Group Importance Estimation Method Based on Group LASSO Regression

Yuki Mori[1][a], Seiji Yamada[2][b] and Takashi Onoda[1][c]
[1]Aoyama Gakuin University School of Science and Engineering, 5-10-1 Huchinobe, Chuo, Sagamihara, Kanagawa, Japan
[2]National Institute of Informatics 2-1-2 Hitotsubashi, Chiyoda, Tokyo, Japan

Keywords: Group LASSO Regression, Machine Learning, LASSO Regression, Group Variable Selection, Estimate Group Importance, Linear Regression Problem, Penalized Least Squares Problem.

Abstract: There has been a rapidly growing interest in penalized least squares problems via l_1 regularization. The LASSO (Least Absolute Shrinkage and Selection Operator) regression, which utilizes l_1 regularization, has gained popularity as a method for model selection and shrinkage estimation. An important extension of LASSO regression is Group LASSO regression, which generates sparse models at the group level. However, Group LASSO regression does not directly evaluate group importance. In this study, we propose a method to assess group importance based on Group LASSO regression. This method leverages regularization parameters to estimate the importance of each group. We applied this method to both synthetically generated data and real-world data, conducting experiments to evaluate its performance. As a result, the method accurately approximated the importance of groups, enhancing the interpretability of models at the group level.

1 INTRODUCTION

Recently, there has been a rapidly growing interest in penalized least squares problems via l_1 regularization(Nardi and Rinaldo, 2008). The LASSO (Least Absolute Shrinkage and Selection Operator) regression(Tibshirani, 1996) is a regularization technique where the penalty for model complexity is the l_1 norm of the estimated coefficients. Originally developed for linear regression models, LASSO regression has gained popularity as a method for model selection and shrinkage estimation. Group LASSO regression(Yuan and Lin, 2006), which selects key explanatory factors in a grouped manner, is an important extension of LASSO regression(Yang et al., 2010).This method has found successful applications in various fields, including birthweight prediction and gene finding(Yuan and Lin, 2006)(Meier et al., 2008). However, while it can yield solutions with sparsity at the group level, it doesn't inherently assess the individual group importance. In this study, we propose to assess group importance based on Group LASSO regression, thereby enhancing interpretability and addressing the limitations of the existing approach.

In the subsequent section of this paper, we delve into Group LASSO regression. In section 3, we detail the proposed method for estimating group importance based on Group LASSO regression. Moving to section 4, we engage in experimental validations, testing the efficacy of our approach using both synthetically generated data and real data. Finally, we conclude by summarizing the key findings of our study and discussing the adaptability of our proposed approach in real-world scenarios.

2 GROUP LASSO REGRESSION

In this section, we present Group LASSO regression(Yuan and Lin, 2006) and proximal gradient algorithms for Group LASSO regression(Tomioka and Scientific, 2015).

Consider a linear model where we have independent and identically distributed observations $(x_{(1)i}, \cdots, x_{(M)i}, y_i), i = 1, \cdots, N$, of a N-dimensional vector $\vec{x}_k = (x_{(k)1}, \cdots, x_{(k)N})^\top$, $\vec{y} = (y_1, \cdots, y_N)^\top$, matrix $X = (\vec{x}_1, \cdots, \vec{x}_M)$ and the parameter vector $\vec{\beta} \in \mathbb{R}^M$, which holds the coefficients for each feature. The relationship between the features and the response in

[a] https://orcid.org/0009-0007-9122-629X
[b] https://orcid.org/0000-0002-5907-7382
[c] https://orcid.org/0000-0002-5432-0646

Mori, Y., Yamada, S. and Onoda, T.
Group Importance Estimation Method Based on Group LASSO Regression.
DOI: 10.5220/0012304800003654

In *Proceedings of the 13th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2024)*, pages 197-204
ISBN: 978-989-758-684-2; ISSN: 2184-4313

this model is given by:

$$\vec{y} = X\vec{\beta} + \vec{\varepsilon} \tag{1}$$

Here, $\vec{\varepsilon} \in \mathbb{R}^N$ is the error vector capturing the deviation from the linear model, assumed to be independent and identically distributed, following a normal distribution with mean 0.

