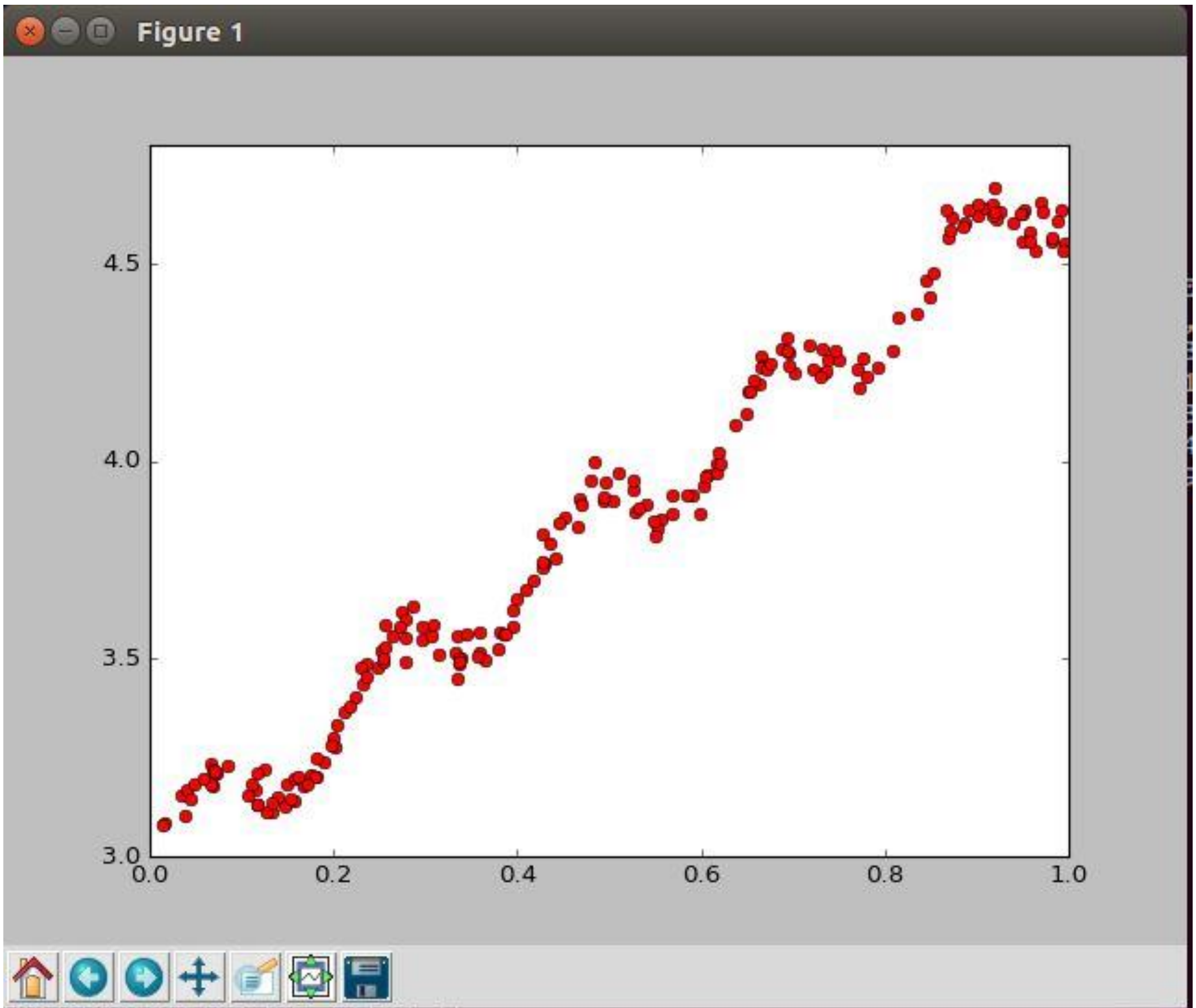
Guthrie Alexander

CS 6316-Machine Learning

