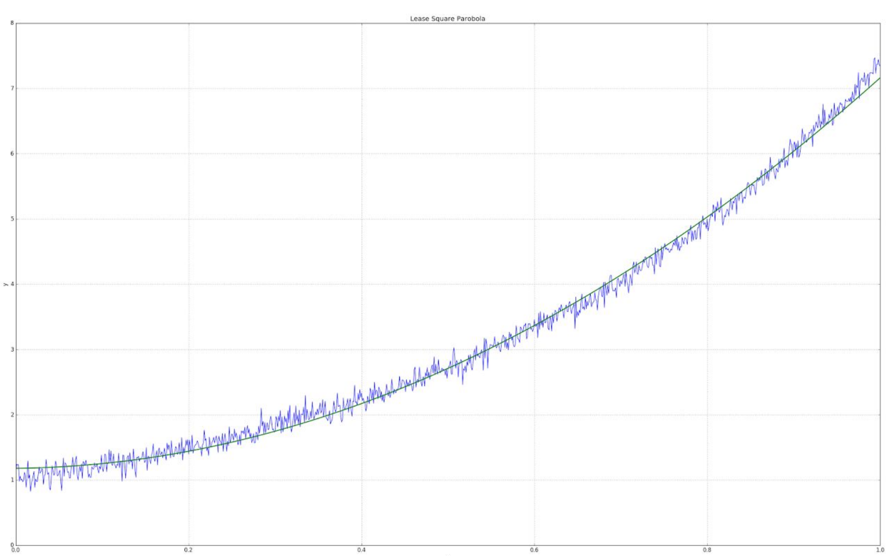
Assignment 4
By: Ravi Patel
A11850926

NOTE: All code for this assignment will be linked separately from this write-up

1. Least Square Parabola
   a. Rewrite the loss function in matrix form to arrive at the result:
   $L(\Theta) = (Y - X\Theta)^2$
   Where
   $$Y = [y_1, y_2, \cdots, y_n]^T, \quad X = \begin{bmatrix} x_1^2 & x_1 & 1 \\ x_2^2 & x_2 & 1 \\ \vdots & \vdots & \vdots \\ x_n^2 & x_n & 1 \end{bmatrix}, \quad \Theta = [\theta_2, \theta_1, \theta_1]^T$$
   Then
   $\Theta^* = argmin_\theta(L(\Theta)) = argmin_\theta((Y - X\Theta)^2)$
   By differentiation, the result follows:
   $\frac{dL}{d\Theta} = -2X^T(Y - X\Theta) \quad \Rightarrow \frac{dL}{d\Theta} = 0$ implies $\Theta^* = (X^TX)^{-1}X^TY$.

   b.



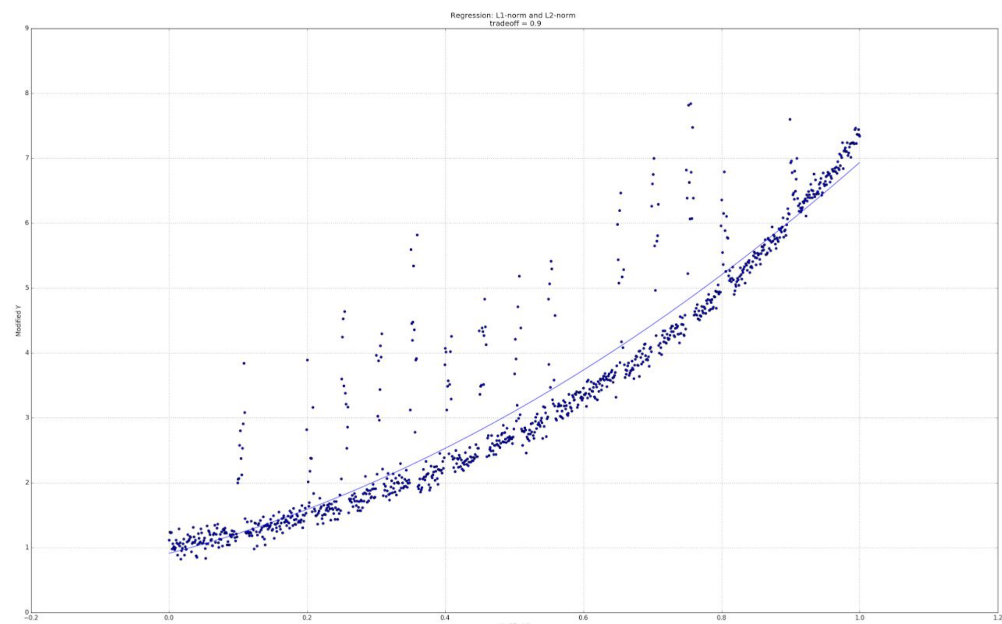
c. Comparing the formulation between linear and quadratic function, the quadratic function captures the trend more accurately than linear function. As a result, the quadratic function is more suitable for this data set.

