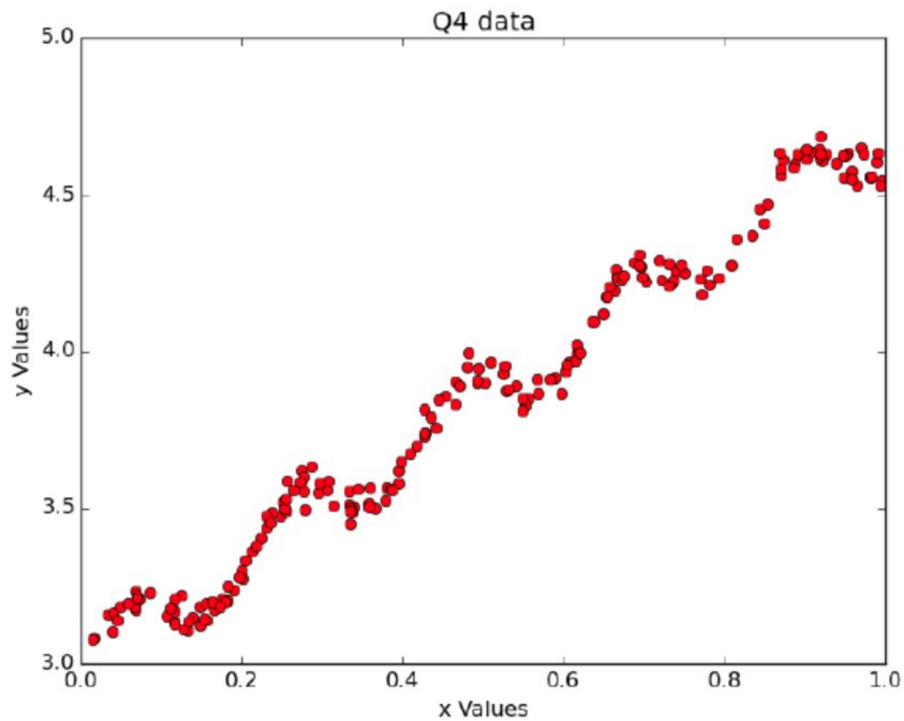
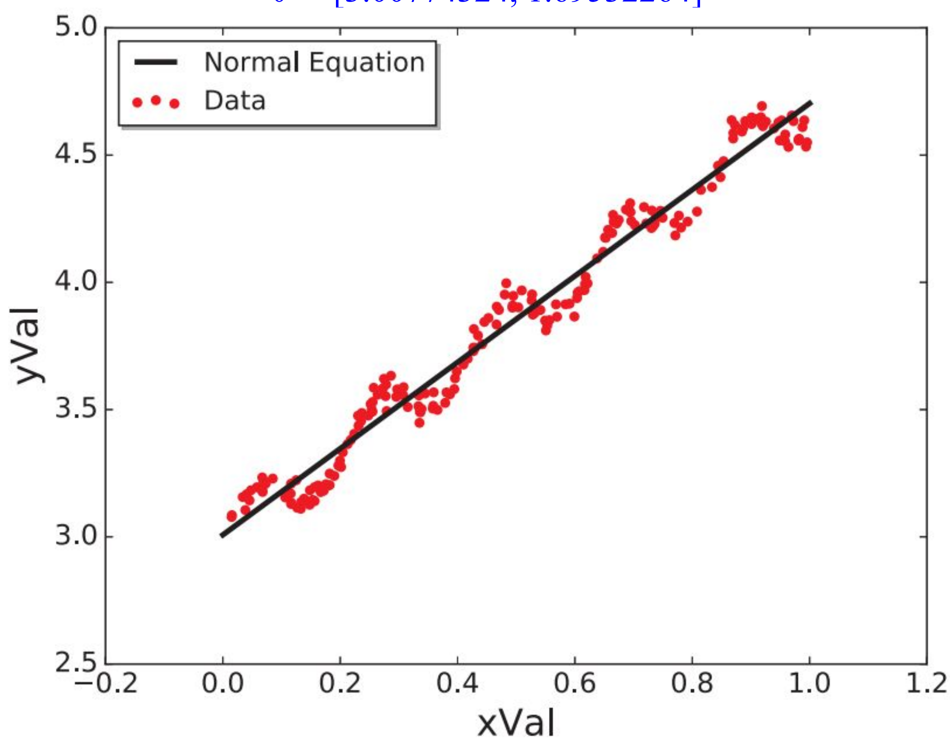
1.1



