Assignment 1 EECS545

Jiazhao Li

2018/1/28

1 Problem1:Gradient Descent

1.1 Stochastic Gradient Descent (SGD)

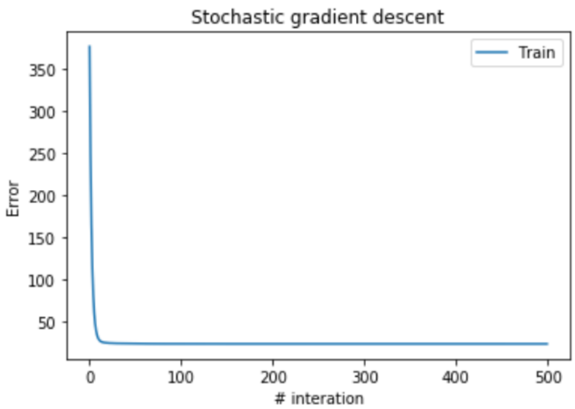

Figure 1: Train Error Vs Iteration.

Learning rate = 5e-4
epoch = 500
Weighted Vector: 22.946. -0.947. 1.198. 0.203. 0.681. -2.119. 2.763. 0.298
-3.121,2.952,-2.45, -2.016,0.91,-4.077
Bias Term : 22.946
Train Error : 23.199
Test Error: 11.0068

1.2 Batch Gradient Descent (BGD)

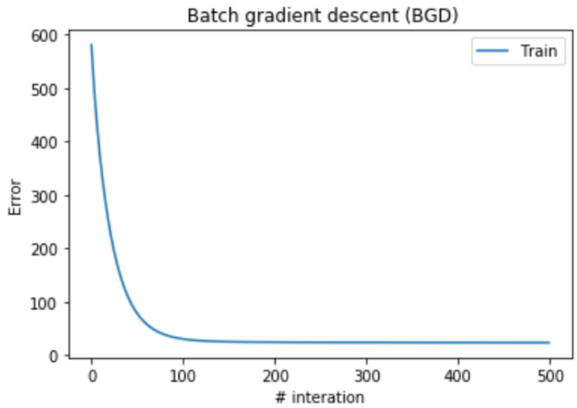

Figure 2: Train Error Vs Iteration.

Learning rate = 5e-5
eopch = 500
Weighted Vector: 22.940 -0.800 0.980 -0.158 0.735 -1.710 2.948 0.155
-2.922 1.6112 -1.167 -1.873 0.9180 -3.9381
Bias term: 22.940
Train Error: 23.454
Test Error: 9.7016

1.3 Closed form solution(pinv)

They are alomst the same as previous result.
Bias term: 22.9410
Train Error: 23.1917
Test Error: 10.966
Weighted Vector: 22.9410 -0.936 1.189 0.218 0.669 -2.105 2.751
-3.123 2.961 -2.454 -2.007 0.9055 -4.057

1.4 Extra e

It is ture that we can have a similiar or smaller Mean training error.
Mean training error: 21.781
Mean test error: 24.087

2 Problem2: Oder and Training Size

2.1 Different Orders

With order list: 0,1,2,3,4
Train Error : 5.994. 3.311. 7.13. 9.4279. 130.469 Test Error : 9.4826 . 4.815.

