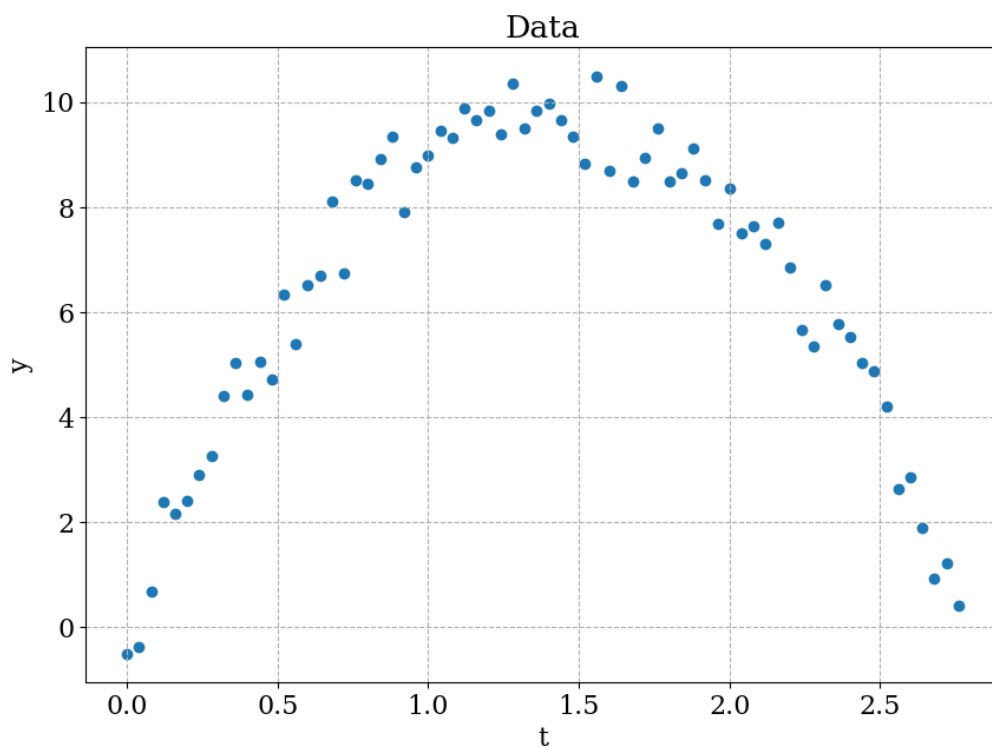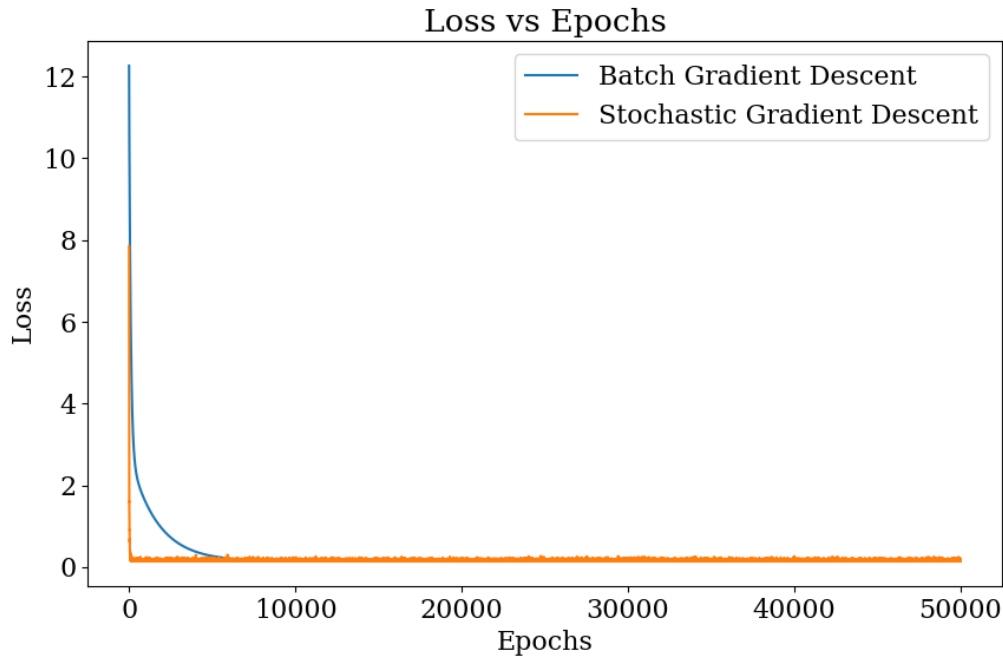# Assignment 2 Plots

Harikesh Kushwaha

Assignment 2

Problem 1

The plot shows time on x axis and height reached by the ball on y axis. We can easily see a parabola. See section 3.1.1 for mathematical explanation.


