CMPS 142 HW 2

Isaiah Solomon

February 2017

1 Question 1

i)

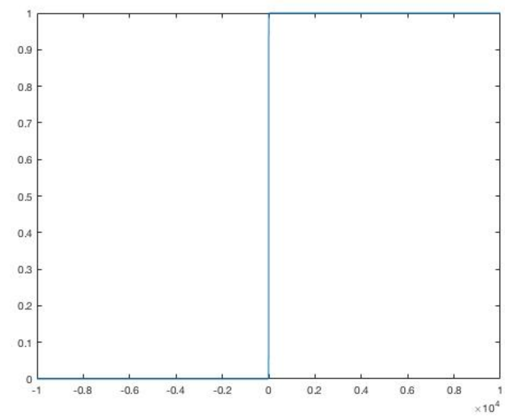
First, we will derive $\frac{d\sigma(z)}{dz} = \sigma(z)(1-\sigma(z))$:

$$\begin{aligned}
\frac{d\sigma(z)}{dz} &= \sigma(z)(1-\sigma(z)) \\
\sigma(z) &= \frac{1}{1+e^{-z}} \\
\frac{d\sigma(z)}{dz} &= \frac{e^{-z}}{(1+e^{-z})^2} \\
&= \sigma(z)(1-\sigma(z))
\end{aligned}$$

Next, we will find the gradient of total loss function w.r.t.:

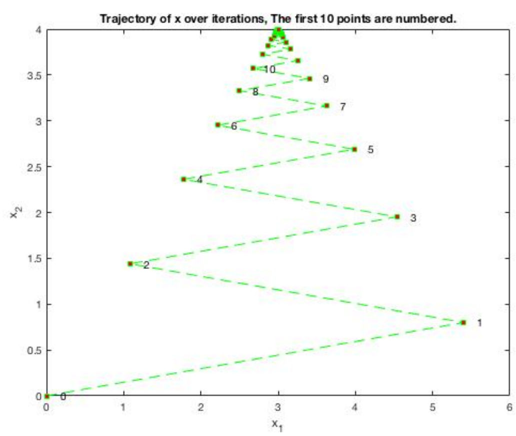
$$\begin{aligned}
\nabla_w l &= X^T(\hat{y} - y) \\
l &= \sum_{i=1}^{t} L(\hat{y}_i, y_i) \\
\nabla_w l &= \nabla_w (\sum_{i=1}^{t} L(\hat{y}_i, y_i)) \\
&= \sum_{i=1}^{t} \frac{\delta L(\hat{y}, y)}{\partial \sigma(z_i)} \frac{\partial \sigma(z_i)}{\partial z_i} \frac{\partial z_i}{\partial w} \\
&= (X^T)(\sigma(z)(1-z))(\frac{\hat{y} - y}{\hat{y} - \hat{y}^2}) \\
&= X^T(\hat{y}(1-\hat{y}))\frac{\hat{y} - y}{\hat{y}(1-\hat{y})} \\
&= X^T(\hat{y} - y)
\end{aligned}$$

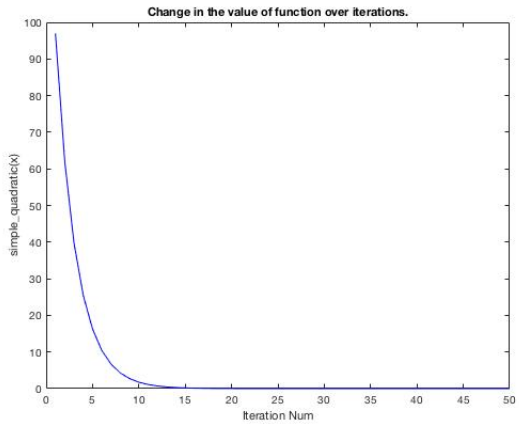
ii)



2 Question 2

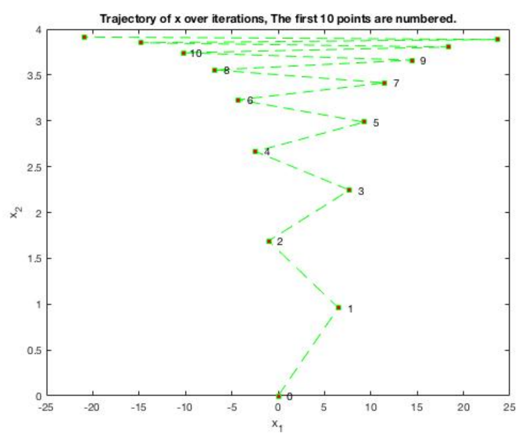
c)



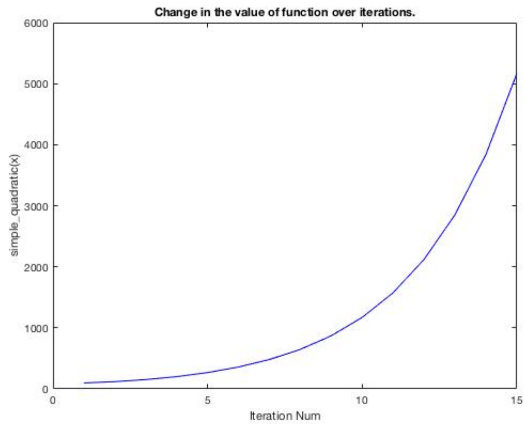


Change in the value of function over iterations.
simple_quadratic(x)
Iteration Num

d)



Trajectory of x over iterations, The first 10 points are numbered.
x_2
x_1



x_2 is converging to the optimal value; however, x_1 is diverging further away. This is due to the step size being larger and fluctuates at a faster pace, thus ends up diverging