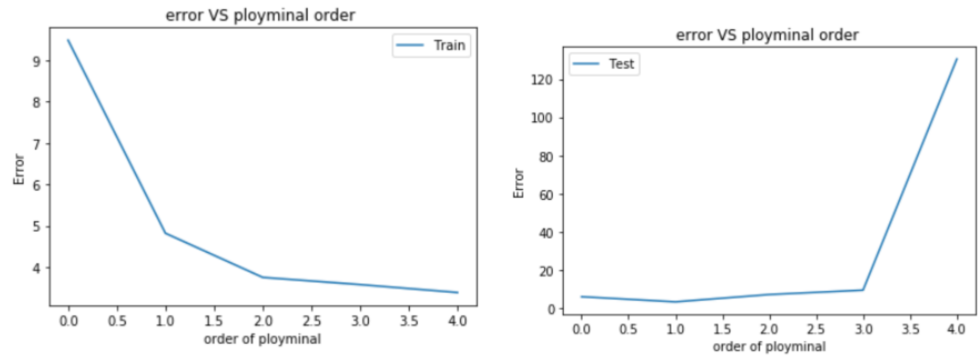

Figure 3: Error Vs Orders

3.7481. 3.5746. 3.3830

2.2 Different Size of Training Set

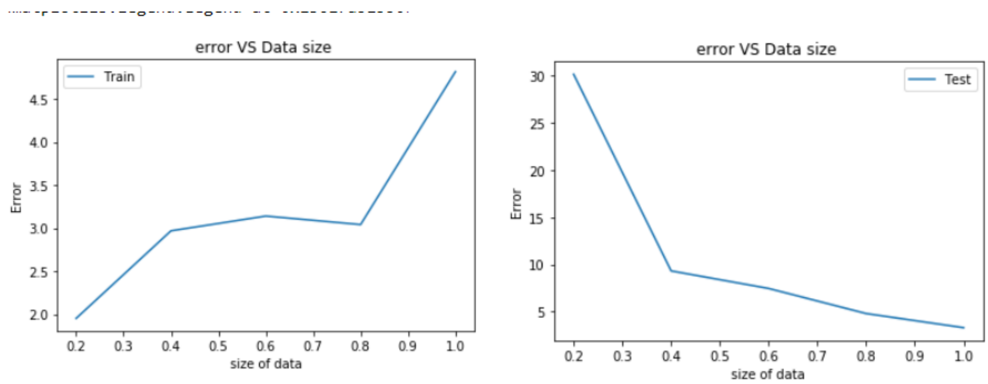

Figure 4: Error vs Size of Train

Sequence train set with size of 20%,40%,60%,80%,100%
Train Error: 1.9514, 2.9680, 3.1393, 3.0408, 4.8157
Test Error:30.1269, 9.3213, 7.4633, 4.8084, 3.3115

3 Problem3: Regularized Least Squares

3.1 Solution

$$E(\omega) = \frac{1}{2N} \sum_{i=1}^{N} (\omega^T \phi(x_n) - t_n)^2 + \frac{\lambda}{2} \parallel \omega \parallel^2$$

$$E(\omega) = \frac{1}{2N} (\Phi\omega - t)^T (\Phi\omega - t) + \frac{\lambda}{2} \omega^T \omega$$

$$E(\omega) = \frac{1}{2N}(\omega^T\Phi^T\Phi\omega - 2t^T\Phi\omega + t^Tt) + \frac{\lambda}{2}\omega^T\omega$$

$$\nabla E(\omega) = \frac{1}{2N}(2\Phi^T\Phi\omega - 2\Phi^Tt) + \lambda\omega$$

Let $\nabla E(\omega) = 0$:

$$\omega_{opt} = (\Phi^T\Phi + \lambda NI)^{-1}\Phi^Tt$$

3.2 Using Validation Set

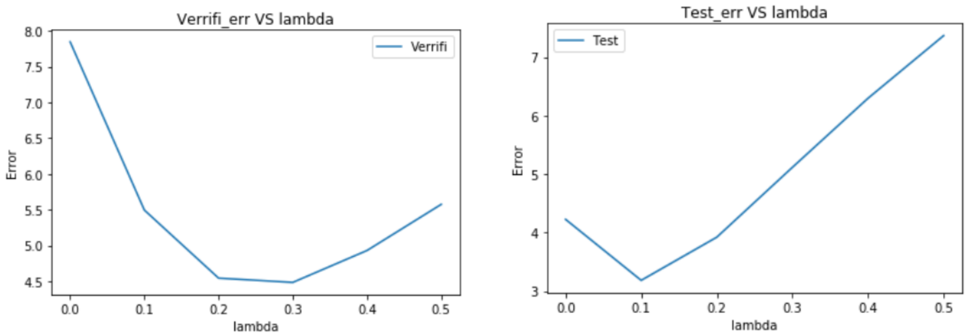

Figure 5: Error vs Lambda

Validation Error : 7.845, 5.495, 4.543, 4.484, 4.9302, 5.575
The lambda corresponding to lowest RMSE lambda = 0.3
Validation Error 4.48.
Foe this lambda the test RMSE is 5.116