In the context of Group LASSO Regression, we extend this basic linear model to accommodate a structure where the M features are partitioned into J distinct groups. Let m_j denote the number of features in the j-th group. Consequently, matrix X can be expressed as $X = (X_1, \cdots, X_J)$ where $X_j \in \mathbb{R}^{N\times m_j}$. The coefficient vector $\vec{\beta}$ is similarly decomposed into $\vec{\beta} = (\vec{\beta}_1, \cdots, \vec{\beta}_J)^\top$, where $\vec{\beta}_j = (\beta_{\{j\}1}, \cdots, \beta_{\{j\}m_j})$ for $j = 1, \cdots, J$.

The optimization problem for Group LASSO, considering the above model, is defined as:

$$\min_{\vec{\beta}} \frac{1}{2}\|\vec{y} - X\vec{\beta}\|_2^2 + \lambda \sum_{j=1}^{J} \|\vec{\beta}_j\|_2 \tag{2}$$

In this formulation, $\|\vec{\beta}_j\|_2$ represents the l_2 norm of the coefficients of the j-th group, calculated as $\|\vec{\beta}_j\|_2 = \sqrt{\sum_{p=1}^{m_j} \beta_{\{j\}p}^2}$. The parameter λ is a non-negative regularization term that introduces group-wise sparsity into the model, effectively encouraging the model to reduce the coefficients of less relevant groups to zero. This results in a more parsimonious and interpretable model that highlights the most significant group-based features for predicting the target variable.

We also introduce the proximal gradient algorithm, an iterative optimization technique used for addressing both convex and non-convex optimization problems, including LASSO regression and Group LASSO regression. It updates the differentiable components similarly to gradient descent and handles the non-differentiable components using the proximal operator. The update formula for the proximal gradient algorithm is given by the following expression, which incorporates the proximal operator:

$$\vec{\beta_{(t+1)}} = \mathrm{prox}(\vec{\beta_{(t)}} - \gamma \nabla f(\vec{\beta_{(t)}})) \tag{3}$$

where, $\vec{\beta_{(t)}}$ represents the regression coefficients $\vec{\beta}$ updated at the t-th iteration, γ is the step size and $\nabla f(\vec{\beta})$ denotes the gradient of the function $f(\vec{\beta})$, which is a differentiable convex function. The convergence criteria used in the proximal gradient algorithm are as follows:

$$\frac{||F(\vec{\beta_{(t+1)}}) - F(\vec{\beta_{(t)}})||_1}{||F(\vec{\beta_{(t+1)}})||_1} < e \approx 0 \tag{4}$$

Input: X, $\vec{y}$, λ, γ
Initialization: $t = 0$ and $\vec{\beta_{(t)}} = \vec{\beta_{(0)}}$
while *not satisfied with Equation (4)* **do**
  Calculate the gradient: $\nabla f(\vec{\beta_{(t)}})$
  Update the $\vec{\beta}$ based on Equation (3) and Equation (7)
  Update $t = t + 1$
end
Output: $\vec{\beta_{(t)}}$

Algorithm 1: Proximal Gradient Algorithm for Group LASSO Regression.

where, $F(\vec{\beta})$ represents the objedtive function, which is a convex function and e denotes a small constant close to zero. In Group LASSO regression, $F(\vec{\beta})$ and $f(\vec{\beta})$ are defined as:

$$F(\vec{\beta}) = \frac{1}{2}||\vec{y} - X\vec{\beta}||_2^2 + \lambda \sum_{j=1}^{J} ||\vec{\beta}_j||_2 \tag{5}$$

$$f(\vec{\beta}) = \frac{1}{2}||\vec{y} - X\vec{\beta}||_2^2 \tag{6}$$

Furthermore, in Group LASSO regression, the proximal operator is defined using the proximity point $\vec{\beta'} \in \mathbb{R}^M$ as follows:

$$\mathrm{prox}(\vec{\beta'_j}) = \begin{cases} \vec{\beta'_j} - \frac{\gamma\lambda}{||\vec{\beta'_j}||_2}\vec{\beta'_j} & ||\vec{\beta'_j}||_2 \geq \gamma\lambda \\ 0 & ||\vec{\beta'_j}||_2 < \gamma\lambda \end{cases} \tag{7}$$

Based on the equations presented in section 2, Algorithm 1 shows the flow of the proximal gradient algorithm for Group LASSO regression.

