4033/5033 Assignment: Lasso

Sasank Sribhashyam

In this assignment, we will solve an optimization problem with an absolute value function. We will also implement Lasso and evaluate it on the Crimes and Communities data set. Split the data set into a training set S and a testing set T

Task 1. Find all critical points of the following function $L(b)$. Treat x, y, λ as (unknown) constants in analysis, and represent any condition as a constraint on x e.g., when $x > \ldots$, the critical point is ...

$$L(b) = 2(b-x)^2 + \lambda|b-y|. \tag{1}$$

To find the critical points of the function $L(b)$, we'll first take the derivative of $L(b)$ with respect to b and then set it equal to zero. We'll also consider any constraints on x and b that might affect the critical points.

Let's start by finding the derivative of $L(b)$ with respect to b:

$$L(b) = 2(b-x)^2 + \lambda|b-y|$$

Now, take the derivative:

$$\frac{dL(b)}{db} = 2 \cdot 2(b-x) \cdot 1 + \lambda \cdot \frac{d|b-y|}{db}$$

The derivative of the absolute value function $|b-y|$ is not defined when $b = y$. However, we can split the derivative into two cases:

1. When $b < y$:

$$\frac{d|b-y|}{db} = -1$$

2. When $b > y$:

$$\frac{d|b-y|}{db} = 1$$

Now, let's consider both cases separately:

Case 1: When $b < y$:

$$\frac{dL(b)}{db} = 2 \cdot 2(b-x) - \lambda$$

Case 2: When $b > y$:

$$\frac{dL(b)}{db} = 2 \cdot 2(b-x) + \lambda$$

To find the critical points, we set each derivative equal to zero:

Case 1:

$$2 \cdot 2(b-x) - \lambda = 0$$

Solve for b:

$$4(b - x) = \lambda b - x = \frac{\lambda}{4} b = \frac{\lambda}{4} + x$$

Case 2:

$$2 \cdot 2(b - x) + \lambda = 0$$

Solve for b:

$$4(b - x) = -\lambda b - x = -\frac{\lambda}{4} b = -\frac{\lambda}{4} + x$$

Now, let's consider the constraints on x:

If $x > \frac{\lambda}{4}$, the critical point is given by:

$$b = \frac{\lambda}{4} + x$$

If $x < -\frac{\lambda}{4}$, the critical point is given by:

$$b = -\frac{\lambda}{4} + x$$

So, we have found the critical points of the function $L(b)$ subject to the specified conditions.

Task 2. Implement Lasso based on coordinate descend (CD) from scratch. The algorithm is in lecture note.

Task 3. Apply Lasso to train a linear model on S and evaluate it on T. Pick a proper λ yourself and report (i) testing error and (ii) model sparsity versus the number of CD updates in Figure 1 and 2 respectively. (Define model sparsity as the number of zero elements in β divided by β's size.) Both figures should have the same range for x-axis, starting from 1 and ending at any proper value you picked.

Task 4. Apply Lasso to train a linear model on S and evaluate it on T. Report (i) (converged) testing error and (ii) model sparsity versus λ in Figure 3 and 4 respectively. Pick 5 values of λ yourself.