Loss vs Epochs

Here, we have plotted the loss with number of epoch. Here are some of the observations found and difference between the SGD and BGD model:

1. For BGD, the loss is decreasing only up to 2000 epoch. After that, change in loss is almost negligible.
2. For SGD this starts happening even before 200 epoch!
3. SGD starts with much lower loss that the BGD.
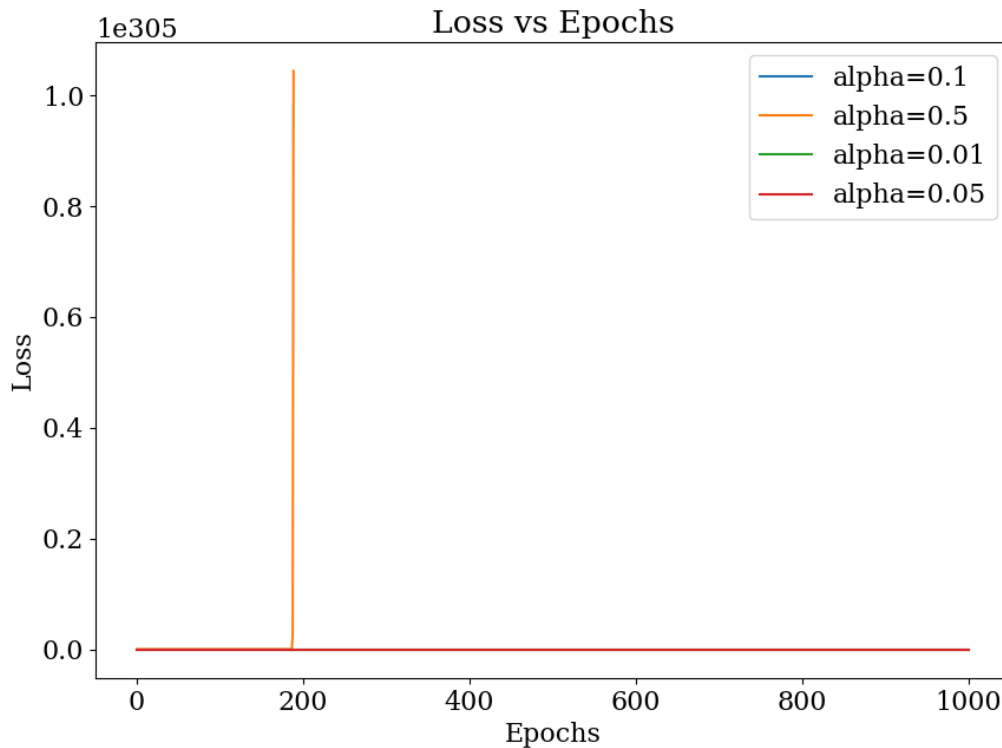4. SGD loss is fluctuating