1.2

Linear regression using normal equations

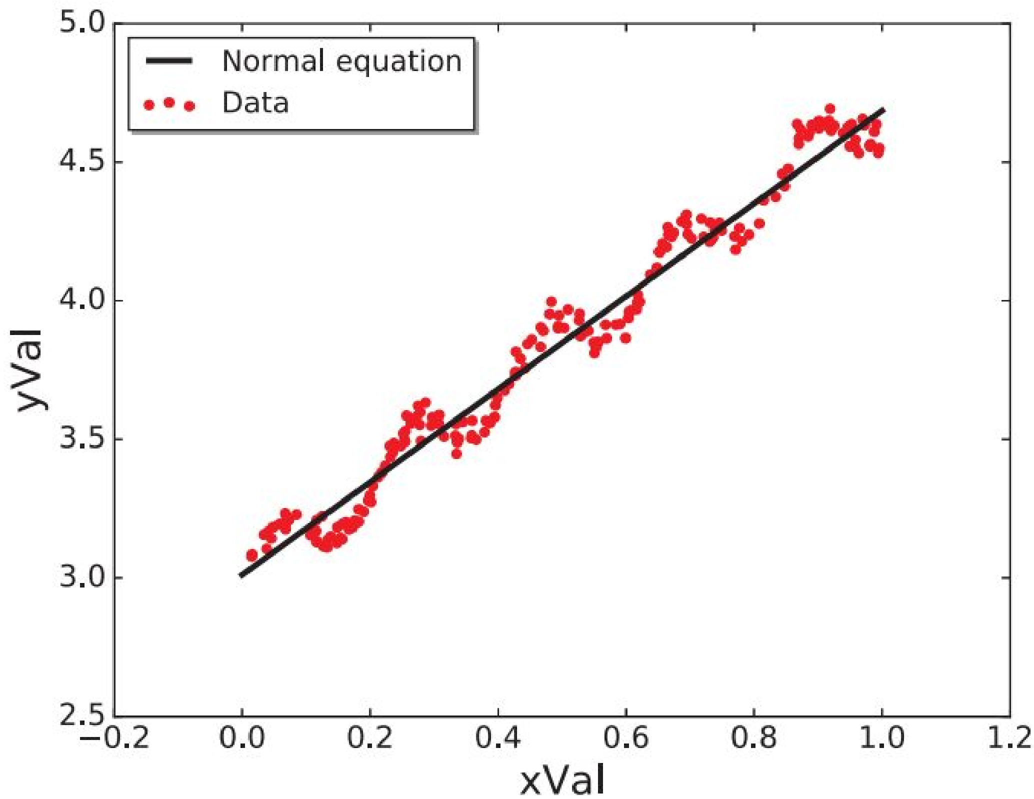
$\theta = [3.00774324, 1.69532264]^T$



1.3

2nd order polynomial regression using normal equations

$\theta = [3.01139245, 1.67414864, 0.02065965]^T$



2.1

$$J(\beta) = (y - X\beta)^T(y - X\beta) + \lambda\beta^T\beta$$
$$\nabla_\beta J(\beta) = -2X^Ty + 2X^TX\beta + 2\lambda I\beta$$
$$\text{Let } \nabla_\beta J(\beta)|_{\beta=\beta^*} = 0$$
$$\beta^* = (X^TX + \lambda I)^{-1}X^Ty$$

