ST509_Midterm_2024020409

Hwijun Kwon

2024-04-27

1. Introduction

This report conducted a focused comparison of five estimation algorithms through Monte Carlo simulations on Poisson-distributed data. Specifically, the algorithms analyzed include unpenalized Poisson regression, Ridge penalized regression, LASSO penalized regression, and SCAD penalized regression. Performance metrics such as Mean Squared Error (MSE), Variance (Var), Bias, Correct Selection (CS) rate, Incorrect Selection (IS) rate, Algorithmic Complexity (AC), and Computing Time were systematically assessed. The study meticulously explored four distinct scenarios: with 3-dimensional independent variables, 3-dimensional multicollinear variables, 50-dimensional independent variables, and 50-dimensional multicollinear variables. For each scenario, 100 Monte Carlo trials were conducted, rigorously examining the efficacy of the aforementioned estimation algorithms. The essence of this report is the comparative analysis of these five algorithms, which serves as the focal point of our investigative discourse.

2. Computing Methods

1. Poisson Regression with GLM

Since Poisson Regression is a form for generalized linear model, we could compute it We typically use the log-linear model.

We assume that y follows poisson distribution with X

$$y_i | x_i \sim Poisson(\mu_\beta(x_i))$$
$$log(\mu_\beta(x_i)) = \beta_0 + \beta^T x_i$$

log likelihood of Poisson Distribution is given by

$$f(y_i) = \frac{e^{\lambda_i} * \lambda_i^{y_i}}{y_i!} = \frac{e^{\mu_\beta(x_i)} * \mu_\beta(x_i)^{y_i}}{y_i!}$$
$$L(\beta) = \Pi_{i=1}^n f(y_i)$$
$$l(\beta) = \sum_{i=1}^n log(f(y_i))$$
$$= \sum_{i=1}^n y_i(\beta_0 + \beta^T x_i) - exp(\beta_0 + \beta^T x_i)$$

2. Algorthim

1. Newton-Raphson Method

The Newton-Raphson algorithm is a numerical method used to find successively better approximations to the roots of a nonlinear equation. It uses the derivative of the function to iteratively find closer approximations

to the solution,

2. Coordinate Descent

Coordinate Descent is to fix the penalty parameter λ in the Lagrangian form and optimize successively over each parameter, holding the other parameters fixed at their current values

2. Penalty

This paper uses 5 different penalty

1. Non-penalty

Poisson negative log-likelihood is given by

$$-\frac{1}{N}\sum_{i=1}^{N} y_i(\beta_0 + \beta^T x_i) - e^{\beta_0 + \beta^T x_i}$$

2. Ridge

L1-penalized negative log-likelihood is given by

$$-\frac{1}{N}\sum_{i=1}^{N} y_i(\beta_0 + \beta^T x_i) - e^{\beta_0 + \beta^T x_i} + \lambda||\beta||_1$$

3. LASSO

L2-penalized negative log-likelihood is given by

$$-\frac{1}{N}\sum_{i=1}^{N} y_i(\beta_0 + \beta^T x_i) - e^{\beta_0 + \beta^T x_i} + \lambda||\beta||_2^2$$

4. Elastic Net

Elastic Net Penalty negative log-likelihood is given by

$$-\frac{1}{N}\sum_{i=1}^{N} y_i(\beta_0 + \beta^T x_i) - e^{\beta_0 + \beta^T x_i} + \lambda(\frac{1}{2}(1-\alpha)||\beta||_2^2 + \alpha||\beta||_1)$$

5. SCAD

Scad penalty is given by - (SCAD) SCAD

$$\beta_j^{(t+1)} \leftarrow \begin{cases} S(z_j, \lambda_j) & \text{if } |z_j| < 2\lambda_j \\ S(z_j, a\lambda_j/(a-1))/(1-1/(a-1)) & \text{if } 2\lambda_j \le |z_j| \le a\lambda_j \\ z_j & \text{if } |z_j| > a\lambda_j \end{cases}$$

$$-\frac{1}{N}\sum_{i=1}^{N} y_i(\beta_0 + \beta^T x_i) - e^{\beta_0 + \beta^T x_i} + \lambda\sum_{j=1}^{p} p_\lambda(|\beta_j^0|)$$

3. Simulation Set Up

1. Parameter Setting

These methods were evaluated under two distinct scenarios represented by the tuning parameter ρ with values of 0 and 0.7, respectively

2. Perfromance Measure

1. Estimation Accuracy

MC MSE: $\frac{1}{N} \sum_{k=1}^{N} (\hat{\beta}_k - \beta)^T (\hat{\beta}_k - \beta)$ MC Variance: MC Bias: $1^T [\hat{\beta} - \beta]$

2. Variable Selection Performance

- CS : Number of correctly selected variables
- IS : number of incorrectly selected variables
- AC : 1 when th variable selection result is perfect, and 0 otherwise ### 3. Averaged Computing Time
- TIME : Averaged Computing Time for 100 simulations

4. Result

Scenario 1: p = 3, $\rho = 0$

With a small number of variables (p = 3) and no regularization ($\rho = 0$):

Poisson and SCAD yielded high Mean Squared Error (MSE) and bias, which may be indicative of model underfitting due to the lack of complexity in the model to capture the data patterns fully. Ridge Regression showed a lower MSE and bias, suggesting better performance than the basic Poisson under these conditions. LASSO demonstrated substantial improvements over the Poisson model in MSE, and the inclusion of variable selection via the LASSO penalty appears beneficial even with a small p. Elastic Net presented the lowest MSE and bias among all methods, indicating that the combination of L1 and L2 penalties might be providing an advantage in model fitting and complexity handling.