---

Zoomed in Loss vs Epochs
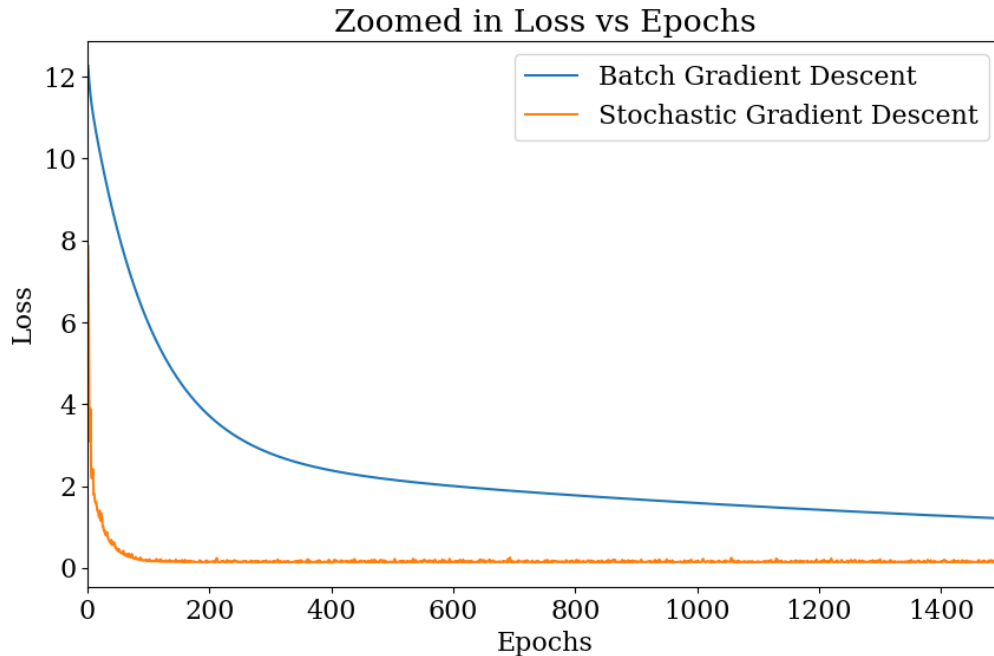
This figure shows the losses at earlier epochs. It is clear from the figure that SGD is converging faster than BGD.

Data with $h(X, \mathbf{w})$

$$y_{sgd} = -5.04t^2 + 14.12t - 0.12$$

$$y_{bgd} = -5.04t^2 + 14.20t - 0.14$$
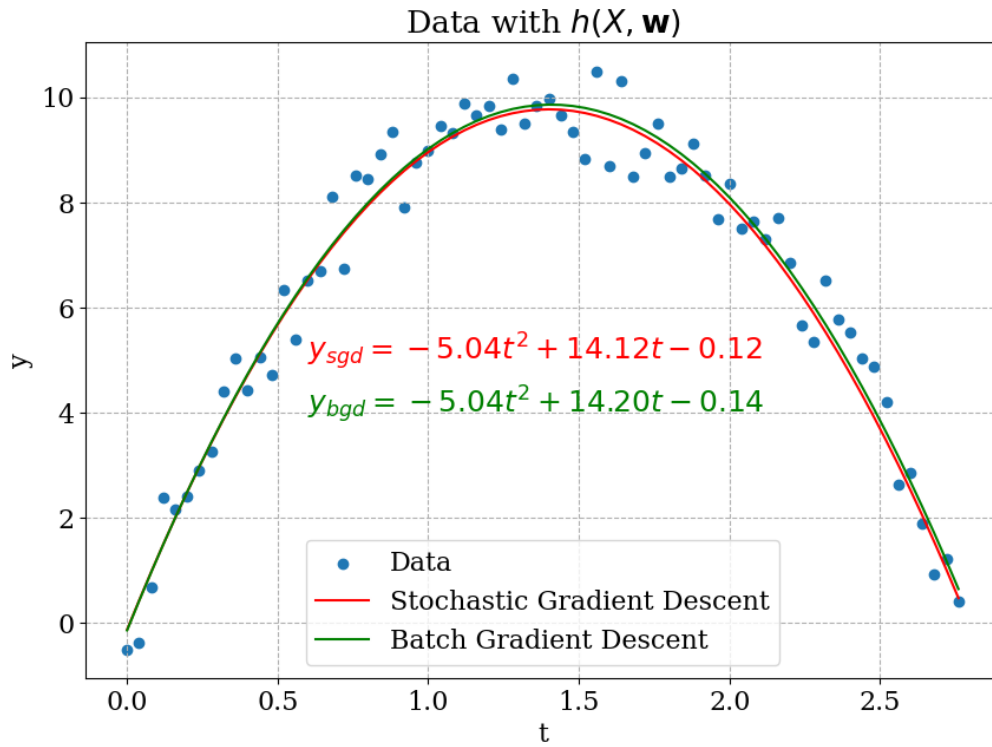
The figure shows the original data with the hypothesis functions learned by the models on top of them. We can see quite a good fit!


