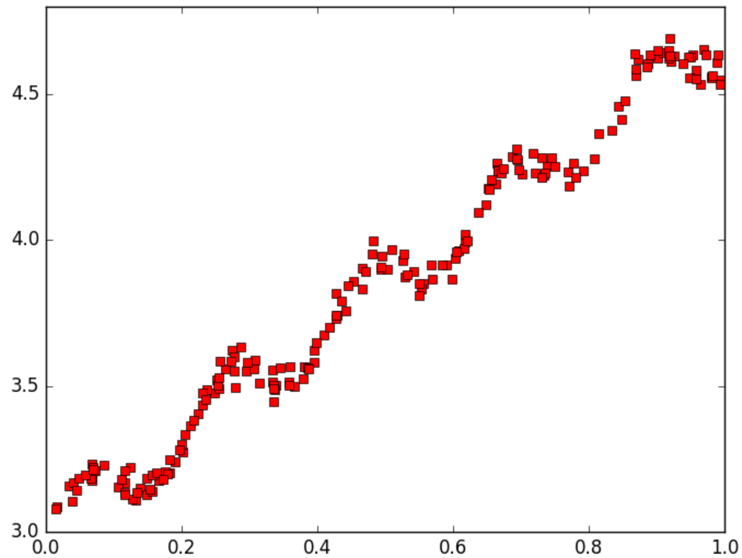
Eamon Collins
Ec3bd

Note: I could not get 3d plotting to work, it was being quite thorny. I realize I should have tried it sooner and worked out the kinks, but as it is please look at my data/methods and see if they would have graphed correctly. Thank you.

Question 1

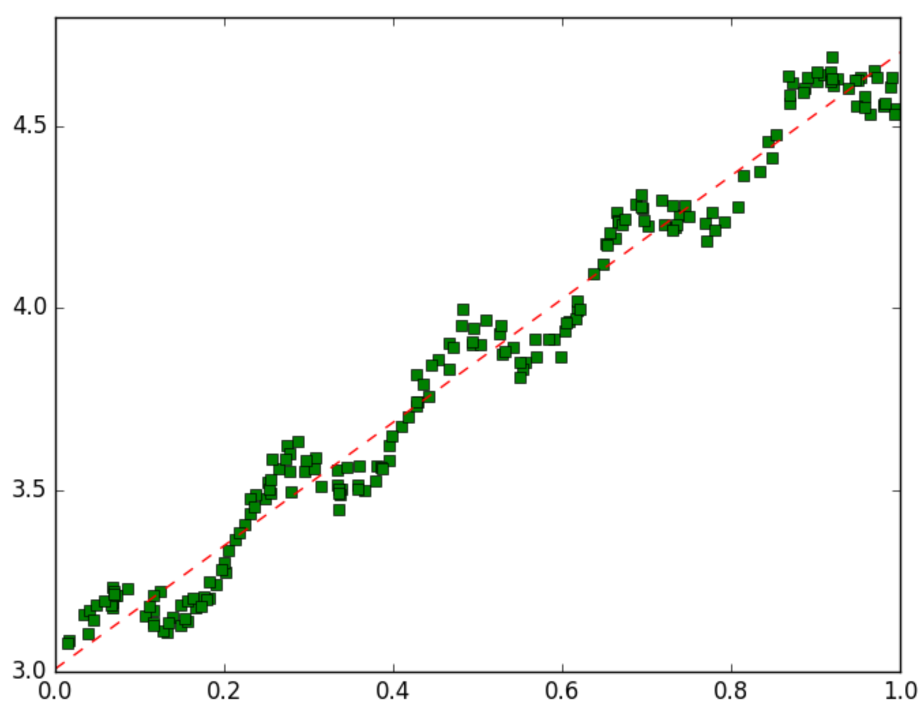
Initial data:



Standard Linear Regression:

Theta = [[3.00774324]

[1.69532264]]