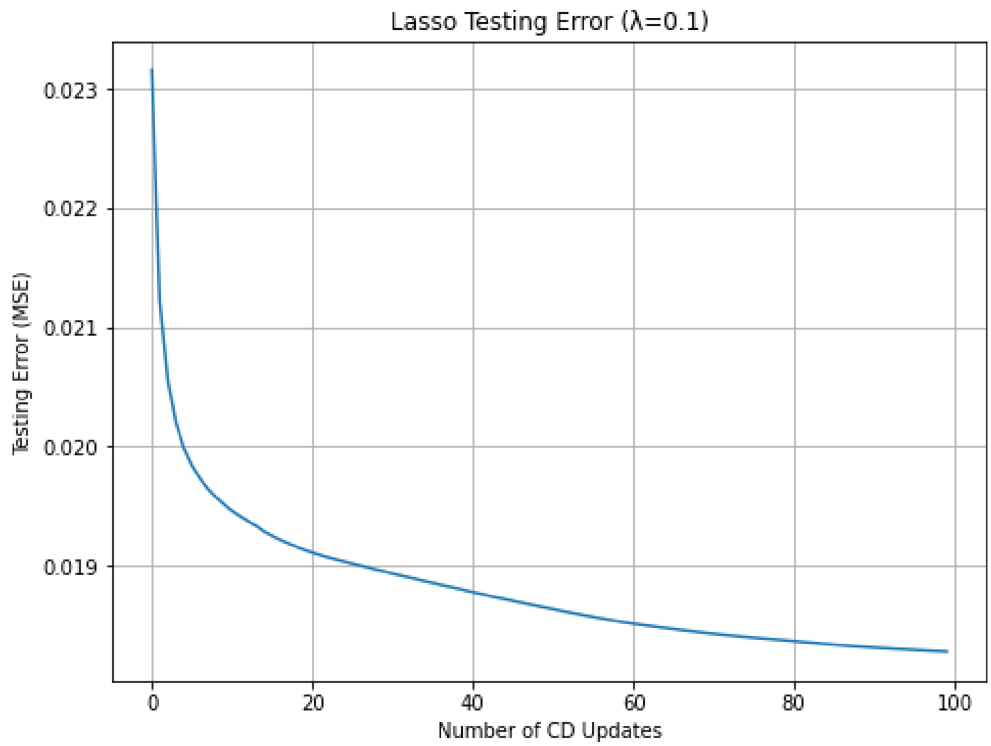

Fig. 1. Testing Error versus CD Updates with $\lambda = 0.1$

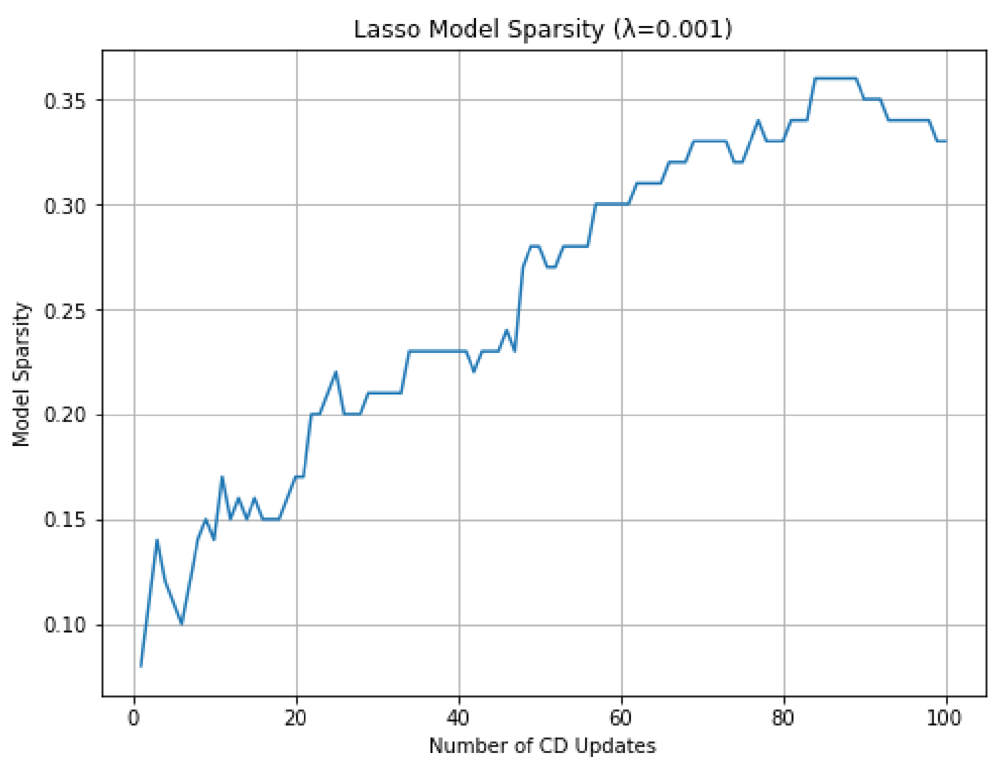

Fig. 2. Model Sparsity versus CD