Loss vs Epochs

The plot of the loss J with epochs is shown here. It seems we are getting overflow for some values of α. This is happening because the value of α is too high and the gradient descent is diverging. So, this is what happening for $\alpha = 0.5$.
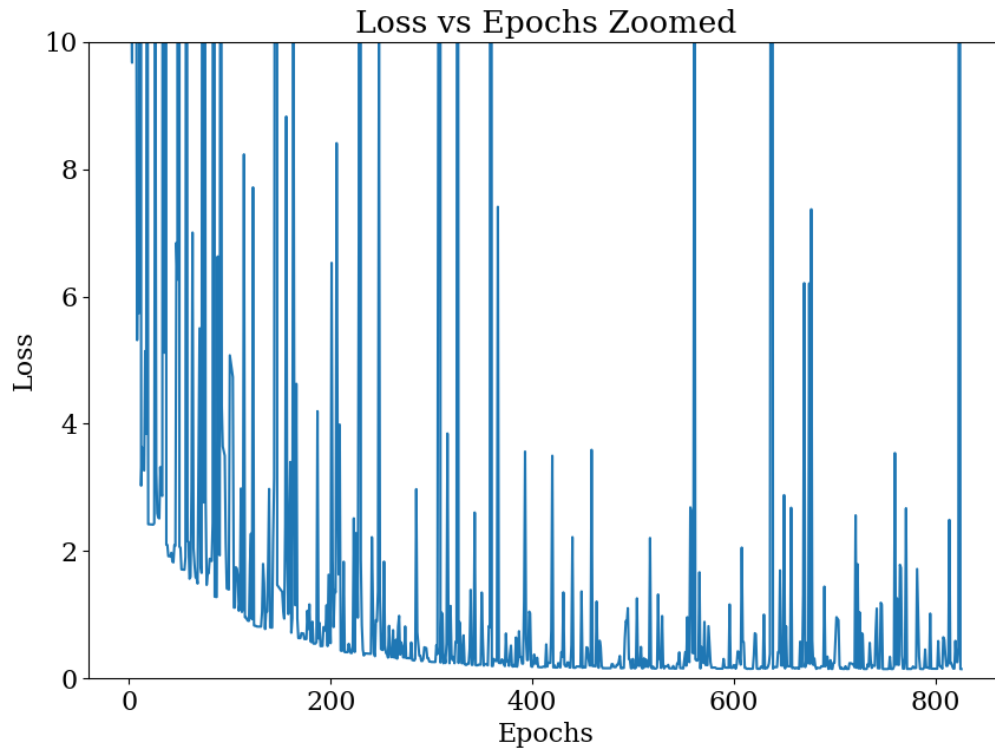
Loss vs Epochs
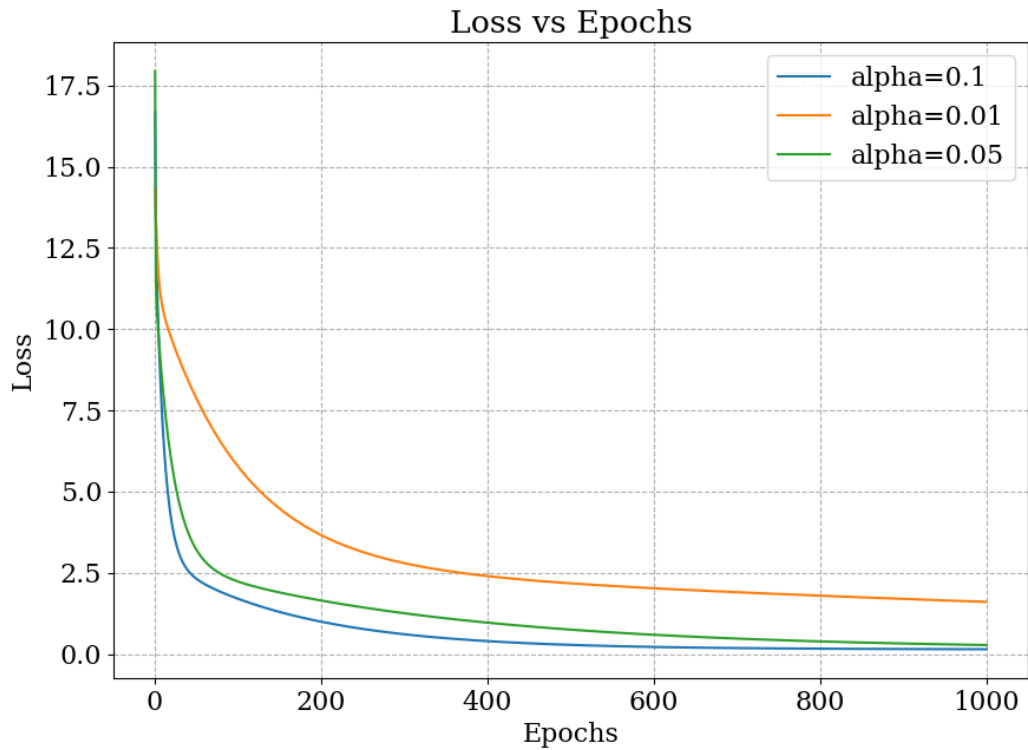
Since the loss is diverging for $\alpha = 0.5$, this plot exclude $\alpha = 0.5$. We can see that the loss is decreasing most steeply for $\alpha = 0.1$
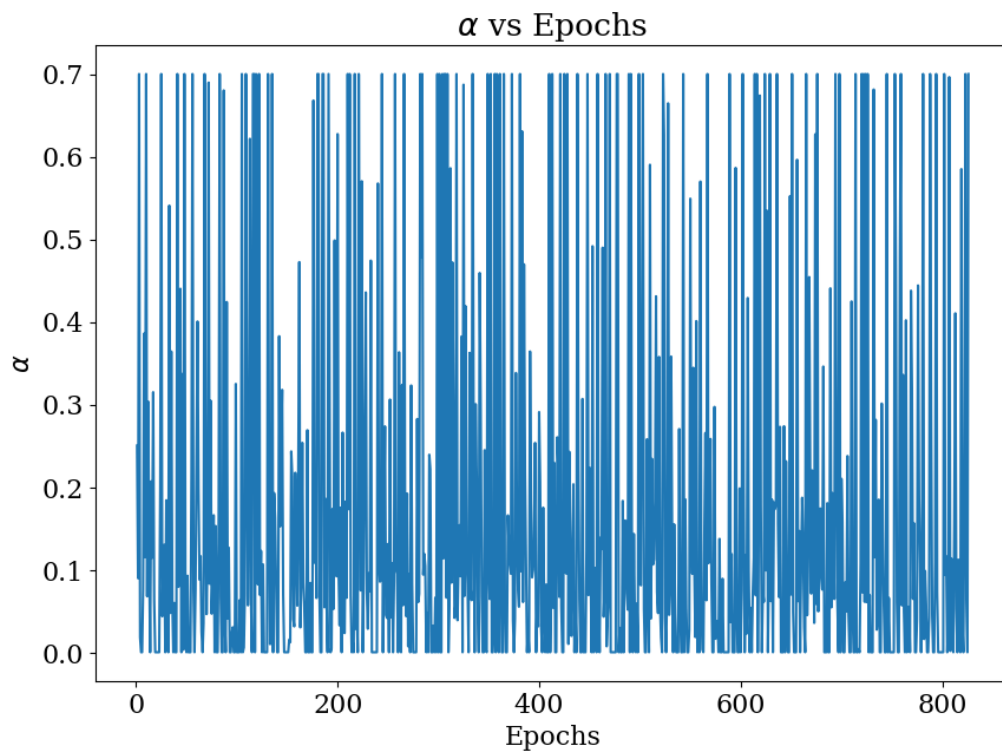

