University of Burgundy

Department of Mathematics

Application of Coordinate Descent to Lasso and Practical Study on the Ames Housing Dataset

Mhamed Halata
Master 1 MIGS 2024/2025

May 18, 2025

Introduction

As part of our Master's program in Mathematics, Computer Science, and Statistical Engineering (MIGS), we are required to conduct an in-depth study on a fundamental optimization mechanism in data science: coordinate descent. This method, while seemingly simple, proves to be very powerful when adapted to real-world problems, particularly the Lasso problem.

The Lasso (Least Absolute Shrinkage and Selection Operator) is a variant of linear regression that introduces an penalty on the model coefficients. This regularization term not only limits the model's complexity but also performs variable selection by setting some coefficients to zero.

This report is divided into two main parts: a theoretical part where we rigorously formulate the Lasso problem and show how coordinate descent can be effectively applied to it, and a practical part where we implement this method on a real dataset: the Ames Housing dataset. This rich and well-documented dataset allows us to apply analysis and optimization tools in a realistic context.

We will also compare the performance of our manual implementation with that of the `scikit-learn` library and discuss the encountered difficulties, particularly regarding convergence.

1 Application to the Lasso Method

1.1 Formulation of the Lasso Problem

The Lasso (Least Absolute Shrinkage and Selection Operator) is a variant of linear regression that introduces an penalty on the model coefficients. This regularization term allows not only to limit the model's complexity but also to perform variable selection by setting some coefficients to zero.

The objective is then to solve the following problem:

$$\min_{\theta \in \mathbb{R}^p} \left(\frac{1}{2} \|y - X\theta\|_2^2 + \lambda \|\theta\|_1 \right) \tag{1}$$

- $\|\theta\|_1 = \sum_{j=1}^{p} |\theta_j|$ is the norm,
- $\lambda > 0$ is a regularization parameter that controls the penalty strength.

This problem is convex but non-differentiable due to the norm, which prevents the direct use of classical gradient-based optimization methods.

1.2 Coordinate Descent Applied to Lasso

Since the norm is separable by coordinate, Lasso is well-suited for a coordinate descent solution. We update one coefficient at a time while keeping the others fixed.

Soft-Thresholding Operator

At each step, the univariate subproblem to solve is:

$$\min_{\theta_j \in \mathbb{R}} \left(\frac{1}{2} \|r_j - X_j\theta_j\|_2^2 + \lambda |\theta_j| \right) \tag{2}$$

Where $r_j = y - \sum_{k \neq j} X_k \theta_k$ is the partial residual. The solution to this problem is given by the soft-thresholding operator, defined as follows:

$$\eta_\lambda(z) = \text{sign}(z) \cdot \max(|z| - \lambda, 0) \tag{3}$$

This operator ("soft thresholding") reduces the absolute value of z by and sets to zero values that are too small.

Complete Algorithm

- Initialize $\theta = 0$, $r = y$
- While has not converged:
  - For each coordinate $j \in \{1, \ldots, p\}$:
    - Compute $z_j = x_j^T r + \|x_j\|_2^2 \theta_j$
    - Update $\theta_j^{\text{new}} = \eta_{\lambda/\|x_j\|_2^2}(z_j/\|x_j\|_2^2)$
    - Update the residual $r \leftarrow r + x_j(\theta_j - \theta_j^{\text{new}})$
    - $\theta_j \leftarrow \theta_j^{\text{new}}$