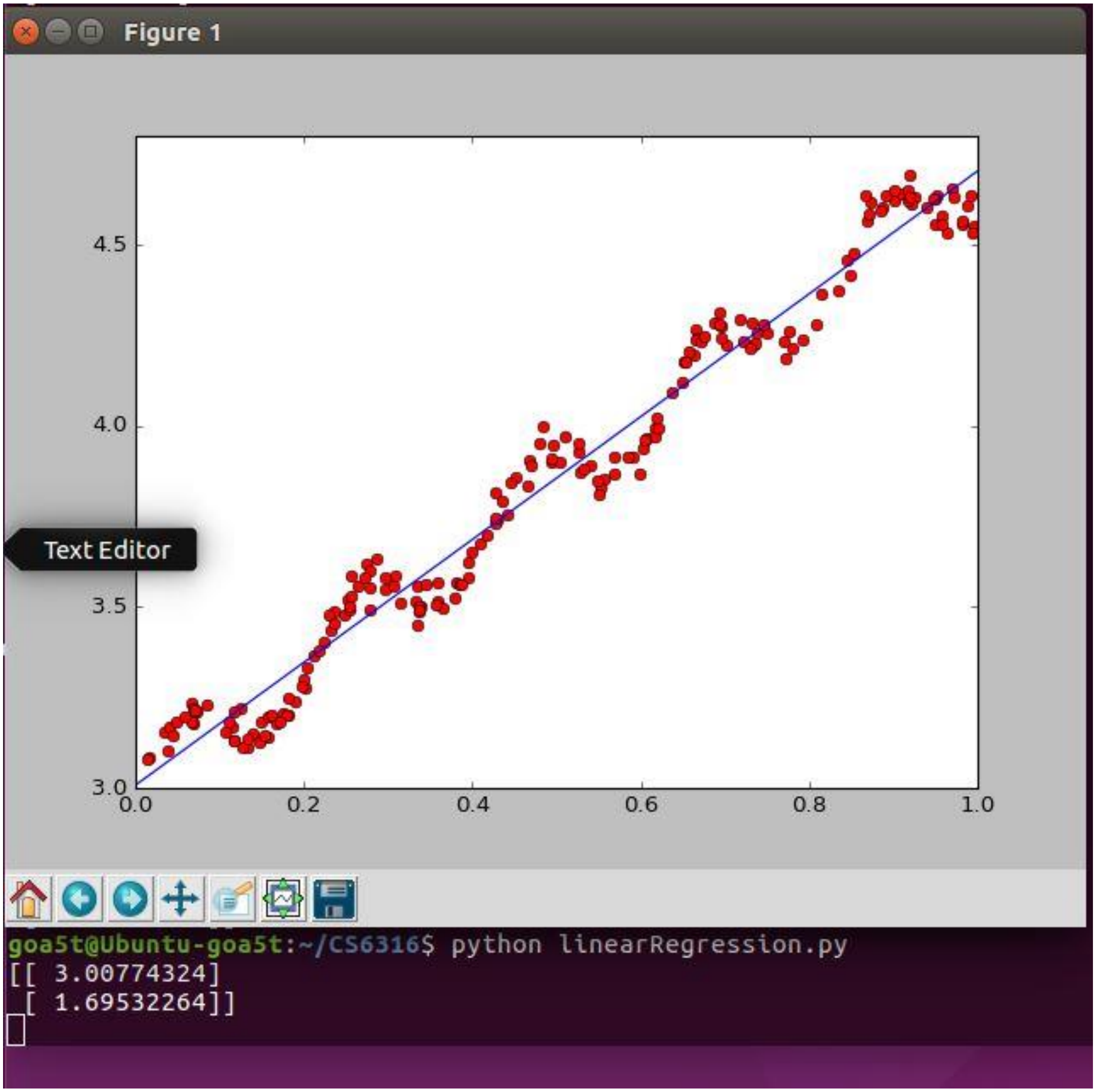
Homework 2

Question 1:

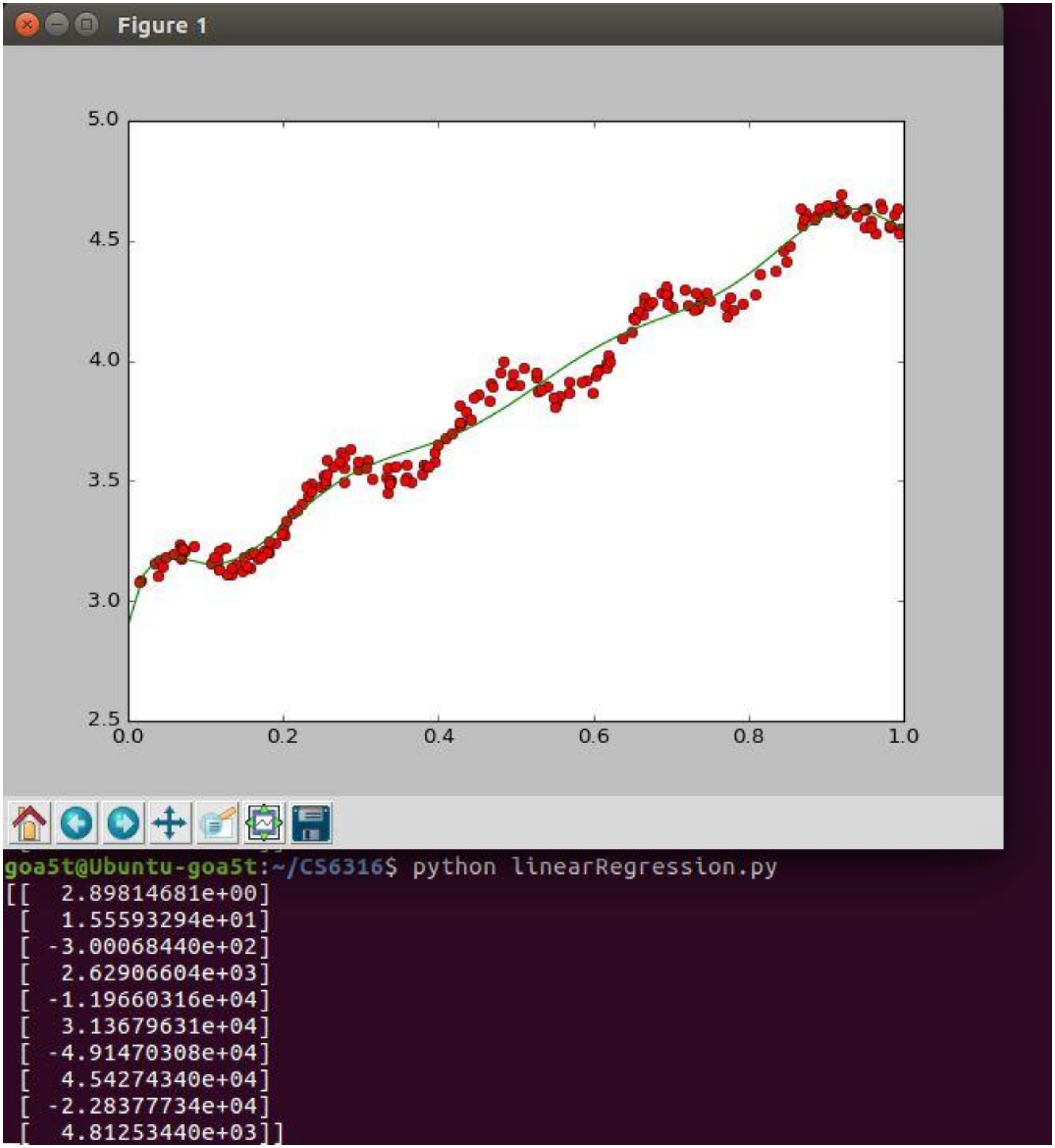
Plot of data:



Plot of data and theta:



Plot of data and polynomial (9th order) regression:



Steps taken to derive this solution: I used the normal equation to derive the linear regression with a beta with 2 values. I then opted to use the 9th order polynomial to plot a best fit line with the above data. This worked better than a simple 2nd order polynomial, which barely had any noticeable curvature in the line at all.