Loss vs Epochs
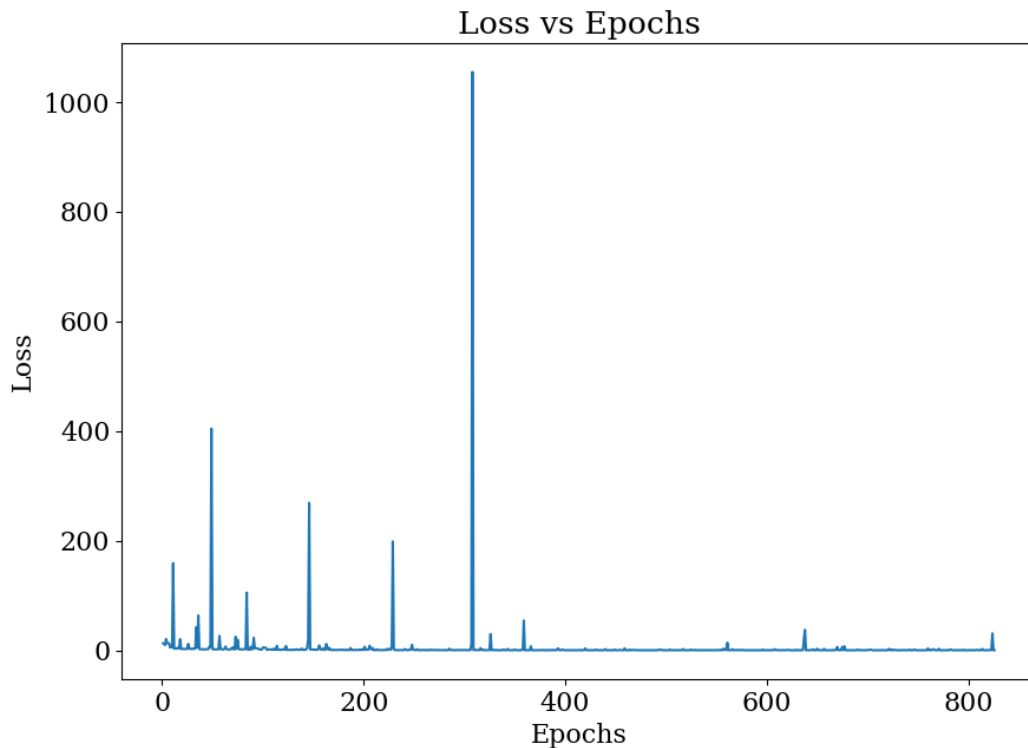
Next, we have the plot of loss with epoch for the method of steepest descent. We can see some