2 Practical Study: Ames Housing Dataset

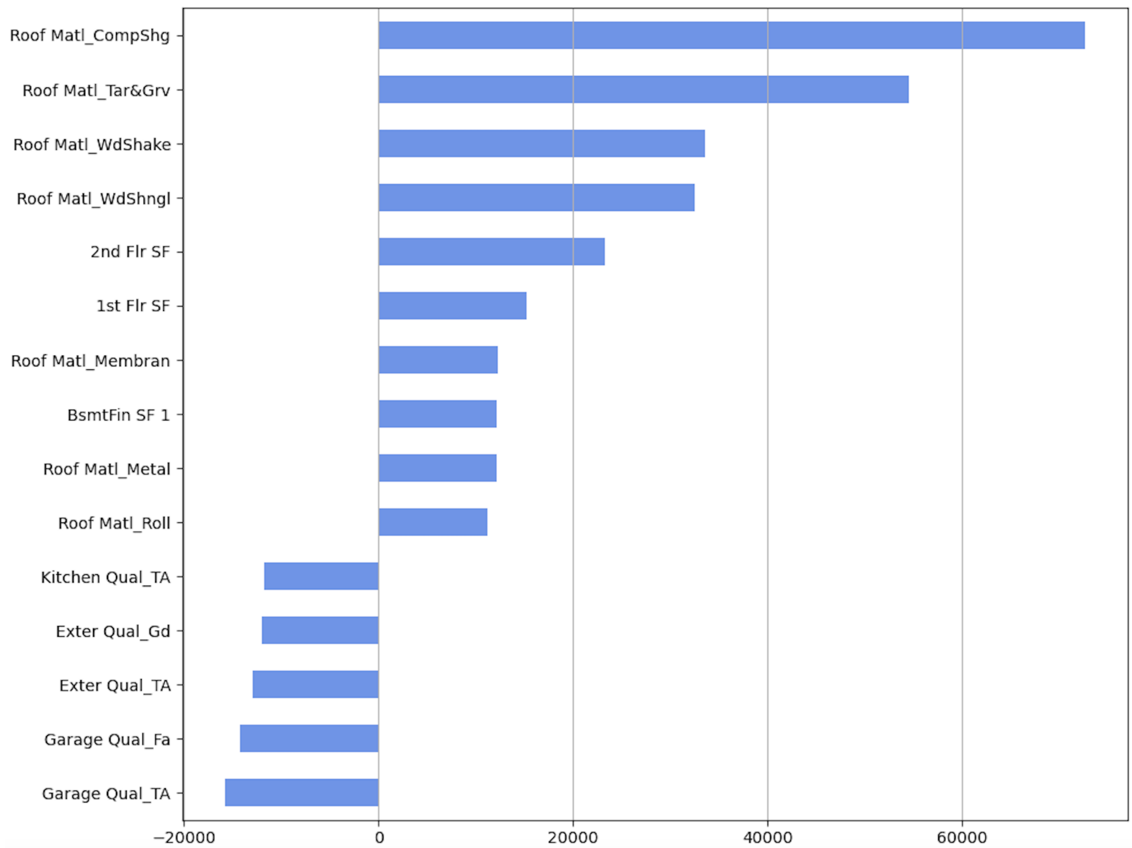
We chose to implement our coordinate descent algorithm on a real dataset: Ames Housing. This dataset is well-known for linear regression as it provides numerous precise characteristics about houses sold in Ames, Iowa (USA).

2.1 Data Preprocessing

- The dataset initially contains 82 columns. We first removed non-informative columns such as the identifier `PID` and `Order`.
- We analyzed missing values:
  - Columns with more than 40% missing values were removed.
  - For remaining columns, NaN values were replaced by the mean (quantitative variables) or the mode (qualitative variables).
- Qualitative variables were transformed using One-Hot Encoding (binary encoding), bringing the total number of columns to 243.
- Explanatory variables were standardized (centered and scaled).