Normal Equation: theta = (X^t * X) ^-1 * X^t * y

Question 2:

1.1 (Scan derivation)

1.1 $RSS = (y - XB)^T (y - XB) + \lambda B^T B$

$\quad = y^T y - y^T X B - B^T X^T y + B^T (X^T X) B + \lambda B^T B$

$y^T X B = (y^T X B)^T = B^T X^T y$

$\quad = y^T y - 2 B^T X^T y + B^T (X^T X) B + \lambda B^T B$

$$\frac{\partial}{\partial \beta} RSS = 0 - \frac{\partial}{\partial B} 2 B^T X^T y + \frac{\partial}{\partial B} B^T (X^T X) B + \frac{\partial}{\partial \beta} \lambda B^T B$$

$$0 = -2 X^T y + 2 (X^T X) \beta + 2 \lambda B$$

$$X^T y = B (X^T X + \lambda I), \quad \boxed{B = (X^T X + \lambda I)^{-1} X^T y}$$

1.2 $X = \begin{bmatrix} 1 & 2 \\ 3 & 6 \\ 5 & 10 \end{bmatrix} \qquad X^T = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 6 & 10 \end{bmatrix}$

$$\Theta = (X^T X)^{-1} X^T Y$$

$$= \begin{bmatrix} 1 & 3 & 5 \\ 2 & 6 & 10 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 6 \\ 5 & 10 \end{bmatrix} = \begin{bmatrix} 35 & 70 \\ 70 & 140 \end{bmatrix}^{-1}$$