jumps in the loss. This is because for that particular epoch, the α has particulary high value.

Loss vs Epochs Zoomed

Since there were a lot of jumps in the original figure above, this figure sets a limit on the y axis to make the figure more readable.
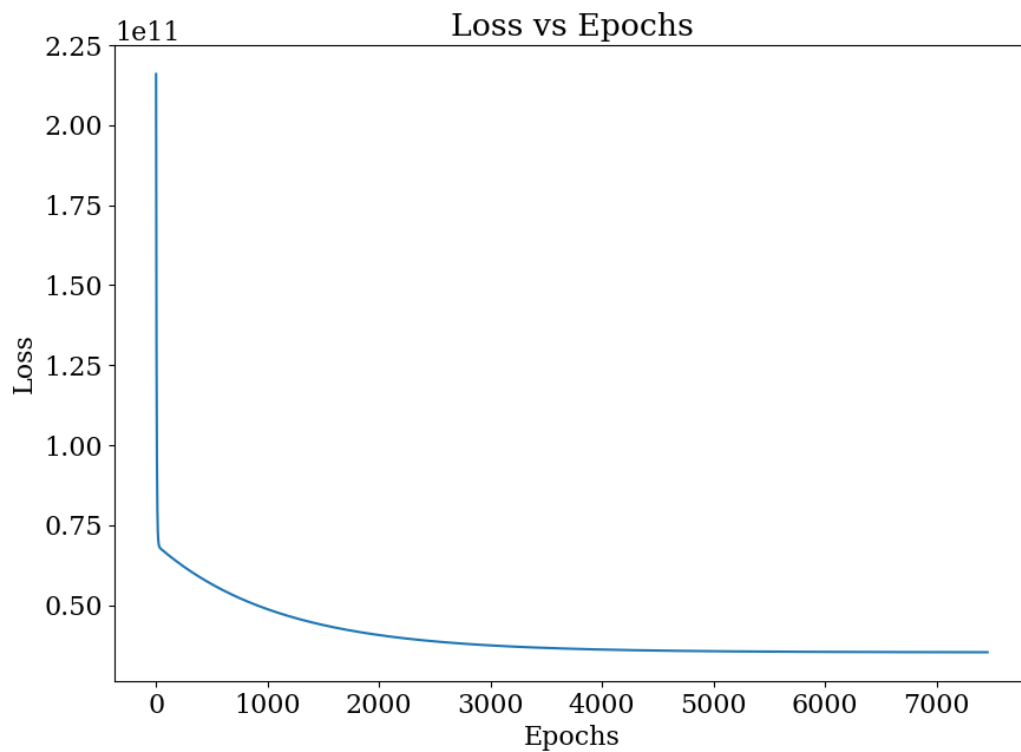

α vs Epochs

Here is the plot of learning rate α with epoch. We can see that there are very few times where the learning rate is not too high, neither too low. This is the reason why the loss is fluctuating so much.

