Coordinate Descent

Nipun Batra

IIT Gandhinagar

August 30, 2025

Coordinate Descent for Unregularized Regression

- Express error as a difference of y_i and $\hat{y}_i$

$$\hat{y}_i = \sum_{j=0}^{d} \theta_j x_i^j = \theta_0 x_i^0 + \theta_1 x_i^1 + \theta_2 x_i^2 ...... + \theta_d x_i^d \tag{1}$$

$$\begin{aligned}
\epsilon_i &= y_i - \hat{y}_i && (2)\\
&= y_i - \theta_0 x_i^0 + \theta_1 x_i^1 + ... + \theta_d x_i^d && (3)\\
&= y_i - \sum_{j=0}^{d} \theta_j x_i^j && (4)
\end{aligned}$$

Coordinate Descent for Unregularized regression

$$\sum_{i=1}^{N} \epsilon^2 = RSS = \sum_{i=1}^{N} \left(y_i - \left(\theta_0 x_i^0 + \dots \quad \theta_j x_i^j + \theta_d x_i^d \right) \right)^2$$

Coordinate Descent for Unregularized regression

$$\sum_{i=1}^{N} \epsilon^2 = RSS = \sum_{i=1}^{N} \left(y_i - \left(\theta_0 x_i^0 + \dots \quad \theta_j x_i^j + \theta_d x_i^d \right) \right)^2$$

$$\frac{\partial \, \mathrm{RSS} \left(\theta_j \right)}{\partial \theta_j} = 2 \sum_{i=1}^{N} \left(y_i - \left(\theta_0 x_i^0 + \dots \quad \theta_j x_i^j + \dots \right) \right) \left(-x_i^j \right)$$

Coordinate Descent for Unregularized regression

$$\sum_{i=1}^{N} \epsilon^2 = RSS = \sum_{i=1}^{N} \left(y_i - \left(\theta_0 x_i^0 + \dots \quad \theta_j x_i^j + \theta_d x_i^d\right)\right)^2$$

$$\frac{\partial \, \mathrm{RSS}\,(\theta_j)}{\partial \theta_j} = 2 \sum_{i=1}^{N} \left(y_i - \left(\theta_0 x_i^0 + \dots \quad \theta_j x_i^j + \dots\right)\right) \left(-x_i^j\right)$$

$$= 2 \sum_{i=1}^{N} \left(y_i - \left(\theta_0 x_i^0 + \dots + \theta_d x_i^d\right)\right) \left(-x_i^j\right) + 2 \sum_{i=1}^{N} \theta_j (x_i^j)^2$$

Coordinate Descent for Unregularized regression

$$\sum_{i=1}^{N} \epsilon^2 = RSS = \sum_{i=1}^{N} \left(y_i - \left(\theta_0 x_i^0 + \dots \quad \theta_j x_i^j + \theta_d x_i^d \right) \right)^2$$

$$\frac{\partial \, \mathrm{RSS}\,(\theta_j)}{\partial \theta_j} = 2 \sum_{i=1}^{N} \left(y_i - \left(\theta_0 x_i^0 + \dots \quad \theta_j x_i^j + \dots \right) \right) \left(-x_i^j \right)$$

$$= 2 \sum_{i=1}^{N} \left(y_i - \left(\theta_0 x_i^0 + \dots + \theta_d x_i^d \right) \right) \left(-x_i^j \right) + 2 \sum_{i=1}^{N} \theta_j (x_i^j)^2$$

where:

$$\hat{y}_i^{(-j)} = \theta_0 x_i^0 + \dots + \theta_d x_i^d$$

is $\hat{y}_i$ without θ_j

Coordinate Descent for Unregularized regression

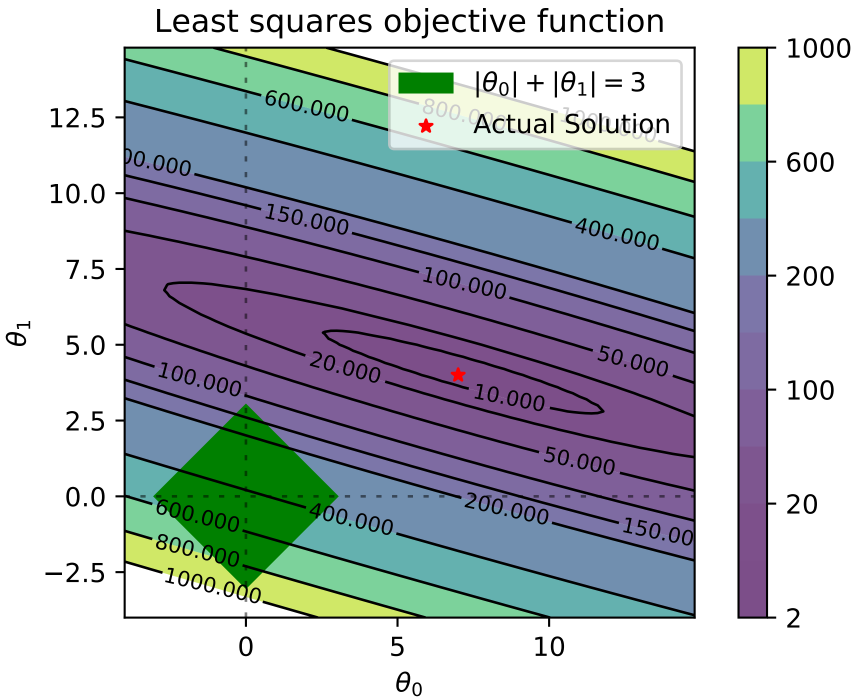
$$Set \frac{\partial \, \mathrm{RSS}\,(\theta_j)}{\partial \theta_j} = 0$$

$$\theta_j = \sum_{i=1}^{N} \frac{\left(y_i - \left(\theta_0 x_i^0 + \ldots + \ldots + \theta_d x_i^d\right)\right)\left(x_i^j\right)}{\left(x_i^j\right)^2} = \frac{\rho_j}{z_j}$$

$$\rho_j = \sum_{i=1}^{N} x_i^j \left(y_i - \hat{y}_i^{(-j)})\right)$$

$$z_j = \sum_{i=1}^{N} \left(x_i^j\right)^2$$

z_j is the squared of ℓ_2 norm of the j^{th} feature


Least squares objective function
$|\theta_0| + |\theta_1| = 3$
Actual Solution
600.000
800.000
400.000
150.000
100.000
50.000
20.000
10.000
200.000
1000.000
θ_1
θ_0
12.5
10.0
7.5
5.0
2.5
0.0
−2.5
0
5
10
1000
600
200
100
20
2

Coordinate Descent for Lasso Regression

$$\underbrace{\text{Minimize} \sum_{i=1}^{N} \epsilon^2 + \delta^2 \{|\theta_0| + |\theta_1| + \dots |\theta_j| + \dots |\theta_d|\}}_{\textit{LASSO OBJECTIVE}}$$

$$\frac{\partial}{\partial \theta_j}(\text{LASSO OBJECTIVE}) = -2\rho_j + 2\theta_j z_j + \delta^2 \frac{\partial}{\partial \theta_j} |\theta_j|$$

$$\frac{\partial}{\partial \theta_j} |\theta_j| = \begin{cases} 1 & \theta_j > 0 \\ [-1, 1] & \theta_j = 0 \\ -1 & \theta_j < 0 \end{cases}$$

Coordinate Descent for Lasso Regression

- **Case 1:** $\theta_j > 0$

$$2\rho_j + 2\theta_j z_j + \delta^2 = 0$$
$$\theta_j = \frac{\rho_j - \frac{\delta^2}{2}}{z_j}$$
$$\rho_j > \frac{\delta^2}{2} \Rightarrow \theta_j = \frac{\rho_j - \frac{\delta^2}{2}}{z_j}$$