4 Problem 4: Weighted Linear Regression

4.1 Proof

$$E(\omega) = \frac{1}{2}\sum_{i=1}^{N} r_n(\omega^Tx_n - t_n)^2$$

$$E(\omega) = \frac{1}{2}\sum_{i=1}^{N}(\sqrt{r_n}\omega^Tx_n - \sqrt{r_n}t_n)^T(\sqrt{r_n}\omega^Tx_n - \sqrt{r_n}t_n)$$

Turn sum to be matrix :

$$E(\omega) = (X\omega - t)^TR(X\omega - t)$$

N= number of data
M= number of feature

$$\text{Matrix } X = \begin{bmatrix} X_0(x_1) & X_1(x_1) & X_2(x_1) & \cdots & X_{M-1}(x_1) \\ X_0(x_2) & X_1(x_2) & X_2(x_2) & \cdots & X_{M-1}(x_2) \\ \vdots & \vdots & \ddots & \vdots & \\ X_0(x_N) & X_1(x_N) & X_2(x_N) & \cdots & X_{M-1}(x_N) \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}, X \subset R^{N\times M}$$

Vector: Weighted $w \subset R^{M\times1}$,

Vector: Target $t = \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_N \end{bmatrix}, t \subset R^{N\times1}$

Matrix $R = \frac{1}{2}\begin{bmatrix} r_1 & 0 & \cdots & 0 \\ 0 & r_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & r_n \end{bmatrix}, R \subset R^{N\times N}$

4.2 Solution: Wopt

$$E(\omega) = \frac{1}{2}\textstyle\sum_{i=1}^{N}(\sqrt{r_n}\omega^T x_i - \sqrt{r_i}t_i)^T(\sqrt{r_i}\omega^T x_i - \sqrt{r_i}t_i)$$

$$E(\omega) = \parallel \sqrt{R}X\omega - \sqrt{R}t \parallel^2$$

$$\text{LS:} \quad \omega_{opt} = (\sqrt{R}X)^\dagger\sqrt{R}t$$

4.3 Maximum Likelihood Estimation

$$\omega_{ML} =: argmax\prod(N(t_n|\omega^T\Phi(x_i)), \sigma_i^2)$$

$$N(t_i|\omega^T\Phi(x_n)), \sigma_i^2) = \frac{1}{\sqrt{2\pi\sigma}}exp(-\frac{(t_i - \omega^T x_i)^2}{2\sigma_i^2})$$

$$ln\frac{1}{\sqrt{2\pi\sigma}}exp(-\frac{(t_i - \omega^T x_i)^2}{2\sigma_i^2}) = -\frac{1}{2}ln(2\pi\sigma_i^2) - \frac{1}{2\sigma_i^2}(\omega^T\Phi(x_i) - t_i)^2$$

$$argmaxf(x) = argmaxln(f(x))$$

$$\omega_{ML} =: argmax\textstyle\sum_{i=1}^{N} -\frac{1}{2\sigma_i^2}(\omega^T\Phi(x_i) - t_i)^2$$

$$\omega_{ML} =: argmin\textstyle\sum_{i=1}^{N} \frac{1}{2\sigma_i^2}(\omega^T\Phi(x_i) - t_i)^2$$

Then after comparing,we can find $r_i = 1/\sigma_i^2$, so if we found the optimal w for weighted linear regression,by doing replacement$r_i = 1/\sigma_i^2$,we could at the same time solve the ML estimation.