3 GROUP IMPORTANCE ESTIMATION METHOD

In this section, we propose the group importance estimation method based on Group LASSO regression. Section 3.1 lays the foundation by explaining the core concepts of our method. Following that, section 3.2 sheds light on the importance of the threshold in our approach. Finally, section 3.3 provides a comprehensive formulation of the proposed method.

3.1 Concept of the Group Importance Estimation

While Group LASSO regression is powerful in achieving sparsity at the group level, it often sidelines the assessment of individual group importance.

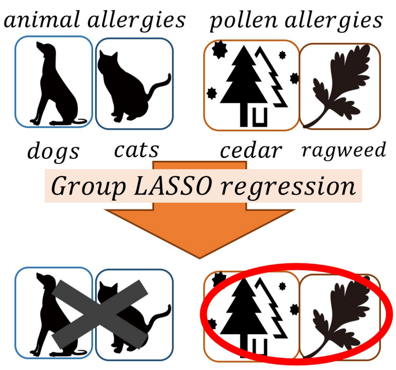

Figure 1: Example of allergy diagnosis for Group LASSO regression.

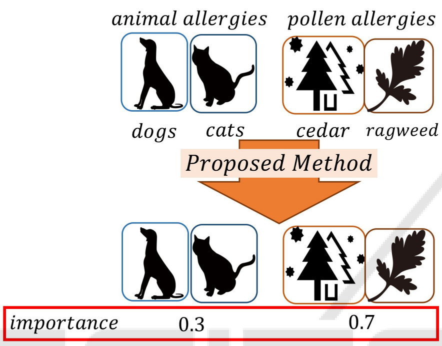

Figure 2: Example of allergy diagnosis for proposed method.

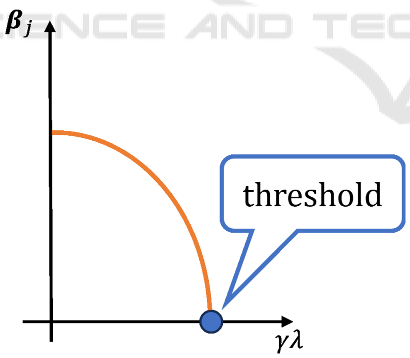

Figure 3: Illustrative representation of the relationship between $\vec{\beta}_j$ and λ.

For illustration, consider an allergy diagnosis scenario where the goal is to predict reactions to specific allergens. The predictions hinge on various symptoms, ranging from allergies to dogs and cats (categorized as animal allergies) to cedar pollen and ragweed (grouped under pollen allergies). When employing Group LASSO regression in such a context, the group variable selection might look like what's shown in Fig. 1. In this figure, ”×” indicates that the regression coefficients for animal allergies are zero, suggesting that this group's variables have been excluded. In contrast, ”◯” signals the presence of non-zero coefficients for pollen allergies, indicating their incorporation in the regression model. A quick look at Fig. 1 might suggest that pollen allergies play a pivotal role in the regression model. However, the current approach falls short in offering a nuanced comparison between the importance of animal and pollen allergies. Addressing this gap, our study proposes a method to assess the importance of such groups, as shown in Fig. 2.

Central to our method is the establishment of a common criterion to assess the relative importance of each group. To achieve this, we explore a methodology that utilizes the threshold for determining whether a group is important or not in predicting the target variable. This threshold serves as a common criterion derived from the feature selection process conducted at the group level in Group LASSO regression. With the general concept of our method outlined, we'll now delve deeper into the pivotal role the threshold plays in our approach.