Updates with $\lambda = 0.001$

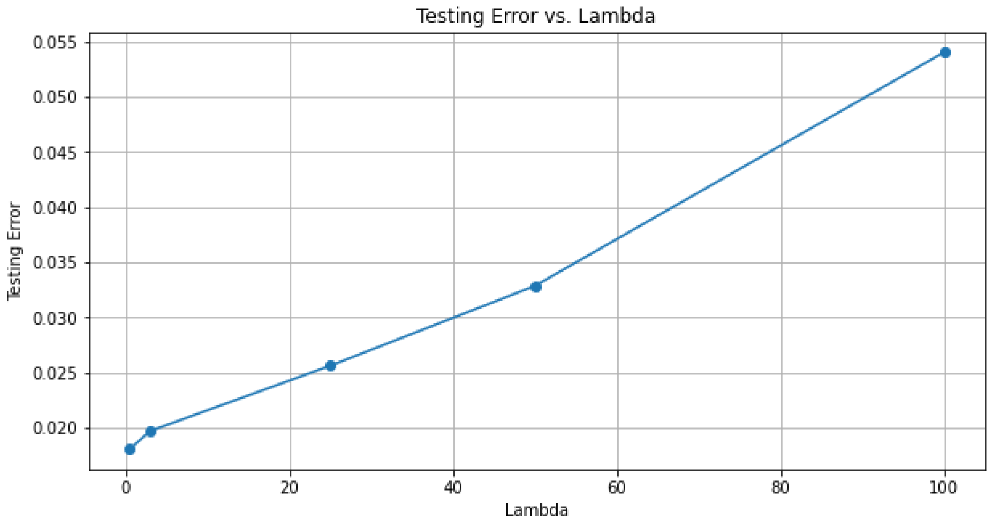

Fig. 3. Testing Error versus λ

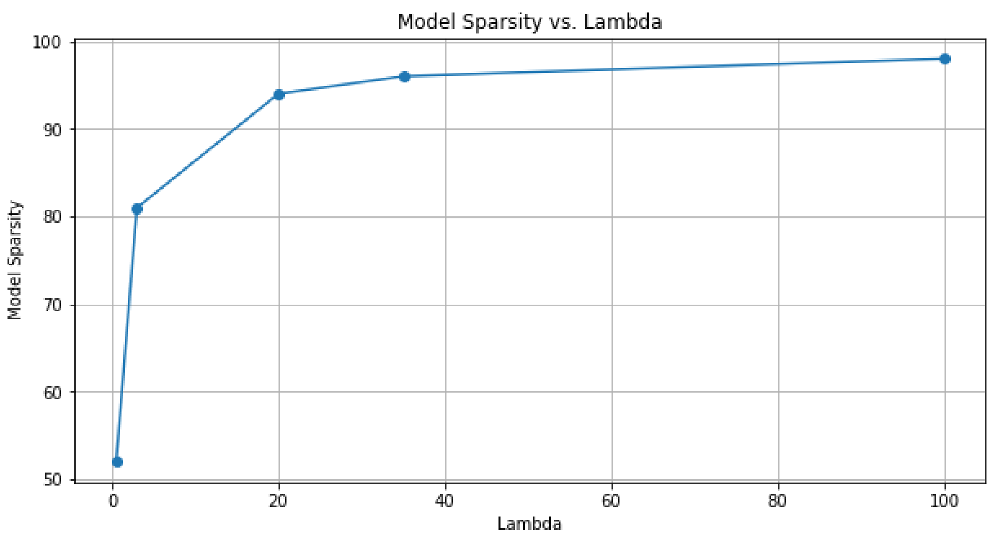

Fig. 4. Model Sparsity versus λ