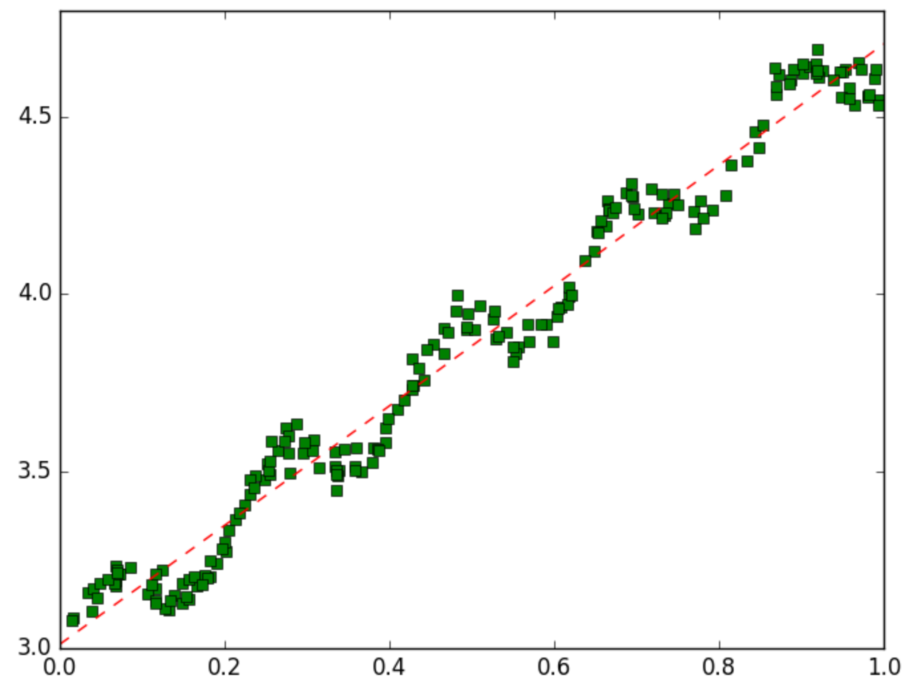


Second Order Linear Regression:

Theta = [[3.01139245]

[1.67414864]

[0.02065965]]



Question 2

Q2

1.1 $\hat{\beta}^{ridge} = \arg\min (y - X\beta)^T(y - X\beta) + \lambda \beta^T \beta$ take gradient, to find argmin

$$\frac{\partial \hat{\beta}^{ridge}}{\partial \beta} = 2X^T(y - \beta^T X) + 2\lambda\beta$$
$$= 2X^T y - 2X^T X\beta + 2\lambda\beta$$
$$X^T y - X^T X\beta + \lambda\beta$$
$$X^T y = (X^T X - \lambda I)\beta$$
$$\beta = (X^T X - \lambda I)^{-1} X^T y$$

Our solution

1.2 $X^T X = \begin{pmatrix} 1 & 3 & 5 \\ 2 & 6 & 10 \end{pmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 6 \\ 5 & 10 \end{bmatrix} = \begin{bmatrix} 35 & 70 \\ 70 & 140 \end{bmatrix}$ $35 \cdot 140 - 70 \cdot 70 = 0$

$X^T X$ is singular, no inverse exists and so linear regression by normal eqn. will not work

1.3 LASSO

LASSO produces a sparse coefficient vector in that many features will be weighted 0

Ridge Regression Lambda = 0

Beta = [[2.97139801]

[-11.00332214]

[6.96229098]]

Ridge Regression after iterative lambda testing:

Beta = [[2.97264929]

[-1.54499492]

[2.23351352]]

Best lambda = .02

Mean squared error = . 739515616958

1.5.

The Beta values for ridge regression are much more accurate than those calculated by linear regression.

Upon implementing a standard regression between x1 and x2, we see that x2 is almost always about 2 times x1. Ridge regression is better when some of the features are highly correlated because it adds some bias to the variables.