3.2 Threshold for the Group Importance Estimation

A group j is considered important in predicting the target variable if $\vec{\beta}_j \neq \vec{0}$, and unimportant if $\vec{\beta}_j = \vec{0}$. As briefly described in Algorithm 1 in Section 2, if the condition $||\vec{\beta}'_j||_2 < \gamma\lambda$ holds according to Equation (7), then $\vec{\beta}_j = \vec{0}$, leading to the elimination of variables within group j. Conversely, if the condition $||\vec{\beta}'_j||_2 > \gamma\lambda$ holds, then $\vec{\beta}_j \neq \vec{0}$, and as a result, variables within group j are not reduced. Essentially, the value of $\gamma\lambda$ has an impact on variable selection at the group level. In this study, Fig. 3 shows the relationship between $\vec{\beta}_j$ and λ. Based on this depiction, we consider the point where $\vec{\beta}_j = \vec{0}$ is first achieved as a threshold common to each group. For this boundary to exist in all groups, the relationship between $\vec{\beta}_j$ and $\gamma\lambda$ must be either monotonically increasing or decreasing. To shed more light on this, we experimentally verify the relationship between $||\vec{\beta}_j||_2$ and $\gamma\lambda$. Since γ is provided as a constant, we use a path diagram to clarify this relationship between $||\vec{\beta}_j||_2$ and λ.

Experimental Conditions.
For the vector $\vec{v}[\theta,\eta]$, we define:

$$\sum_{n=1}^{1000} \vec{v}[\theta,\eta]_n = \theta, \ \sum_{n=1}^{1000} \vec{v}[\theta,\eta]_n{}^2 = \eta$$

From this, we derive the vector $\vec{x_m}$ as:

$$\vec{x_m} = \vec{v}[0,1] \ \ (m = 1, \cdots, 9)$$

The target variable, $\vec{y}$, is then generated as:

$$\vec{y} = X\vec{\beta} + \vec{v}[0,2]$$

The regression coefficient vectors $\vec{\beta_1}$, $\vec{\beta_2}$, and $\vec{\beta_3}$ are defined as:

$\vec{\beta_1} = (1.0, 1.0, 1.0)$
$\vec{\beta_2} = (2.0, 2.0, 2.0)$
$\vec{\beta_3} = (3.0, 3.0, 3.0)$

Combining these, the regression coefficient vector $\vec{\beta}$ is:

$\vec{\beta} = (\vec{\beta_1}, \vec{\beta_2}, \vec{\beta_3})^\top$

Furthermore, standardization is applied to X, and $\vec{x_m}$ is centered (mean 0). By varying the value of the regularization parameter λ in the range $[10^{-1}, 10^5]$, we aim to show the relationship between $||\vec{\beta_j}||_2$ and λ. Fig. 4 shows a path diagram that captures this relationship. Fig. 4 shows that "Overall" represents $||\vec{\beta}||_2$, "Group1" is $||\vec{\beta_1}||_2$, "Group2" is $||\vec{\beta_2}||_2$, and "Group3" is $||\vec{\beta_3}||_2$. It's evident from the figure that $||\vec{\beta_j}||_2$ decreases monotonically as λ grows. Next, based on Algorithm 1, we will explore the relationship between $||\vec{\beta_j}||_2$ and λ. From Equation (3) and Equation (7) in Algorithm 1, it becomes clear that as λ increases, the value of $||\vec{\beta_j}||_2$ is updated to become smaller. Therefore, we can infer that there is a monotonically increasing or decreasing relationship between $||\vec{\beta_j}||_2$ and λ. However, due to the convergence conditions in Equation (4), there might be scenarios where this relationship doesn't hold. Thus, the relationship between $||\vec{\beta_j}||_2$ and λ is not strictly monotonic. However, instances where this relationship does not hold are rare. From the above discussions, it's clear that a general monotonic relationship, either increasing or decreasing, exists between $||\vec{\beta_j}||_2$ and λ. Based on this understanding, in this study, we assume that either a monotonic decreasing or increasing relationship exists between λ and $||\vec{\beta_j}||_2$. Furthermore, based on the above analysis, we use the threshold to estimate the importance of each group in predicting the target variable as a criterion for establishing their relative importance in this study. Notably, the key factor that distinguishes among different groups is the value of the regularization parameter λ at this threshold. Therefore, by leveraging the value of the regularization parameter λ at the threshold for each group, we can estimate the importance of each group, thereby improving the interpretability of group-specific data.

Step 1: Prepare multiple values of the regularization parameter λ to identify the threshold in group j that determines whether it is important or not for predicting the target variable.
Step 2: Define the optimization problem as outlined in Equation 2. Estimate the regression coefficient $\vec{\beta}$ using the proximal gradient algorithm for Group LASSO Regression (Algorithm 1).
Step 3: If the regression coefficient $\vec{\beta_j}$ estimated in Step 2 is not the zero vector, calculate the value of $||\vec{\beta_j}||_2 - \gamma\lambda$. The regularization parameter λ that is closest to calculating $||\vec{\beta_j}||_2 - \gamma\lambda \approx 0$ is defined as λ_j.
Step 4: Convert λ_j into a ratio and estimate group importance using the formula given in Equation (8).

