LASSO: Coordinate Descent Derivation

Step 1: Original Objective (Penalized Lagrangian)

$$\min_{\beta_0,\, \boldsymbol{\beta}\in\mathbb{R}^N} \; \frac{1}{M}\sum_{i=1}^{M}\Big(y_i - (\beta_0 + \boldsymbol{\beta}^\top \boldsymbol{x}_i)\Big)^2 \;+\; \lambda\sum_{j=1}^{N}|\beta_j|. \tag{1}$$

Step 2: Eliminating the Intercept (Centering)

Differentiate with respect to β_0:

$$-\frac{2}{M}\sum_{i=1}^{M}\Big(y_i - \beta_0 - \boldsymbol{\beta}^\top \boldsymbol{x}_i\Big) = 0 \implies \beta_0^* = \bar{y} - \bar{\boldsymbol{x}}^\top \boldsymbol{\beta},$$

where $\bar{y} = \frac{1}{M}\sum_i y_i,\ \bar{\boldsymbol{x}} = \frac{1}{M}\sum_i \boldsymbol{x}_i$.

Define the centered variables:

$$\tilde{y}_i := y_i - \bar{y}, \qquad \tilde{\boldsymbol{x}}_i := \boldsymbol{x}_i - \bar{\boldsymbol{x}}.$$

Then the problem is equivalent to

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^N} \; \frac{1}{M}\sum_{i=1}^{M}\Big(\tilde{y}_i - \boldsymbol{\beta}^\top \tilde{\boldsymbol{x}}_i\Big)^2 \;+\; \lambda\|\boldsymbol{\beta}\|_1. \tag{2}$$

Step 3: Vectorized Notation

Let

$$X = [\,\tilde{\boldsymbol{x}}_1^\top; \ldots; \tilde{\boldsymbol{x}}_M^\top\,] \in \mathbb{R}^{M\times N}, \qquad \boldsymbol{y} = [\,\tilde{y}_1, \ldots, \tilde{y}_M\,]^\top.$$

Then (2) can be written as

$$\min_{\boldsymbol{\beta}} \; \tfrac{1}{M}\|\boldsymbol{y} - X\boldsymbol{\beta}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1.$$

Step 4: One-Dimensional Subproblem in Coordinate Descent

Take the j-th column $x_j \in \mathbb{R}^M$ and define the partial residual excluding j:

$$\boldsymbol{r}^{(-j)} := \boldsymbol{y} - \sum_{k\neq j} x_k\beta_k = \boldsymbol{y} - X\boldsymbol{\beta} + x_j\beta_j.$$

Fixing other coordinates, the subproblem for β_j is:

$$\min_{\beta_j} \ \frac{1}{M} \left\| \boldsymbol{r}^{(-j)} - x_j \beta_j \right\|_2^2 + \lambda |\beta_j|. \tag{3}$$

Expanding (and dropping constants):

$$f(\beta_j) = \frac{a_j}{2} \beta_j^2 - b_j \beta_j + \lambda |\beta_j|,$$

where

$$a_j := \frac{2}{M} \|x_j\|_2^2, \qquad b_j := \frac{2}{M} x_j^\top \boldsymbol{r}^{(-j)}.$$

Step 5: Subgradient First-Order Condition

The derivative of the smooth part is $a_j \beta_j - b_j$, while the subgradient of $|\beta_j|$ is

$$\partial |\beta_j| = \begin{cases} \{+1\}, & \beta_j > 0, \\ [-1, 1], & \beta_j = 0, \\ \{-1\}, & \beta_j < 0. \end{cases}$$

The first-order condition is therefore

$$0 \in a_j \beta_j - b_j + \lambda \, \partial |\beta_j|.$$

Case analysis yields:

$$\beta_j = \begin{cases} \dfrac{b_j - \lambda}{a_j}, & b_j > \lambda, \\ 0, & |b_j| \le \lambda, \\ \dfrac{b_j + \lambda}{a_j}, & b_j < -\lambda. \end{cases}$$

This can be unified as the soft-thresholding operator:

$$\boxed{\beta_j \ \leftarrow \ \frac{S(b_j, \lambda)}{a_j}, \quad S(b, \lambda) = \operatorname{sign}(b) \ \max\{|b| - \lambda, \, 0\}} \tag{1}$$

Step 6: Practical Implementation (Residual Update)

Maintain the global residual:

$$\boldsymbol{r} := \boldsymbol{y} - X\boldsymbol{\beta}.$$

When updating the j-th coordinate:

$$\tilde{\boldsymbol{r}} := \boldsymbol{r} + x_j\beta_j = \boldsymbol{r}^{(-j)}, \quad b_j = \tfrac{2}{M}\, x_j^\top \tilde{\boldsymbol{r}}, \quad \beta_j \leftarrow \tfrac{S(b_j,\lambda)}{a_j}, \quad \boldsymbol{r} \leftarrow \tilde{\boldsymbol{r}} - x_j\beta_j.$$

Step 7: Critical λ

At $\boldsymbol{\beta} = \mathbf{0}$, we have $\boldsymbol{r}^{(-j)} = \boldsymbol{y}$. Thus

$$\lambda_{\max} = \tfrac{1}{M} \max_j |x_j^\top \boldsymbol{y}|.$$

For $\lambda \geq \lambda_{\max}$, the optimal solution is $\boldsymbol{\beta}^* = \mathbf{0}$.

Finally, the intercept is recovered as

$$\beta_0^* = \bar{y} - \bar{\boldsymbol{x}}^\top \boldsymbol{\beta}^*.$$