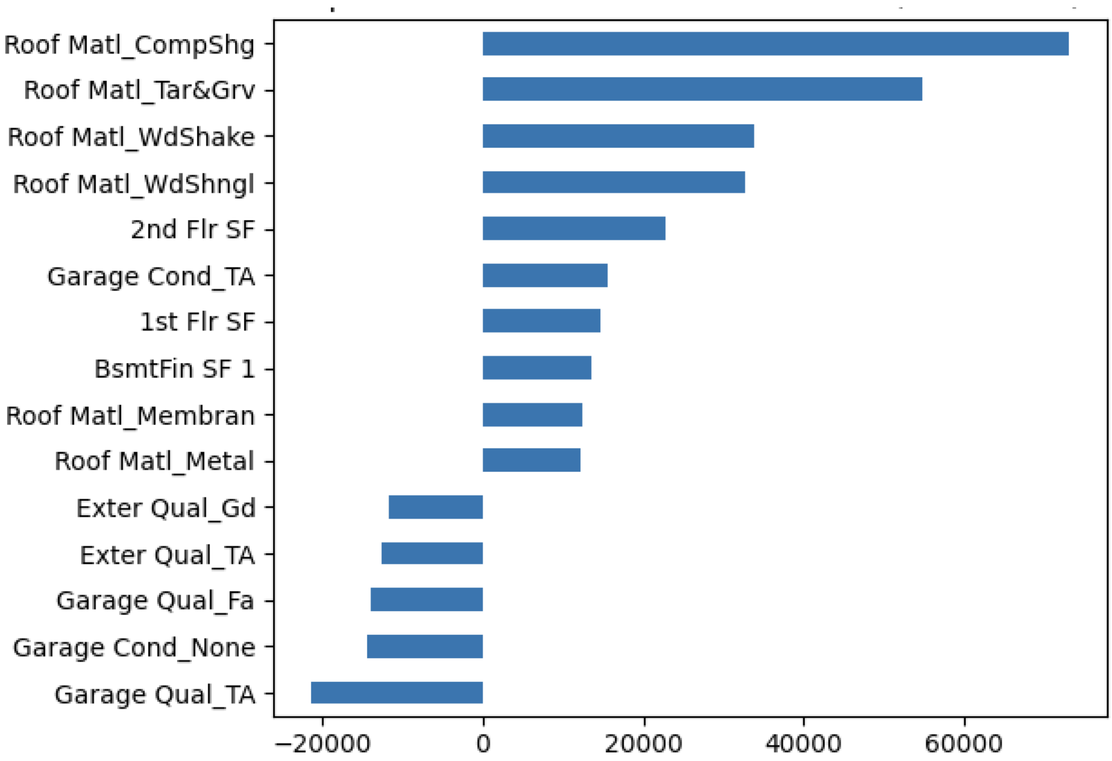
2.2 Manual Implementation of Lasso

- We implemented the coordinate descent algorithm with soft-thresholding.
- The convergence tolerance $\varepsilon = 10^{-6}$ and $\lambda = 0.1$ were used.
- The algorithm did not converge even with a large number of iterations (5000).
- The obtained coefficients were sorted by importance, and we visualized the 15 most influential variables.
- Among these: `1st Flr SF`, `2nd Flr SF`, `Overall Qual`, `Kitchen Qual`, `Roof Matl_CompShg`, etc.



2.3 Application with scikit-learn

- We used `Lasso` from `scikit-learn` with the same processed data.
- The same $\lambda = 0.1$ was used.
- This time, the algorithm converged, and the obtained coefficients were very similar to the manual implementation.



2.4 Results Analysis

- We observe that variables related to area (`1st Flr SF`, `2nd Flr SF`), overall quality (`Overall Qual`), and amenities (`Kitchen Qual`, `Garage Area`) most influence the price.
- Roofing materials have a significant effect, related to extreme weather conditions in the USA (hurricanes, snow, heatwaves).
- The consistency between both methods validates our implementation despite the non-convergence of the manual algorithm.

2.5 Discussion on Non-Convergence

- The manual algorithm did not converge even with $\lambda = 100$.
- According to the community (e.g., StackExchange), this could be explained by a lack of acceleration, a poorly conditioned dataset, or a too dense structure.
- However, we tested our algorithm on a low-dimensional dataset (for example, a quadratic function like $ax^2 + bxy + cy^2$), and this time convergence was observed.
- This non-convergence on Ames Housing does not question the validity of the algorithm itself but rather shows the limits of its stability in real-world high-dimensional cases.

- The results are very similar to those obtained with `scikit-learn`, which validates the quality of the approach.