Algorithm 2: Group Importance Estimation based on Group LASSO Regression.

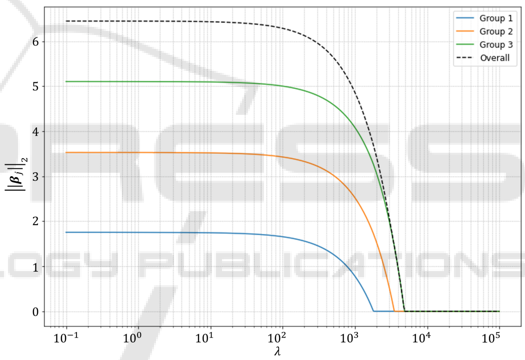

Figure 4: Relationship between the regularization parameter λ and group coefficient $||\vec{\beta_j}||_2$.

3.3 Formulation of the Group Importance Estimation

In the proposed method, we utilize the optimization problem of Group LASSO regression, represented by Equation (2). We consider the point where the group regression coefficient $\vec{\beta_j}$ first becomes zero as the group's threshold, as detailed in section 3.2.

To effectively implement our proposed method, it is essential to identify the threshold for the group by adjusting the value of the regularization parameter λ. However, pinpointing this threshold accurately is a complex task. Specifically, there's a need to search through the values of λ in fine-grained steps to determine the threshold for the group with precision. Such an approach might substantially increase computational time, potentially rendering it impractical for large datasets or real-time applications. Therefore, to address this challenge, instead of using the point at which the regression coefficient for a group, $\vec{\beta}_j$, first becomes zero as the threshold, we propose estimating the threshold near the point where $\vec{\beta}_j$ first becomes zero. Building on this, the essence of our approach can be captured from Equation (7). If $\gamma\lambda > ||\vec{\beta}_j||_2$, $\vec{\beta}_j$ remains non-zero. Conversely, if $\gamma\lambda < ||\vec{\beta}_j||_2$, $\vec{\beta}_j$ becomes zero. This implies that $\gamma\lambda = ||\vec{\beta}_j||_2$ serves as a crucial threshold in our framework. For our study, we adopt $\gamma\lambda \approx ||\vec{\beta}_j||_2$ as the threshold for the group. Further, to estimate this threshold, we compute multiple regression coefficients $\vec{\beta}$ for varied λ values. Our objective is to locate the threshold $||\vec{\beta}_j||_2 \approx \gamma\lambda$ where $\vec{\beta}$ remains non-zero. With γ being constant at this threshold, λ provides a measure of group importance. We define λ_j for each group j such that $||\vec{\beta}_j||_2 \approx \gamma\lambda$. This value, λ_j, is indicative of the importance of group j in our proposed method.

To facilitate a more comprehensive comparison among groups, we introduce an importance measure o_j, which quantifies the relative importance of group j. Formally, it is defined as:

$$o_j = \frac{\lambda_j}{\sum_{k=1}^{J} \lambda_k} \tag{8}$$

A thorough step-by-step explanation of this method is provided in Algorithm 2.

4 EXPERIMENTS

In this section, we conduct experiments using both generated data and real data to demonstrate the efficacy of the estimated group importance using our proposed method. Section 4.1 provides a detailed description of the generated and real-world datasets. Section 4.2 presents the common experimental conditions for both sets of experiments. Finally, section 4.3 shows the results from both the generated and real data experiments.