Data with $h(X, \mathbf{w})$

$y_{sd} = -5.00t^2 + 14.06t - 0.06$
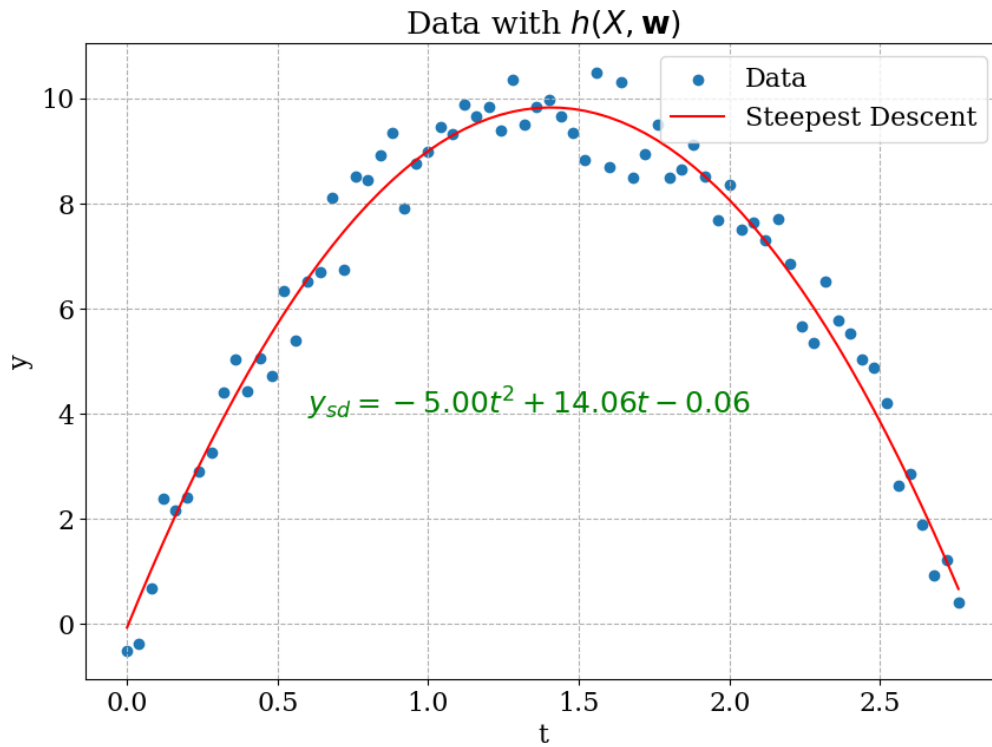
Finally, we have the plot of the hypothesis function learned by the steepest descent model on top of the original data points.
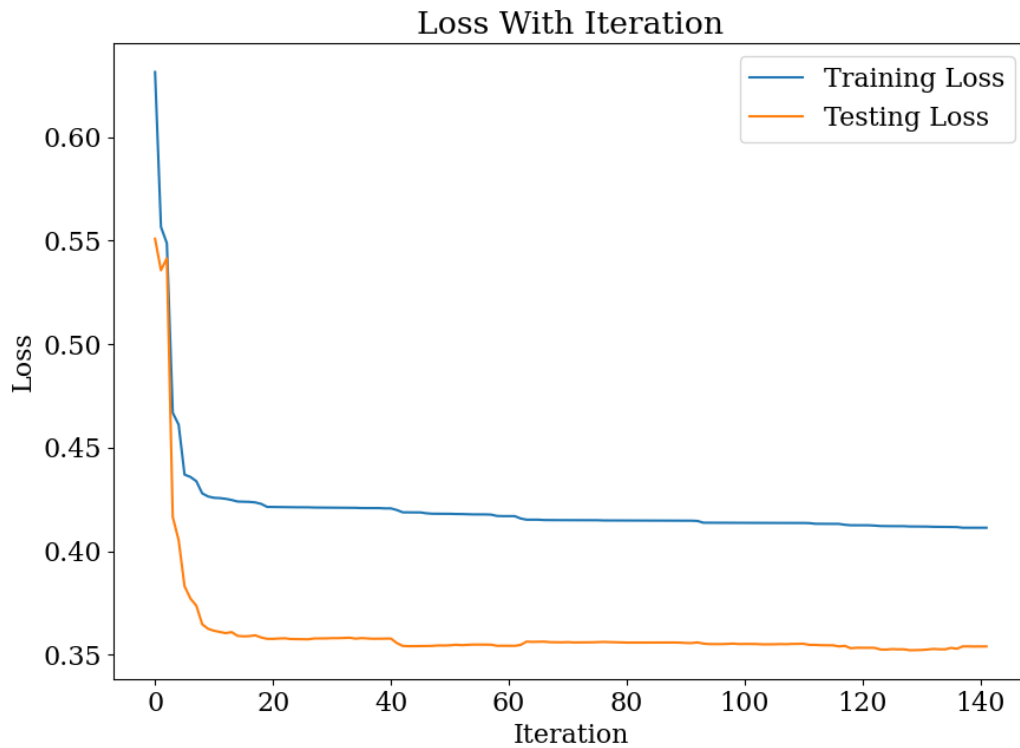