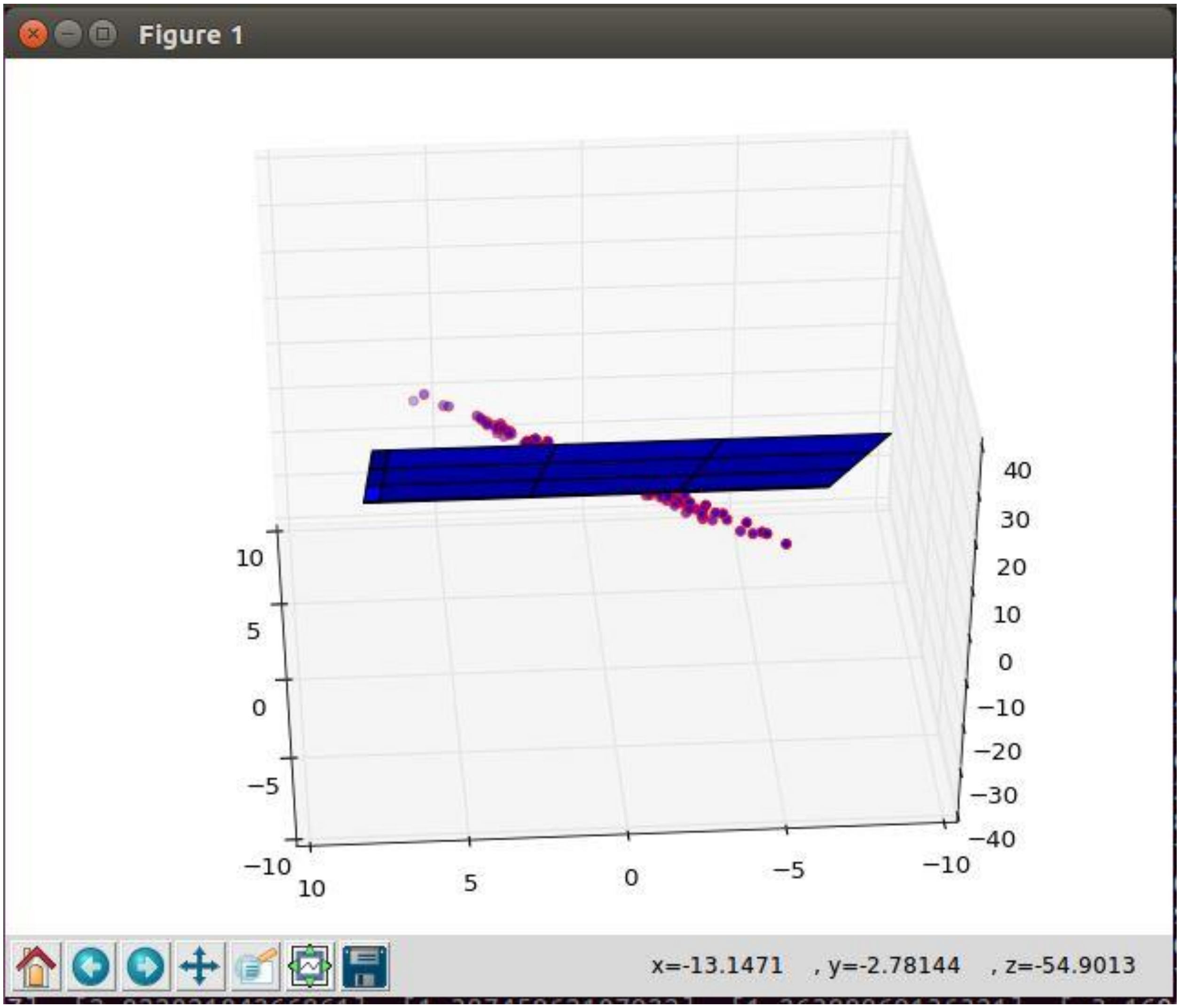
$\det(\,) = 0$

→ determinant is 0, inverse doesn't exist
The matrix isn't full rank, the columns are not linearly independent

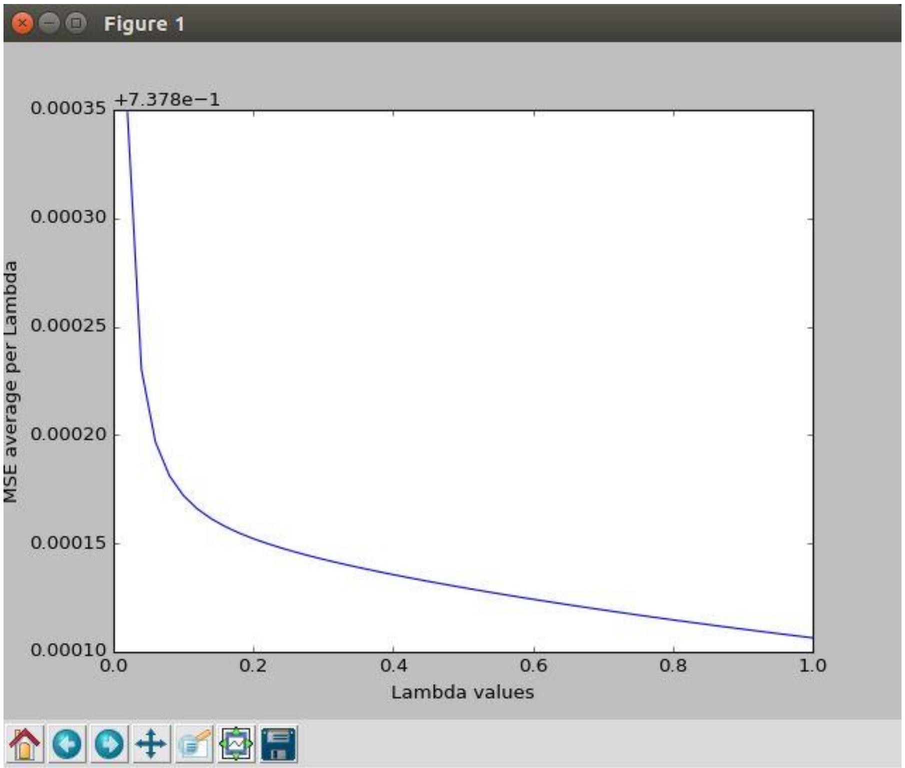
1.2 (see above)

1.3 If coefficient B should be sparse, the best regularized method to solve would be the Lasso – it tends to generate sparse coefficients.

