CSCE 5063-001: Assignment 1

Due 11:59pm Friday, September 17, 2021

Name: Rashmitha Thoranala

I used the data from the data.txt file given in the assignment which contains 17 columns that represent the features and 60 rows that represent examples. Our main goal is to implement linear regression to predict the death rate from several features, including annual precipitation, temperature, population, income, pollution, etc. that was given in the data.txt file.

1.1 Feature Normalization:

Some of the features in this data are 1000 times others thereby, features differ by order of magnitude so, in order to achieve gradient descent convergence more quickly we need to perform feature normalization at first.

1.2 Gradient descent with quadratic regularization:

The loss function for linear regression with quadratic regularization is:

$$J(\theta) = \frac{1}{2m}\sum_{i=1}^{m} (y^{(i)} - h_\theta(x^{(i)}))^2 + \frac{\lambda}{2m}\sum_{j=1}^{n} \theta_j^2$$

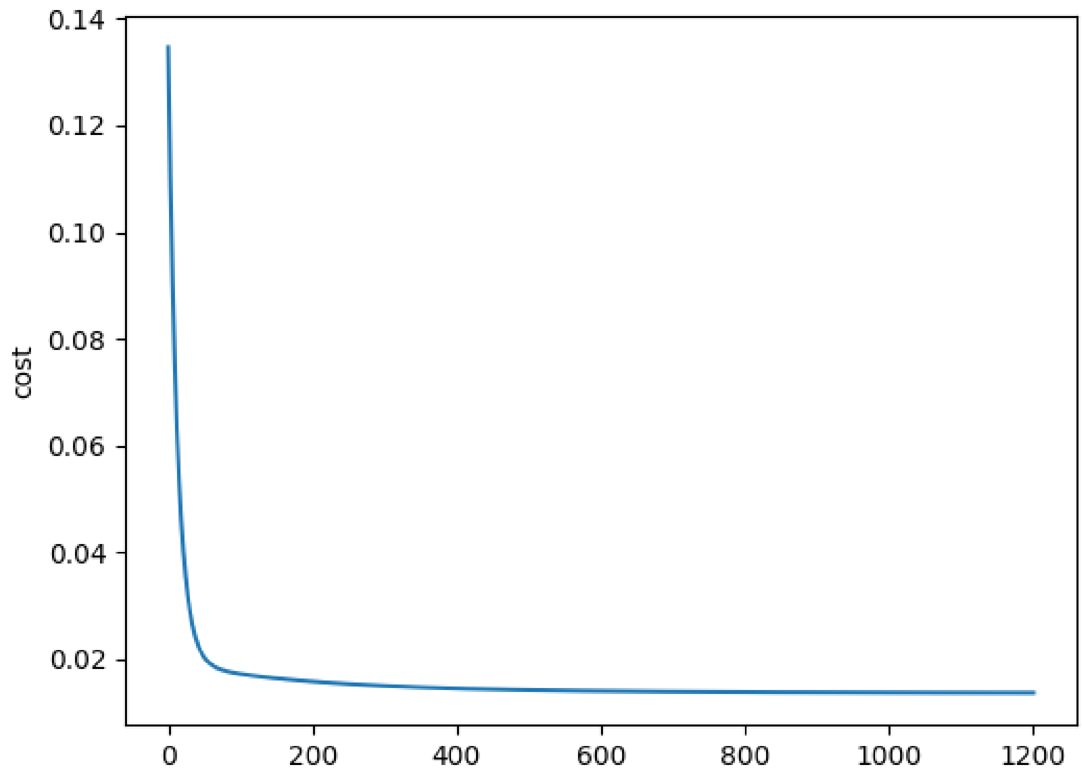
To minimize this function, we need to find gradient with respect to each parameter.

Gradient = $\frac{\partial J(\theta)}{\partial \theta_j} = \frac{1}{m}\sum_{i=1}^{m}(h_\theta\left(x^{(i)}\right) - y^{(i)})x_j^{(i)} + \frac{\lambda}{m} * \theta_j$

In order to this algorithm, we need to first load the data and then split into 80% / 20% training/test data so we can use first 80% of data as training data and remaining 20% of data as testing data. Now, learn the linear regression model on training data.

1. Plot the value of loss function J_k(θ) vs. the number of iterations (k) for Section 1.2, report the squared loss on the test data for Section 1.2

Solution: The below plot shows the value of loss function J_k(θ) vs. the number of iterations (k) using gradient descent with quadratic regularization. The x-axis represents the number of iterations and y-axis represents the cost.