4.1 Experimental Data

4.1.1 Generated Data

We generated data using specific parameters. For the vector $\vec{v}[\theta,\eta]$, we define:

$$\sum_{n=1}^{400} \vec{v}[\theta,\eta]_n = \theta, \ \sum_{n=1}^{400} \vec{v}[\theta,\eta]_n{}^2 = \eta$$

From this, we derive the vector $\vec{x}_m$ as:

$$\vec{x}_m = \vec{v}[0,1] \quad (m = 1, \cdots, 15)$$

The target variable, $\vec{y}$, is then generated as:

$$\vec{y} = X\vec{\beta} + \vec{v}[0,2]$$

The regression coefficients $\vec{\beta}$ are detailed in Table 1. As indicated in Table 1, the coefficients for group 1, $\vec{\beta}_1$, are defined as $\beta_{\{1\}m_j} = 0.4$ for $m_j = 1,2,3$. By setting the elements of the regression coefficients for each group to the same value in this experiment, the values presented in the third row of Table 1 represent the importance of each group.

4.1.2 Real Data

The data utilized in this study is sourced from the open datasets made publicly available by the Ministry of Health, Labour and Welfare in Japan. Our experiments span data points collected from May 10, 2020, to May 8, 2023, totaling 1094 entries. The target variable (or the dependent variable) is the number of daily deaths in Japan due to the novel coronavirus (COVID-19) infection. The independent variables (or explanatory variables) represent the number of daily infections with the COVID-19, broken down by prefecture in Japan. Out of all the prefectures, 12 were selected for this study based on the criterion that they accounted for at least 2% of the total infections in Japan as of May 8, 2023. These prefectures are Hokkaido, Saitama, Chiba, Tokyo, Kanagawa, Shizuoka, Aichi, Kyoto, Osaka, Hyogo, Hiroshima, and Fukuoka. The rationale behind this selection is to ensure the model's appropriateness by avoiding prefectures with significantly low infection rates compared to the national total as of May 8, 2023.

In Japan, prefectures are often categorized into regions: Hokkaido, Tohoku, Kanto, Chubu, Kinki, Chugoku, Shikoku, Kyushu, and Okinawa. To define the group importance in our experiments, we used the total number of deaths due to COVID-19 as of May 8, 2023, in the selected 12 prefectures. We then aggregated these death counts according to the aforementioned regional groupings. The rationale behind using the total number of deaths in each region is to provide an estimate of the group's importance in each area. By utilizing the total death count, we can indicate the severity and impact of the pandemic in each region, which serves as an approximation of the group's importance. The defined group importance is presented in the fourth column of Table 2.

In this study, we conducted statistical tests using Ridge regression to assess whether the real data is suitable for the regression model. Specifically, we evaluated whether the model's residuals were homoscedastic by conducting the Breusch-Pagan test. The results showed a p-value of 0.078, indicating

Table 1: Regression coefficients and Group importance in generated data.

Coefficient	$\vec{\beta}_1$			$\vec{\beta}_2$			$\vec{\beta}_3$			$\vec{\beta}_4$			$\vec{\beta}_5$		
	$\beta_{\{1\}1}$	$\beta_{\{1\}2}$	$\beta_{\{1\}3}$	$\beta_{\{2\}1}$	$\beta_{\{2\}2}$	$\beta_{\{2\}3}$	$\beta_{\{3\}1}$	$\beta_{\{3\}2}$	$\beta_{\{3\}3}$	$\beta_{\{4\}1}$	$\beta_{\{4\}2}$	$\beta_{\{4\}3}$	$\beta_{\{5\}1}$	$\beta_{\{5\}2}$	$\beta_{\{5\}3}$
Importance	0.40			0.70			1.50			0.50			1.70		

Table 2: Group importance as defined in real data.

Region	**Prefecture**	**Total Deaths as of May 8, 2023**	**Regional Total Deaths as of May 8, 2023**
Hokkaido	Hokkaido	4610	4610
Kanto	Saitama	4013	20418
	Chiba	3944	
	Tokyo	8126	
	Kanagawa	4335	
Chubu	Shizuoka	1408	5771
	Aichi	4363	
Kinki	Kyoto	1674	14141
	Osaka	8559	
	Hyogo	3908	
Chugoku	Hiroshima	1373	1373
Kyushu	Fukuoka	3205	3205

insufficient evidence to reject the hypothesis of homoscedastic residuals. Furthermore, the high determination coefficient of 0.821 demonstrates that the model explains a significant portion of the data's variation. Based on these findings, we conclude that it is appropriate to apply the real data to the model in this study.