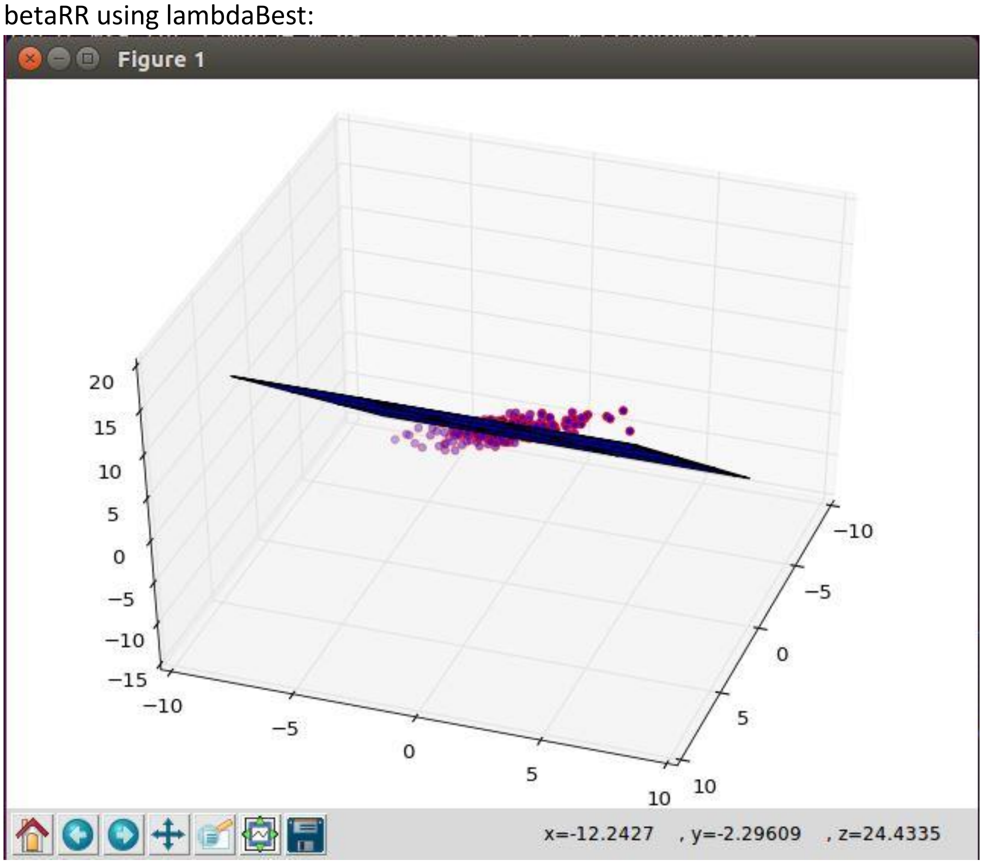
1.4 Lamba=0:



Lamba versus MSE:



betaRR using lambdaBest:



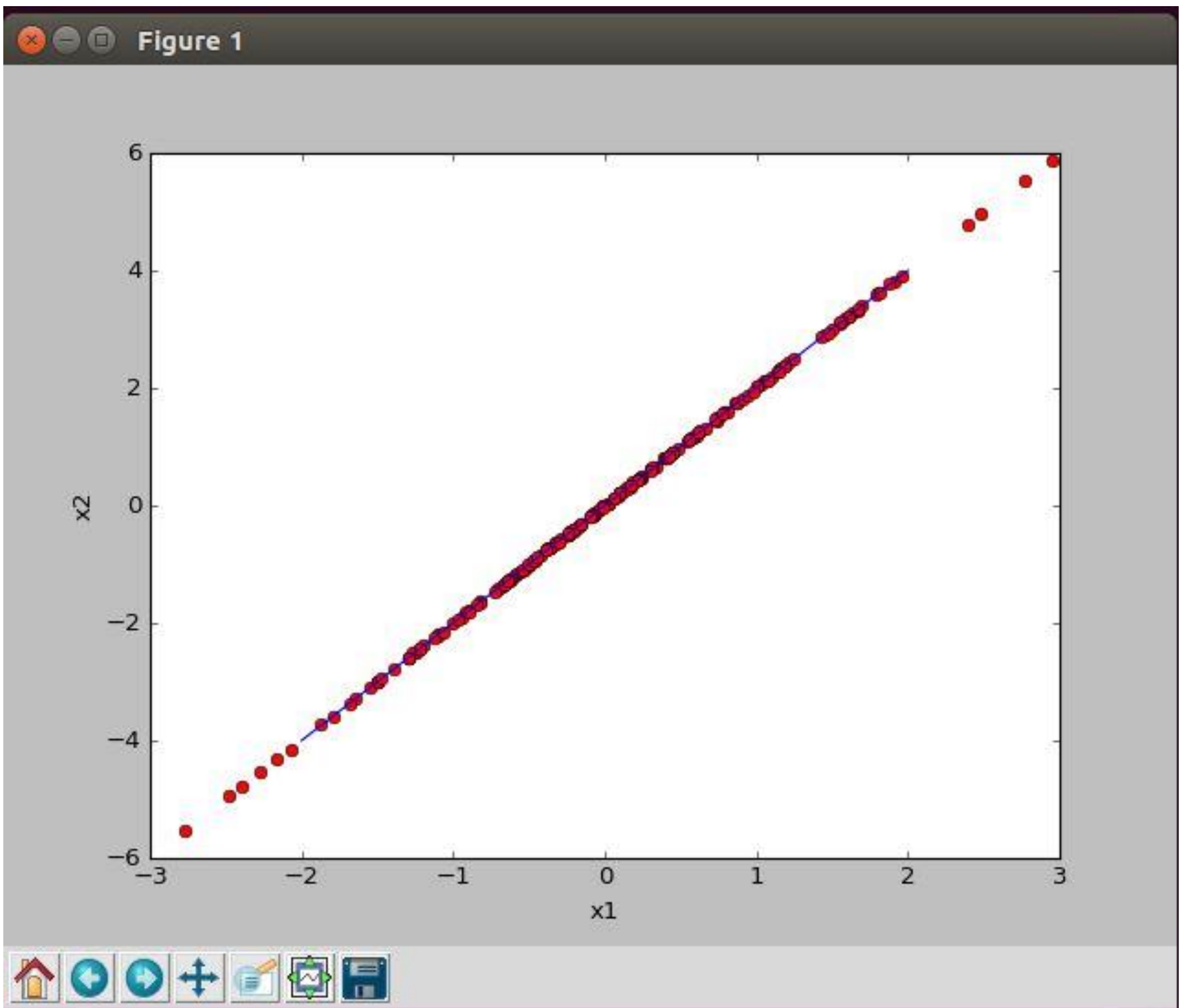
betaRR =

```
betaRR =
[[ 2.97310976]
 [ 0.53192201]
 [ 1.19379912]]
```

1.5 Ridge Regression

1.5.1 This is because the coefficient matrix isn't full rank; not all of the columns are linearly independent.

Plot of x1 versus x2 data points, and linear regression:



As you can see, the data points for x1 and x2 can be described with a single linear line. This means that inversion of the full matrix leads to instability in the system. Adding regularization fixes this, as the inverse function from the normal equation does not lead to instability.