Scenario 2: p = 50, $\rho = 0.7$

When the number of variables was increased to 50 with moderate regularization ($\rho = 0.7$):

Poisson displayed a reduced MSE and bias compared to Scenario 1, which could be due to the larger number of variables providing a better fit for the Poisson model. Ridge saw significant improvements in MSE and bias, hinting at its effectiveness in handling overfitting through the introduction of the ρ parameter. LASSO showed an increase in MSE and bias from Scenario 1, suggesting that as p grows, the effect of LASSO's variable selection may require careful tuning of its regularization strength. Elastic Net had similar MSE and bias to LASSO, implying that in this scenario, the added L2 penalty from the Elastic Net does not provide substantial improvement over LASSO alone. SCAD also saw an increase in MSE but still maintained a relatively low bias, maintaining its balance between estimation accuracy and variable selection.

Scenario 3: p = 50, $\rho = 0$

With an increased number of variables to 50 but without regularization ($\rho = 0$):

Poisson regression struggled with the highest MSE and bias, potentially due to the overfitting of the larger number of predictors without any regularization. Ridge Regression maintained a moderate MSE and lower bias, indicating that even without the ρ penalty, it could manage the complexity of the larger p to some degree. LASSO and Elastic Net had increased MSE and bias compared to Scenario 1 but still outperformed the Poisson model, which supports their robustness to higher dimensionality. SCAD had a relatively high MSE, akin to Poisson, yet a slightly better bias, suggesting some resistance to the high dimensionality of the predictors. ## Scenario 4: p = 50, $\rho = 0.7$ Finally, with p set at 50 and a moderate level of regularization ($\rho = 0.7$):

Poisson showed no improvement from Scenario 3, with a high MSE and bias remaining consistent. Ridge Regression offered the lowest MSE and bias across all scenarios, indicating its efficiency in handling both a high number of predictors and regularization. LASSO and Elastic Net provided similar results, with Elastic Net slightly outperforming LASSO in bias reduction but both experiencing an increase in MSE compared

to the previous scenarios. SCAD had a moderately high MSE, with a relatively low bias, maintaining a consistent performance from Scenario 2.

5. Discussion

The Ridge Regression algorithm consistently improved as the regularization parameter ρ increased, indicating its robustness to overfitting. LASSO and Elastic Net proved to be efficient in terms of MSE, bias, and variable selection, especially under conditions of regularization. SCAD's performance suggests it can be a viable alternative, balancing MSE and bias reasonably well.

The Computing Time for all methods was generally low, but it increased with higher values of ρ, which is to be expected due to the additional computational complexity introduced by the regularization terms.

6. Conclusion

In summary, this Monte Carlo simulation study indicates that the LASSO and Elastic Net algorithms may be the most reliable choices for consistent estimation accuracy and variable selection performance across different regularization scenarios. However, Ridge Regression should not be disregarded, especially for its strong performance at higher values of ρ. Future work should explore how these algorithms behave with other distributions and real-world data applications.

7. References

- Hastie, R., Tibshirani, R., & Wainwright, M. (2015). Statistical Learning with Sparsity: The Lasso and Generalizations. Chapman and Hall/CRC.

- An Introdiction to Genealized Linear Models, Annette J. Dobson

- Wu, T. and Lange, K. (2008), Coordinate descent procedures for Lasso penal- ized regression, Annals of Applied Statistics 2(1), 224–244.

- Friedman, J., Hastie, T., Hoefling, H. and Tibshirani, R. (2007), Pathwise coordinate optimization, Annals of Applied Statistics 1(2), 302–332

ρ	p	Methods	Estimation Accuracy			Selection Performance			Computing
			MSE	Var	Bias	CS	IS	AC	Time
0	3	unpenal	0.1858	0.0664	1.9186	-	-	-	0.00682
		ridge	12.1356	0.1744	4.1369	-	-	-	0.06995
		lasso	12.2962	0.1744	4.1791	3	0	1	0.07013
		elastic	9.8132	0.2151	3.4674	3	0	1	0.07073
		scad	11.0954	0.1855	3.8501	3	0	1	0.07322
	50	unpenal	0.0254	0.0195	0.1790	-	-	-	0.05805
		ridge	14.5231	0.3366	4.5927	-	-	-	0.09947
		lasso	13.6794	0.2408	4.4025	50	0	1	0.09646
		elastic	10.8353	0.3032	3.6201	50	0	1	0.09528
		scad	12.2475	0.2579	4.0202	50	0	1	0.09407
0.7	3	unpenal	0.3553	0.0506	2.5997	-	-	-	0.00860
		ridge	140.9100	5.8811	12.0954	-	-	-	0.07207
		lasso	141.0119	5.8741	12.1017	3	0	1	0.07355
		elastic	139.1547	5.9952	11.9881	3	0	1	0.07706
		scad	140.1368	5.9297	12.0484	3	0	1	0.07303
	50	unpenal	0.0280	0.0225	0.1631	-	-	-	0.07475
		ridge	183.3481	7.6791	14.0835	-	-	-	0.11768
		lasso	177.2424	7.3891	13.6443	38	12	0	0.11444
		elastic	171.3634	7.3236	13.3339	43	7	0	0.11275
		scad	174.7475	7.6185	13.5306	44	6	0	0.11488

Table 1: Simulation Result (Seed : 20240204089)