2.2

$$det(X^T X) = 35 \times 140 - 70 \times 70 = 0 \tag{2}$$

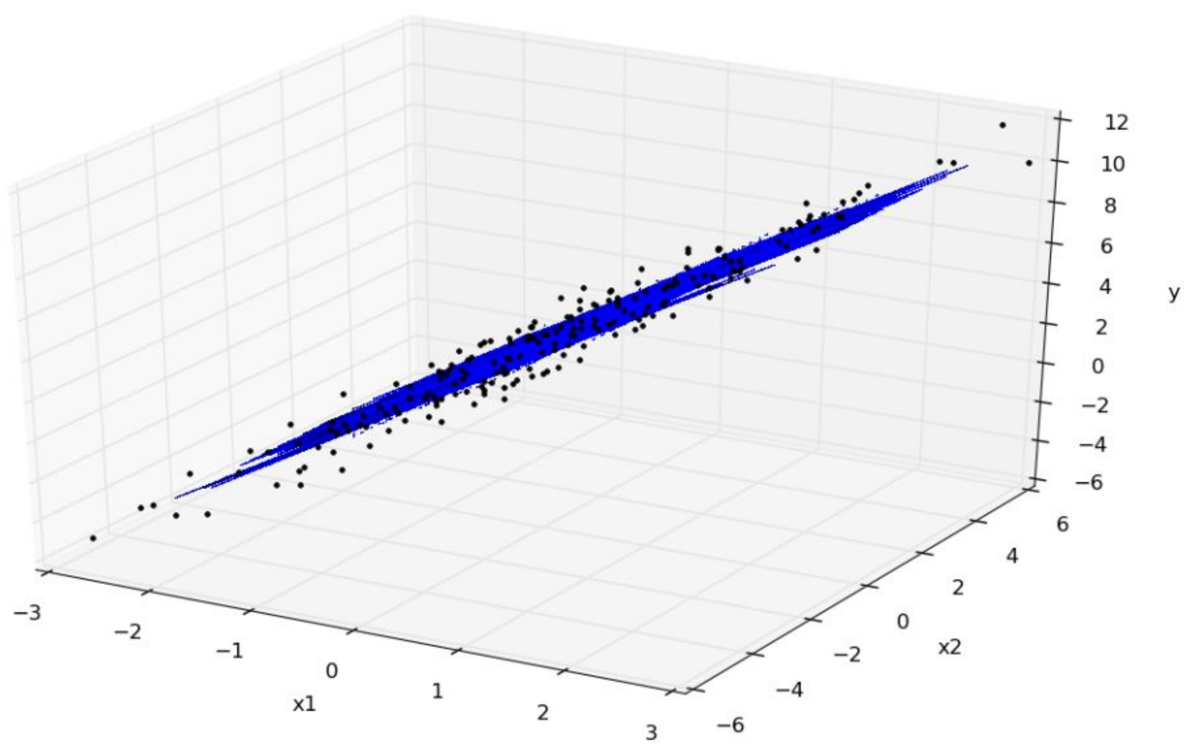
This means that $X^T X$ is not invertable. Therefore, you can not compute β^* through $\beta^* = (X^T X + \lambda I)^{-1} X^T y$.

