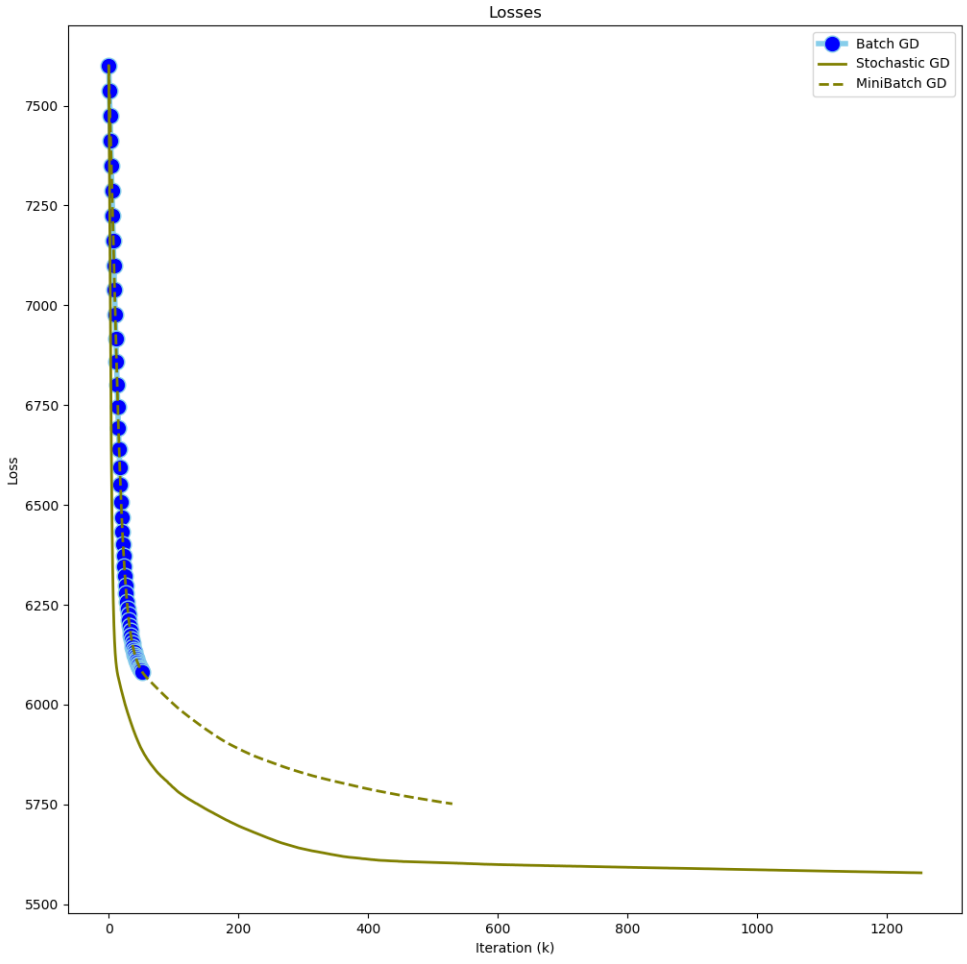
Problem a

Gradient with respect to b

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

Problem b

- Comparison of Batch, Stochastic and Mini-Batch Gradient Descent.



- Convergence Times:
  - Batch Gradient Descent: 1.52661204338 seconds
  - Stochastic Gradient Descent: 35.0849909782 seconds
  - Mini-Batch Gradient Descent: 13.6391267776 seconds