2. Regression

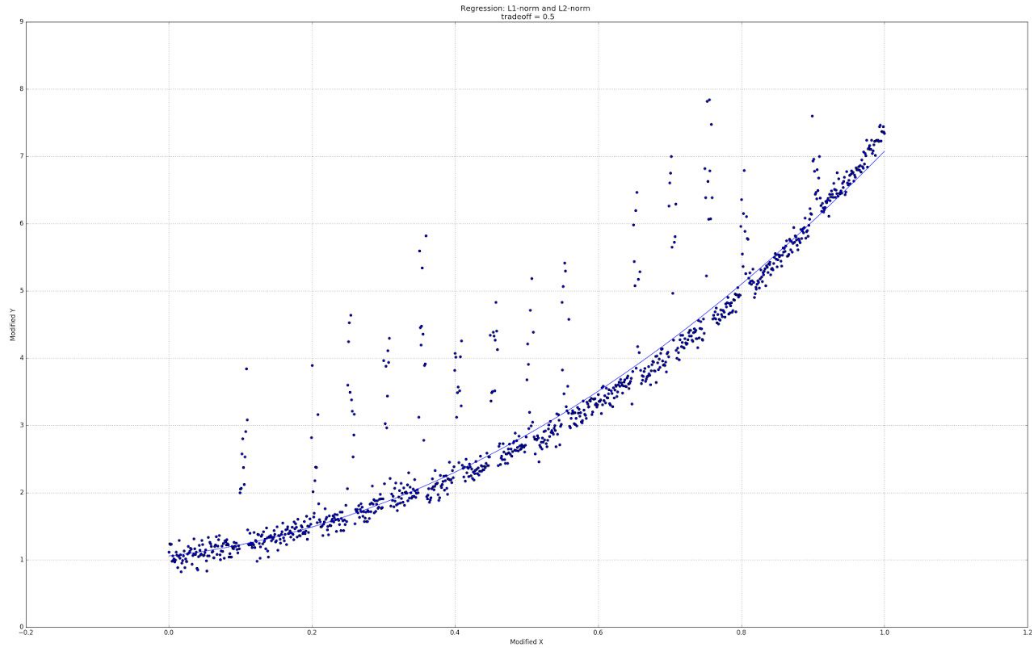
a. Define X, Y, Θ as above, then the gradient of the loss function is a combination of L1-norm and L2-norm loss function:
$\frac{dL}{d\Theta} = -\lambda X^T(Y - X\Theta) - (1-\lambda)X^T \cdot sign(Y - X\Theta)$, and the formulation of the update parameter via gradient descent would be:

$$\Theta_{n+1} = \Theta_n - \mu \frac{dL}{d\Theta}\bigg|_{\Theta=\Theta_n}$$, where μ is the learning rate.

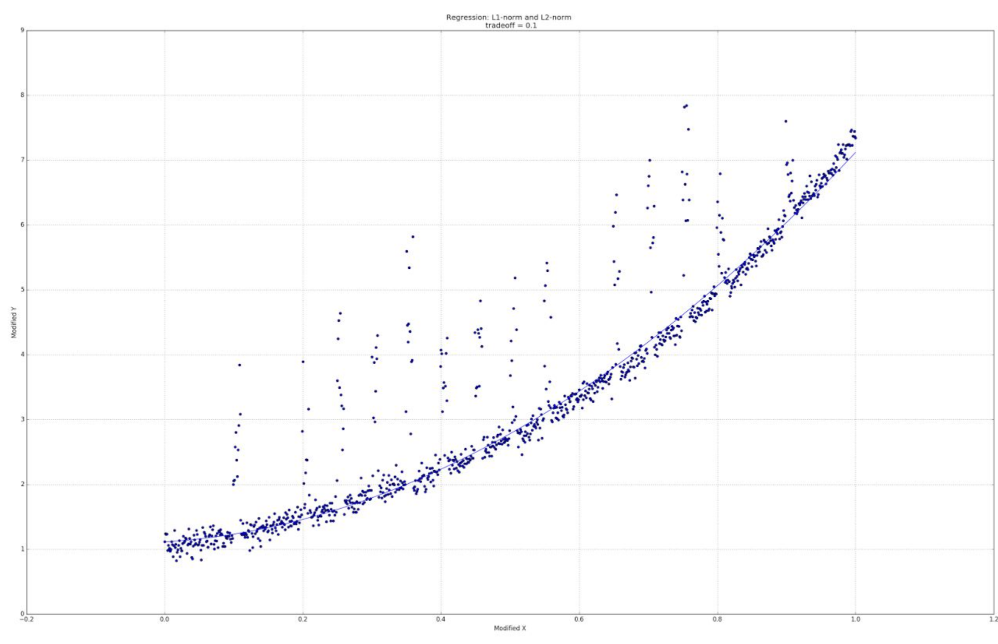
b.



$\Theta = [\theta_2, \theta_1, \theta_0]^T = [3.291, 2.731, 0.912]^T$ when tradeoff = 0.9



$\Theta = [\theta_2, \theta_1, \theta_0]^T = [4.835, 1.181, 1.058]^T$ when tradeoff = 0.5



$\Theta = [\theta_2, \theta_1, \theta_0]^T = [5.337, 0.669, 1.111]^T$ when tradeoff = 0.

3. Logistic Regression

$$\Theta^* = [\theta_2, \theta_1, \theta_0] = [0.03428048,\ -0.26204465,\ 0.43163758]$$

4. Linear Discriminative Analysis