2.3

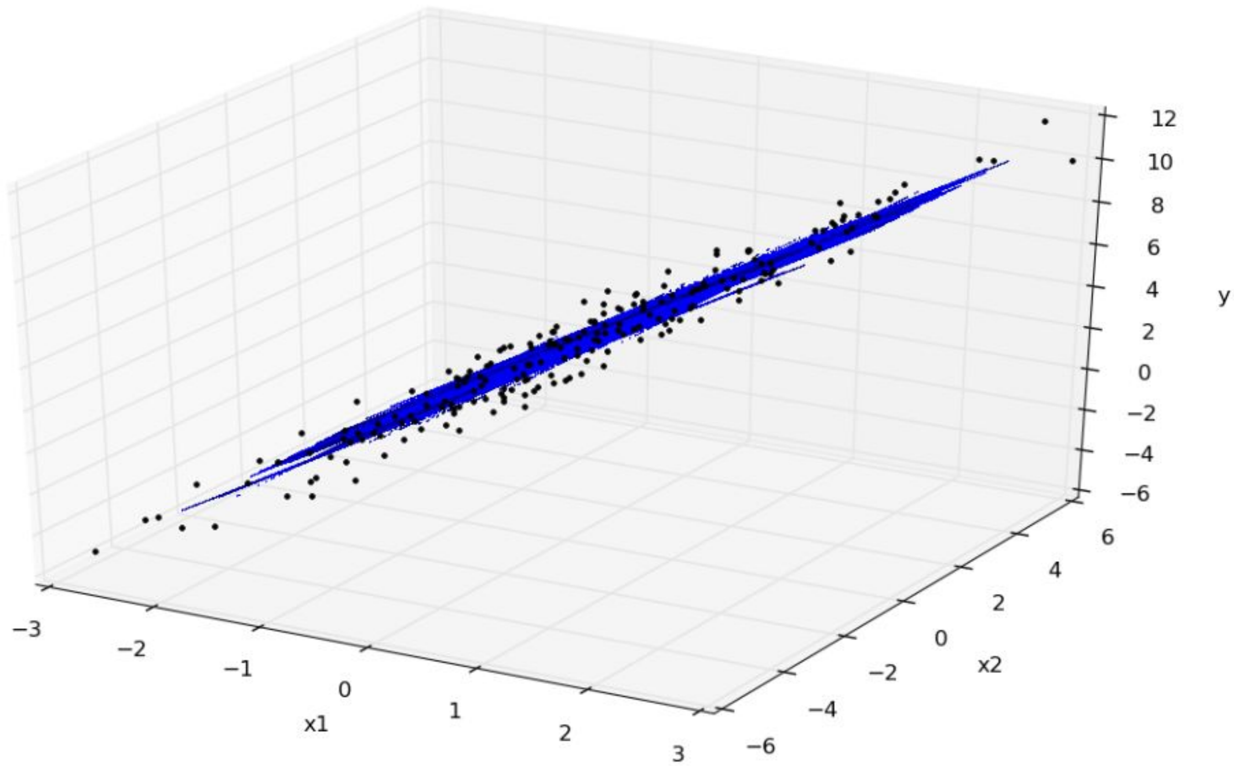
Lasso

2.4

With $\lambda = 0$, $\beta_{LR} = [2.97139801, -11.00332214, 6.96229098]^T$

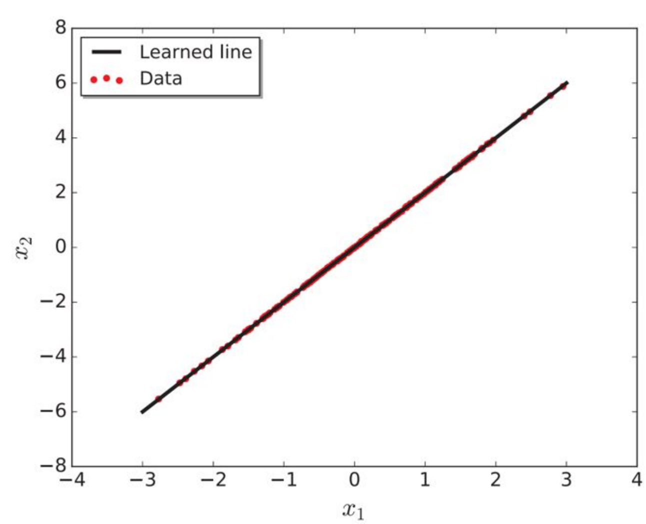


With $\lambda = 1$, $\beta_{RR} =$ $[3.17523191, 0.53352905, 1.1864448]^T$



2.5

If the true coefficient is $\beta = (3, 1, 1)T$, then the ridge regression estimate at $[3.17523191, 0.53352905, 1.1864448]^T$ performs better. Since the x_1 vs x_2 looks like the following,



we can see that the relationship between x_1 and x_2 appears linear. Thus it is likely that X is not full rank, and therefore X^TX is not invertable, making linear regression perform worse.