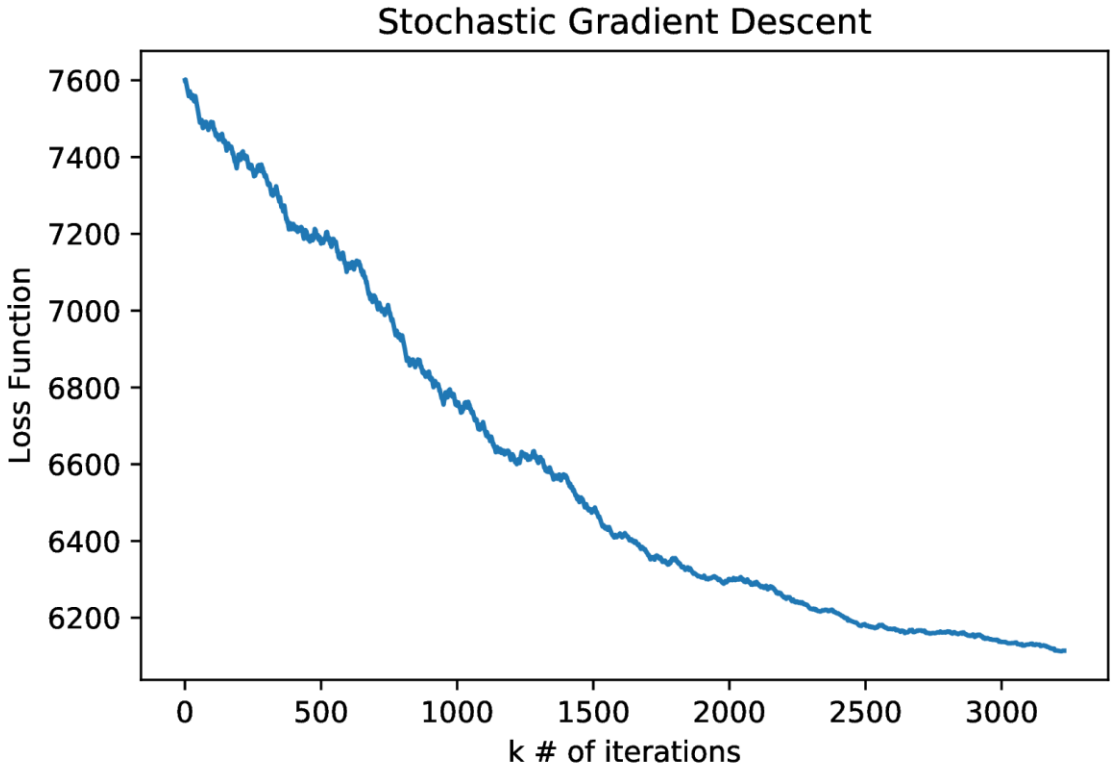
$$\nabla_b J(\boldsymbol{w}, b) = \frac{\partial J(\boldsymbol{w}, b)}{\partial b} = C \sum_{i=1}^{m} \frac{\partial L\left(x^{(i)}, y^{(i)}\right)}{\partial b}$$

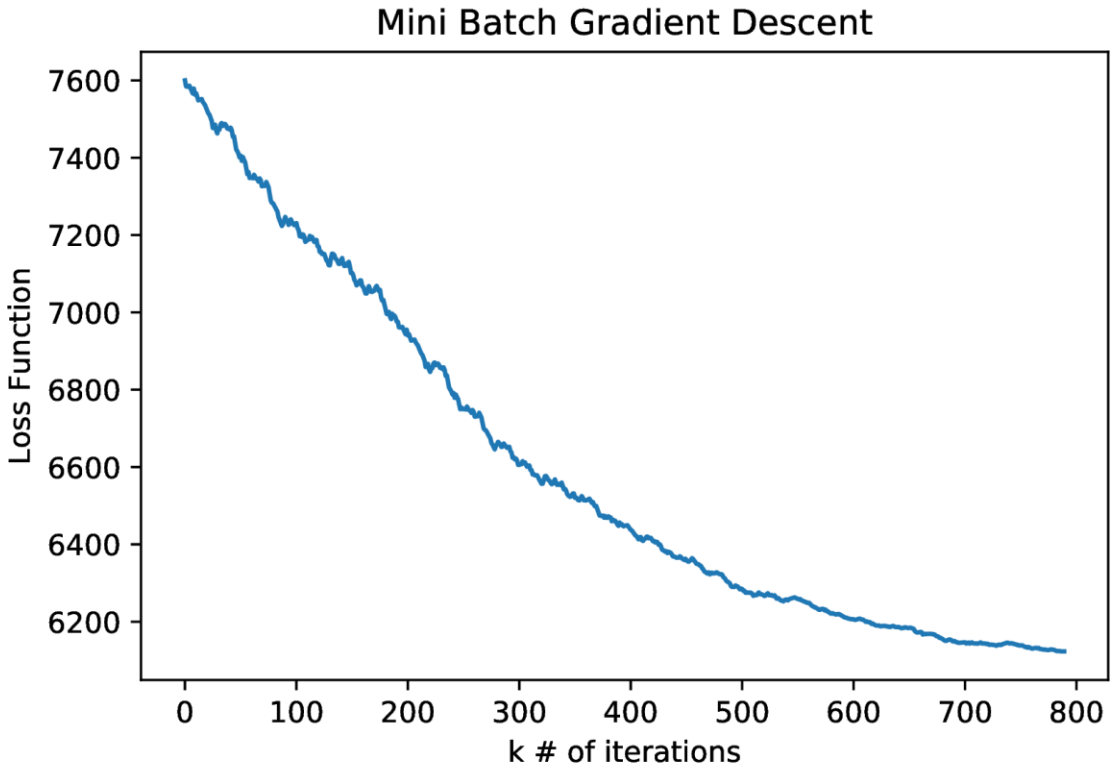
$$\frac{\partial L\left(x^{(i)}, y^{(i)}\right)}{\partial b} = \begin{cases} 0, & y^{(i)}(\boldsymbol{x}^{(i)}\boldsymbol{w} + b) \geq 1 \\ -y^{(i)}, & otherwise \end{cases}$$

2. Plots of Cost function vs the number of iterations (k) for all methods together a




Batch Gradient Descent
Loss Function
7600
7400
7200
7000
6800
6600
6400
6200
0
10
20
30
40
50
k # of iterations


Stochastic Gradient Descent
Loss Function
7600
7400
7200
7000
6800
6600
6400
6200
0
500
1000
1500
2000
2500
3000
k # of iterations



3. Batch Gradient Descent needs 52 iterations to converge. Stochastic Gradient Descent needs 3231 iterations to converge. Mini Batch Gradient Descent need 790 iterations to converge. Number of required iterations for Mini Batch and Stochastic methods is subjected to change since for every run we shuffle the data.
4. The python source codes are attached separately.