The value of squared loss function without quadratic regularization on test data is 0.09146.

1.3 Gradient descent with lasso regularization

The loss function for linear regression with lasso regularization is given by

$$J(\theta) = \frac{1}{2m}\sum_{i=1}^{m} (y^{(i)} - h_\theta(x^{(i)}))^2 + \frac{\lambda}{2m}\sum_{j=1}^{n} |\theta_j|$$

2. Equation for the gradient of Eq. (2).

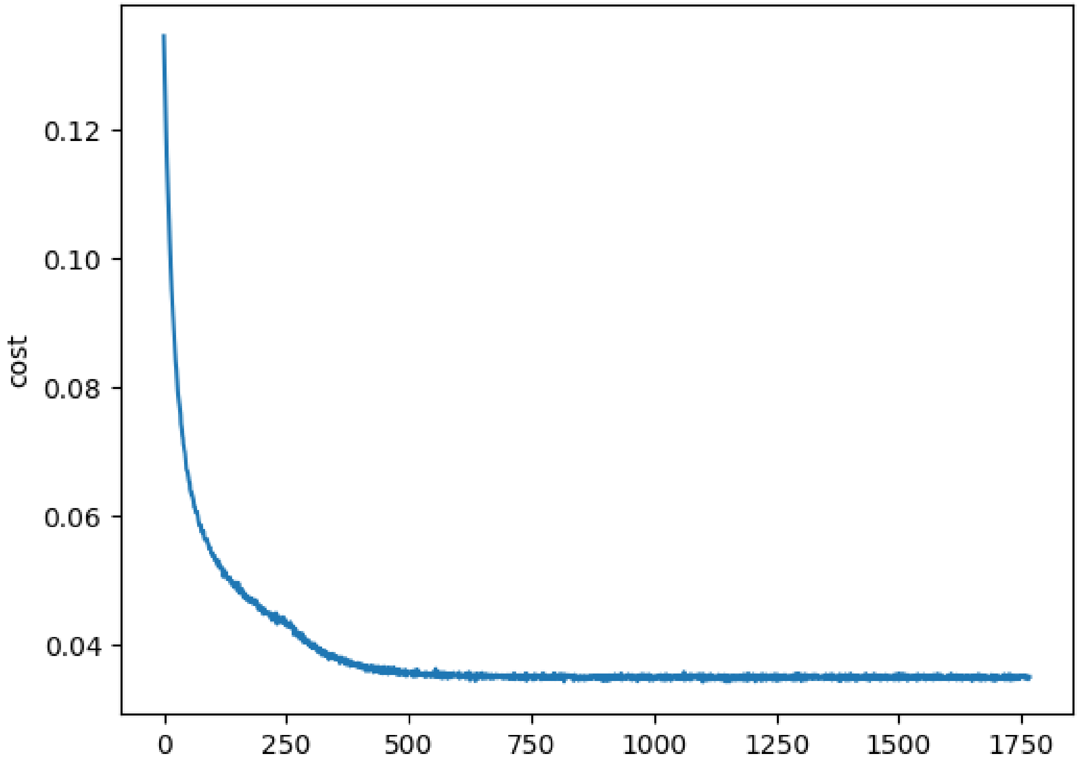
To minimize the loss function we need to find gradient by obtaining partial derivative of the above function with respect to θ_j.

Gradient = $\frac{\partial J(\theta)}{\partial \theta_j} = \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_j^{(i)} - \frac{n^*\lambda}{2m}$ if $\theta_j < 0$

Gradient= $\frac{\partial J(\theta)}{\partial \theta_j} = \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_j^{(i)} + \frac{n^*\lambda}{2m}$ if $\theta_j >= 0$

3. Plot the value of loss function $J_k(\theta)$ vs. the number of iterations (k) for Section 1.3, report the squared loss on the test data for Section 1.3.

The below plot shows the value of loss function $J_k(\theta)$ vs. the number of iterations (k) using gradient descent with lasso regularization. The x-axis represents the number of iterations and y-axis represents the cost.



The value of squared loss function without lasso regularization on test data is 0.09146.

4. Numbers of zero parameters of the models obtained in Sections 1.2 and 1.3.

Answer: The number of zero parameters obtained using quadratic regularization is 3 and the number of zero parameters obtained using lasso regularization is 15. Hence, compared to quadratic regularization, the lasso regularization produces a greater number of zero parameters.