4.2 Experimental Setups

In this section, we elucidate the experimental conditions that are consistent across both the generated and real data experiments. We compare four methods: LASSO regression, Group LASSO regression, Ridge regression (Hoerl and Kennard, 1970), and our proposed method.

For both generated and real data, the explanatory variables undergo standardization to have a mean of 0 and a variance of 1. To avoid discussions about the intercept $\vec{\varepsilon}$ with regard to the target variable $\vec{y}$, we center the target variable in both datasets. In the proposed method, the regularization parameter λ ranges from 10^{-1} to 10^{4}, and is divided into 5000 equidistant values. On the other hand, for the LASSO regression, Group LASSO regression, and Ridge regression methods, the same range of λ values is used to determine the optimal λ. This determination is performed using Leave-one-out cross-validation, and the identified λ is then utilized to estimate the regression model using the entire dataset.

4.3 Experimental Results

In this section, we present the experimental results obtained from both generated and real data. Following the results, a discussion based on the results of each experiment is provided. All methods utilized in this study were executed using custom-made programs in Python. The computations for all methods were performed on a PC equipped with an Intel(R) Core(TM) i7-8700 CPU @3.20GHz.

4.3.1 Generated Data

This section presents the experimental results utilizing the generated data delineated in section 4.1.1. Table 3 lists the regression coefficients $\vec{\beta}$ estimated by LASSO regression, Group LASSO regression, and Ridge regression. As evident from Table 3, multiple regression coefficients are zero for both LASSO and Group LASSO regression. Thus, assessing the group importance through LASSO and Group LASSO is infeasible. As Ridge regression estimates values for all regression coefficients, for this experiment, we consider the sum of regression coefficients for each group as an indicator of the group importance. Table 4 lists the group importance as determined by our proposed method and Ridge regression, alongside the group importance described in section 4.1.1. The proposed method estimates the group importance in percentages, so all values of group importance in Table 4 are conveyed in percentages. From Table 4, it's discernible that the group importance estimated

Table 3: LASSO, Group LASSO, and Ridge regression coefficients for generated data.

Coefficient/Method	$\beta_{\{1\}1}$	$\beta_{\{1\}2}$	$\beta_{\{1\}3}$	$\beta_{\{2\}1}$	$\beta_{\{2\}2}$	$\beta_{\{2\}3}$	$\beta_{\{3\}1}$	$\beta_{\{3\}2}$	$\beta_{\{3\}3}$	$\beta_{\{4\}1}$	$\beta_{\{4\}2}$	$\beta_{\{4\}3}$	$\beta_{\{5\}1}$	$\beta_{\{5\}2}$	$\beta_{\{5\}3}$
LASSO	0.000	0.000	0.000	0.000	0.000	0.000	0.432	0.000	0.652	0.000	0.000	0.000	0.663	0.677	0.788
Group LASSO	0.000	0.000	0.000	0.000	0.000	0.000	0.805	0.499	0.936	0.000	0.000	0.000	1.050	1.082	1.132
Ridge	0.300	0.325	0.191	0.505	0.499	0.369	1.161	0.845	1.294	0.253	0.287	0.386	1.241	1.306	1.370

Table 4: Group importance as estimated by Proposed method and Ridge regression for generated data.

Group/Method	Group 1	Group 2	Group 3	Group 4	Group 5
Defined Importance (%)	8.33	14.58	31.25	10.42	35.42
Proposed Method (%)	8.59	13.96	30.97	9.58	36.90
Ridge Regression (%)	7.89	13.30	31.94	8.96	37.91

Table 5: Execution times between the Proposed method and Ridge regression for genereted data.

Method	Time (seconds)
Proposed Method	33
Ridge Regression	1040

Table 6: LASSO, Group LASSO, and Ridge regression coefficients for real data.

Coefficient/Method	$\beta_{Hokkaido}$	$\beta_{Saitama}$	β_{Chiba}	β_{Tokyo}	$\beta_{Kanagawa}$	$\beta_{Shizuoka}$	β_{Aichi}	β_{Kyoto}	β_{Osaka}	β_{Hyogo}	$\beta_{Hiroshima}$	$\beta_{Fukuoka}$
LASSO	0.000	0.000	0.000	0.000	0.000	4.620	0.000	0.000	0.000	0.000	1.891	0.000
Group LASSO	0.000	0.212	0.247	0.220	2.658	4.620	1.694	0.000	0.000	0.000	0.000	0.000
Ridge	0.087	0.378	0.821	-0.652	0.539	1.849	0.663	0.146	-0.153	0.682	1.745	0.227

Table 7: Group importance as estimated by Proposed method and Ridge regression for real data.

