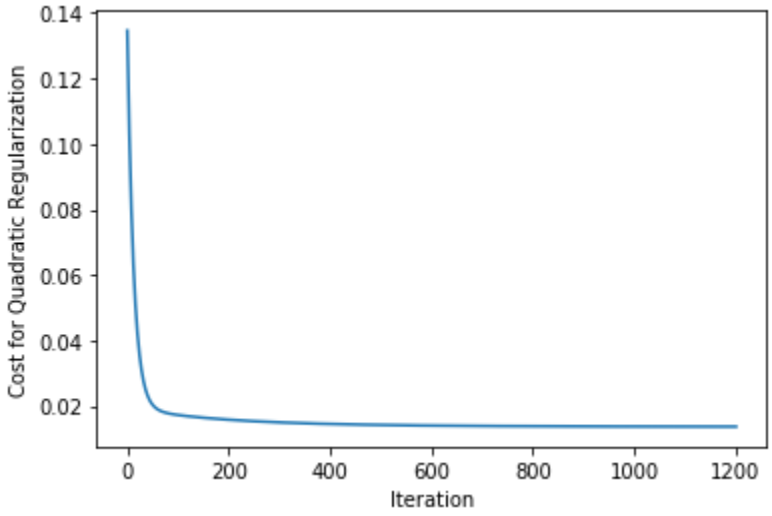
1. Plot the value of loss function $J_k(\theta)$ vs. the number of iterations (k) for Section 1.2, report the squared loss on the test data for Section 1.2.

Task: Load the dataset in data.txt, split it into 80% / 20% training/test (the dataset is already shuffled so you can simply use the first 80% examples for training and the remaining 20% for test.), and learn the linear regression model using the training data. Plot the value of loss function $J_k(\theta)$ vs. the number of iterations (k). After the training completes, compute the squared loss (w/o regularization function) on the test data.

Solution: Here is the plot (cost function $J_k(\theta)$ vs. the number of iterations (k)) for quadratic regularization:



The squared loss (w/o regularization function) is computed using following equation:

$$J(\boldsymbol{\theta}) = \frac{1}{m}\sum_{i=1}^{m}\left(y^{(i)} - h_{\boldsymbol{\theta}}(\boldsymbol{x}^{(i)})\right)^2$$

$$= \frac{1}{m}\sum_{i=1}^{m}\left(y^{(i)} - \boldsymbol{\theta}^T\boldsymbol{x}^{(i)}\right)^2$$

The value for the testing data set using quadratic regularization parameters is 0.02474187700900062.

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

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

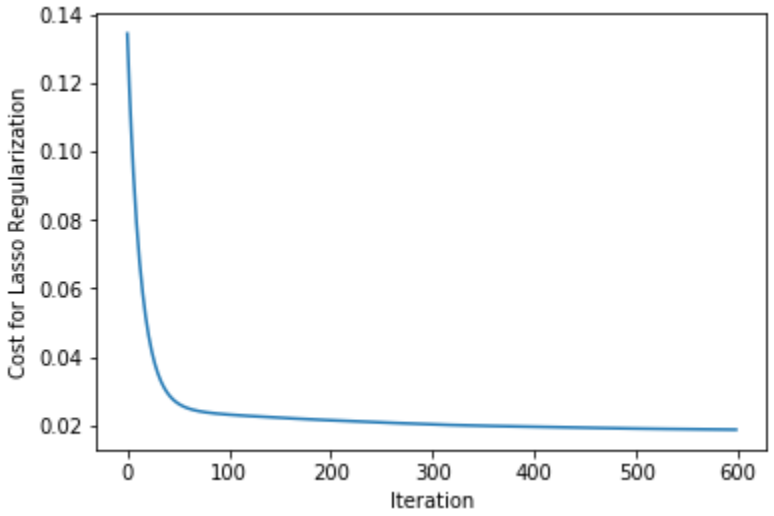
Solution:

$$\frac{\partial J(\theta)}{\partial \theta_j} = \begin{cases} \frac{1}{m}\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right)x_j^{(i)} + \frac{\lambda}{2m} & ;\ \theta_j \geq 0 \\ \frac{1}{m}\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right)x_j^{(i)} - \frac{\lambda}{2m} & ;\ \theta_j < 0 \end{cases}$$

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.

Task: Load the dataset in data.txt, split it into 80% / 20% training/test, and learn the linear regression model using the training data. Plot the value of loss function $J_k(\theta)$ vs. the number of iterations (k). After the training completes, compute the squared loss (w/o regularization function) on the test data.

Solution: Here is the plot (cost function $J_k(\theta)$ vs. the number of iterations (k)) for lasso regularization:



The squared loss value for the testing data set using lasso regularization parameters is 0.025920793835615637.

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

	Quadratic Regularization	**Lasso Regularization**
Number of zero parameters	3	6

5. Source Code file (.py) is attached.