6

Problem 2

Though, we are not required to plot J versus epochs for this problem, I've done it anyway to show how the loss is changing with epochs. We can see that the loss is decreasing very steeply at first then after epoch somewhere near 1500, the loss is decreasing very slowly.
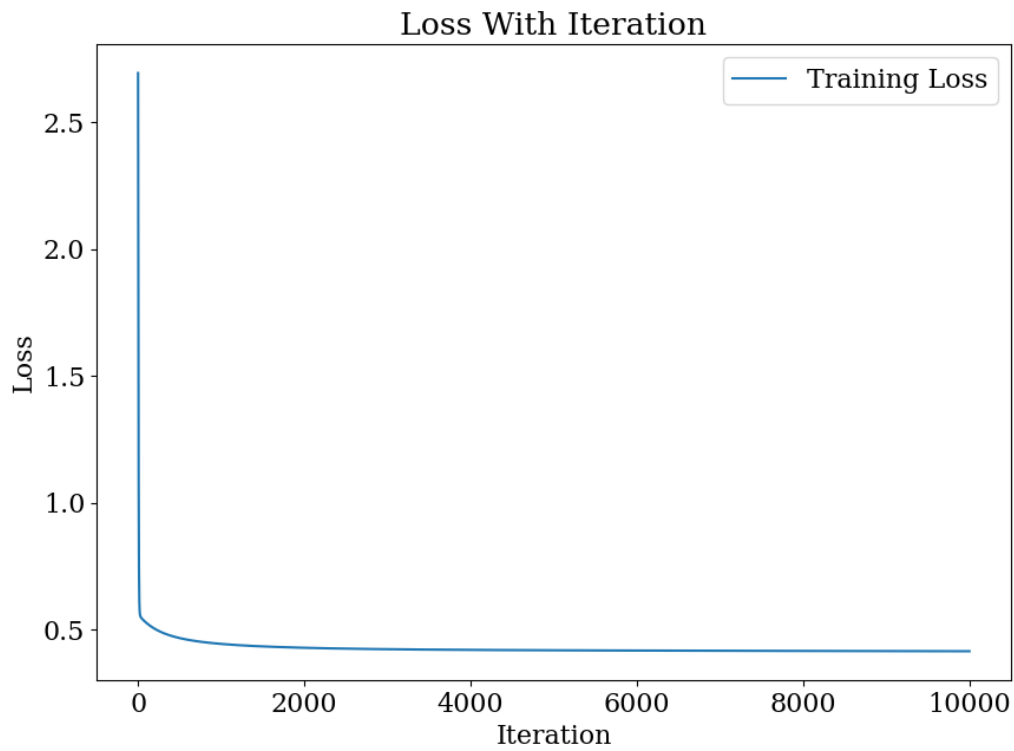

Problem 3

Here, the loss, which is nothing but accuracy, is plotted against the number of iterations. We observe that:

1. The loss is not decreasing too much after epoch 10 or so.
2. Testing loss is always much lower that the training loss, showing that the model is not overfitting.

This figure shows the loss with number of epochs for the `LogisticGradientDescent` which I implemented while implementing gradient descent algorithms. This was not required for this problem.