Group/Method	Hokkaido	Kanto	Chubu	Kinki	Chugoku	Kyushu
Defined Importance (%)	9.31	41.23	11.65	28.56	2.77	6.47
Proposed Method (%)	3.92	46.16	18.57	23.78	3.50	4.07
Ridge Regression (%)	1.10	30.10	31.63	12.35	21.97	2.86

Table 8: Execution times between the Proposed method and Ridge regression for real data.

Method	Time (seconds)
Proposed Method	1598
Ridge Regression	7682

by the proposed method more closely mirrors the defined group importance compared to that estimated by Ridge regression. Additionally, Table 5 documents the execution times for estimating the group importance by our proposed method and Ridge regression, where Ridge regression estimations are performed using the Leave-one-out cross-validation method. The results show that our proposed method is more time-efficient compared to Ridge regression, highlighting its practicality, especially in scenarios requiring quick model evaluations.

4.3.2 Real Data

This section presents the experimental results utilizing the real data delineated in section 4.1.2. Table 6 lists the regression coefficients $\vec{\beta}$ estimated by LASSO regression, Group LASSO regression, and Ridge regression. Analogous to the results from the generated data, it's apparent from Table 6 that multiple regression coefficients are zero for both LASSO and Group LASSO regression. Hence, assessing the group importance through LASSO and Group LASSO is infeasible. As with the generated data, since Ridge regression estimates values for all regression coefficients, we consider the summation of regression coefficients for each group as a metric for the group importance. Table 7 lists the group importance as determined by our proposed method and Ridge regression, alongside the group importance described in section 4.1.2. Similarly, since our proposed method estimates the group importance in percentages, all values in Table 7 are conveyed in percentages. It's discernible from Table 7 that the group importance estimated by our proposed method more closely mirrors the defined group importance compared to that estimated by Ridge regression. Additionally, Table 8 documents the execution times for estimating the group importance by our proposed method and Ridge regression, where Ridge regression estimations are performed using the Leave-one-out cross-validation method. The results show that our proposed method is more time-efficient compared to Ridge regression, highlighting its practicality, especially in scenarios requiring quick model evaluations.

5 CONCLUSION

In this study, we introduced a method for estimating the group importance based on Group LASSO regression. This method addresses the primary limitation of Group LASSO regression, which is its focus on sparsity only at the group level and often neglecting the assessment of group importance. Our experiments with both generated and real data showed that our method consistently demonstrates values closer to the defined group importance compared to existing methods, highlighting the efficacy of our approach. However, our method does have certain limitations, which are important to consider:

- **Necessity to Predefine Multiple Regularization Parameters.**
  The method requires a careful selection and predefinition of a range of regularization parameter values, which can be time-consuming and challenging, especially for datasets with varying characteristics and complexities.
- **Possibility of Extended Execution Times.**
  For large datasets or complex models, our method might need more computation per regularization parameter, potentially increasing execution time. However, tests with generated and real data suggest it's generally more time-efficient than Ridge regression's Leave-one-out cross-validation. This efficiency isn't always consistent across different scenarios. Considering execution time is crucial when comparing models and datasets.

Moving forward, it would be beneficial to further validate the accuracy of group importance estimated by our method. This necessitates conducting experiments with a broader range of generated and real datasets to establish its credibility and robustness more comprehensively.

REFERENCES

Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67.

Meier, L., Van De Geer, S., and Bühlmann, P. (2008). The group lasso for logistic regression. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 70(1):53–71.

Nardi, Y. and Rinaldo, A. (2008). On the asymptotic properties of the group lasso estimator for linear models.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1):267–288.

Tomioka, R. and Scientific, K. (2015). *Machine Learning with Sparsity Inducing Regularizations*. MLP Machine Learning Professional Series. Kodansha.

Yang, H., Xu, Z., King, I., and Lyu, M. R. (2010). Online learning for group lasso. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pages 1191–1198.

Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 68(1):49–67.