Coordinate Descent for Lasso Regression

- **Case 1:** $\theta_j > 0$

$$
\begin{aligned}
2\rho_j + 2\theta_j z_j + \delta^2 &= 0 \\
\theta_j &= \frac{\rho_j - \frac{\delta^2}{2}}{z_j} \\
\rho_j > \frac{\delta^2}{2} \Rightarrow \theta_j &= \frac{\rho_j - \frac{\delta^2}{2}}{z_j}
\end{aligned}
$$

- **Case 2:** $\theta_j < 0$

$$
\rho_j < \frac{\delta^2}{2} \Rightarrow \theta_j = \frac{\rho_j + \delta^2/2}{z_j} \tag{5}
$$

Coordinate Descent for Lasso Regression

- **Case 3:** $\theta_j = 0$

$$\frac{\partial}{\partial \theta_j}(\text{LASSO OBJECTIVE}) = -2\rho_j + 2\theta_j z_j + \delta^2 \underbrace{\frac{\partial}{\partial \theta_j}|\theta_j|}_{\text{[-1,1]}}$$

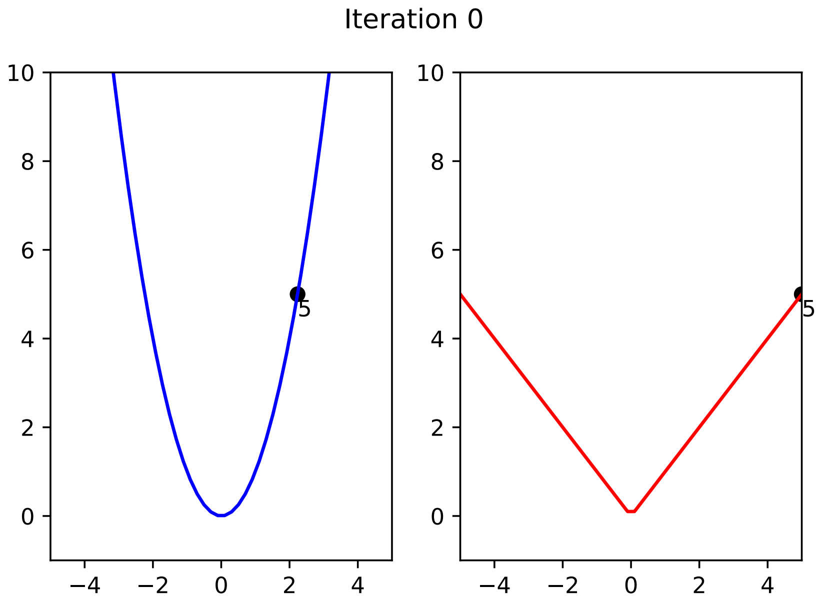
$$\epsilon \underbrace{\left[-2\rho_j - \delta^2, -2\rho_j + \delta^2\right]}_{\{0\}\text{ lies in this range}}$$

$$-2\rho_j - \delta^2 \leq 0 \text{ and } -2\rho_j - \delta^2 \leq 0$$

$$-\frac{\delta^2}{2} \leq \rho_j \leq \frac{\delta^2}{2} \Rightarrow \theta_j = 0$$

Summary of Lasso Regression

$$\theta_j = \begin{bmatrix} \frac{\rho_j + \frac{\delta^2}{2}}{z_j} & \textit{if} & \rho_j < -\frac{\delta^2}{2} \\ 0 & \textit{if} & -\frac{\delta^2}{2} \leq \rho_j \leq \frac{\delta^2}{2} \\ \frac{\rho_j - \frac{\delta^2}{2}}{z_j} & \textit{if} & \rho_j > \frac{\delta^2}{2} \end{bmatrix} \tag{6}$$



Iteration 0
10
8
6
4
2
0
−4
−2
0
2
4
5
10
8
6
4
2
0
−4
−2
0
2
4
5