

Employing particle swarm optimization algorithm for shrinkage parameter estimation in generalized Liu estimator

Qamar Abdulkareem Abdulazeez¹, Zakariya Yahya Algamal^{1*}

¹ Department of Statistics and Informatics, University of Mosul, Mosul, Iraq

*Corresponding author E-mail: zakariya.algamal@uomosul.edu.iq

Abstract

It is well-known that in the presence of multicollinearity, the Liu estimator is an alternative to the ordinary least square (OLS) estimator and the ridge estimator. Generalized Liu estimator (GLE) is a generalization of the Liu estimator. However, the efficiency of GLE depends on appropriately choosing the shrinkage parameter matrix which is involved in the GLE. In this paper, a particle swarm optimization method, which is a metaheuristic continuous algorithm, is proposed to estimate the shrinkage parameter matrix. The simulation study and real application results show the superior performance of the proposed method in terms of prediction error.

Keywords: Multicollinearity; Shrinkage Parameter; Generalized Liu Estimator; Particle Swarm Optimization.

1. Introduction

Regression modeling is a widely applied strategy for studying several real data problems. In linear regression model, the response variable is considered as a continuous and reasonably assumed to follow normal distribution. In linear regression models, it is assumed that the correlations among the explanatory variables are not high (Alheety & Kibria, 2014; Alkhamisi & Shukur, 2007; Dorugade, 2014). However, this assumption is not always hold in practice. In linear regression model, the ordinary least squares (OLS) estimator is the best estimator among all linear and unbiased estimators. However, under multicollinearity, OLS estimator becomes unhelpful due to their large variance.

The ridge estimator (RE) (Hoerl & Kennard, 1970) has been consistently demonstrated to be an attractive and alternative to the OLS, when the multicollinearity exists. RE can shrink all the regression coefficients toward zero to reduce the large variance (Asar & Genç, 2015).

Liu (1993) suggested a Liu estimator (LE) for the linear regression model and found out that the drawback in the ridge estimator is that it provides a complicated non-linear function of the biasing parameter k which has the range 0 to ∞ . So, the LE is considered as an alternative to the OLS estimator to overcome the problem of multicollinearity. The advantage of the LE is that it is a linear function of the Liu parameter, d , which has a range of $0-1$.

Therefore, many researchers have used the LE instead of the RE due to the linear function of the d (Qasim et al., 2018).

Generalized Liu estimator (GLE) has also been considered as a generalization of the LE. The performance of the GLE is fully depending on the values of the shrinkage parameter matrix. Accordingly, appropriate choosing of the shrinkage parameter matrix is an important part of applying GLE.

In recent years, numerous natural-inspired algorithms have been successfully introduced and applied as random search strategies for solving a number of optimization problems. Particle swarm optimization algorithm is a comparatively recent population-based algorithm that is inspired by swarm.

In this paper, the Particle swarm optimization algorithm is proposed to estimate the values of the shrinkage parameter matrix in GLE. Our proposed approach will efficiently help to find the best values with high prediction accuracy. The superiority of our proposed approach in different simulated examples and a real data application is proved.

2. Generalized Liu estimator

Suppose that we have a data set $\{(y_i, \mathbf{x}_i)\}_{i=1}^n$ where $y_i \in \mathbb{R}$ is a response variable and $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip}) \in \mathbb{R}^p$ represents a p -dimensional explanatory variable vector. Without loss of generality, it is assumed that the response variable is centered and the explanatory variables are standardized.

Consider the following linear regression model,

$$y = X\beta + \varepsilon, \quad (1)$$

Where y is an $n \times 1$ vector of observations of the response variable, $X = (x_1, \dots, x_p)$ is an $n \times p$ known design matrix of explanatory variables, $\beta = (\beta_1, \dots, \beta_p)$ is a $p \times 1$ vector of unknown regression coefficients, and ε is an $n \times 1$ vector of random errors with mean 0 and variance σ^2 . Using OLS method, the parameter estimation of Eq. (1) is given by

$$\hat{\beta}_{OLS} = (X^T X)^{-1} X^T y. \quad (2)$$

OLS estimator is unbiased and it has minimum variance among all linear unbiased estimators. However, in the presence of multicollinearity, the $X^T X$ matrix is nearly singular that makes OLS estimator unstable due to their large variance. To reduce the effects of the multicollinearity, RE (Hoerl & Kennard, 1970), which is the most commonly used method, adds a positive shrinkage parameter, k , to the main diagonal of the $X^T X$ matrix. The RE is defined as

$$\hat{\beta}_{RE} = (X^T X + kI)^{-1} X^T y, \quad (3)$$

Where I is the identity matrix with dimension $p \times p$. The estimator $\hat{\beta}_{RE}$ is biased but more stable and has less mean square error. The shrinkage parameter, k , controls the shrinkage of β toward zero. The OLS estimator can be considered as a special estimator from the RE with $k = 0$. For larger value of k , the RE yields greater shrinkage approaching zero (Yang & Emura, 2017).

The LE which is proposed by Liu (1993) for linear regression model is an alternative estimator to overcome multicollinearity. It is defined as

$$\hat{\beta}_{LE} = (X^T X + I)^{-1} (X^T X + dI) \hat{\beta}_{OLS}, \quad (4)$$

Where $0 < d < 1$ is a constant parameter known as Liu-biasing parameter. Rewriting Eq. (1) as (Alheety & Kibria, 2014)

$$y = Z\alpha + \varepsilon \quad (5)$$

Where $Z = XW$, where W is a matrix $p \times p$ so that $ZZ = WX^T XW$ will implies $ZZ = \Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$ where Λ is a diagonal matrix with the Eigen values of $X^T X$ and $\alpha = W^T \beta$, then OLS estimator of α is given by (Alheety & Kibria, 2009):

$$\hat{\alpha}_{OLS} = \Lambda^{-1} Z^T y \quad (6)$$

And

$$\hat{\alpha}_{LE} = (\Lambda + I)^{-1} (\Lambda + dI) \hat{\alpha}_{OLS} \quad (7)$$

The mean square error (MSE) is

$$MSE(\hat{\beta}_{LE}) = \hat{\sigma}^2 \sum_{i=1}^p \frac{(\lambda_i + d)^2}{\lambda_i (\lambda_i + 1)^2} + (d-1)^2 \sum_{i=1}^p \frac{\alpha_i^2}{(\lambda_i + 1)^2} \quad (8)$$

The difference between LE and GLE is there are ith values of d , such that (Hoerl & Kennard, 1970)

$$\hat{\beta}_{GLE} = (X^T X + I)^{-1} (X^T X + D) \hat{\beta}_{OLS}, \quad (9)$$

Where $D = \text{diag}(d_1, d_2, \dots, d_p)$. Since the Liu parameter is the key to reduce the multicollinearity, there are multi ways to determine this value, the researcher suggest several ways to choose the optimal d such as (Hocking et al., 1976), (Nomura, 1988), (Troskie & Chalton, 1996), (Firinguetti, 1999), (Alkhamisi & Shukur, 2007), (Batah et al., 2008), (Al-Hassan, 2010), (Dorugade & Kashid, 2010), (Månsson et al., 2010), (Dorugade, 2014), (Asar et al., 2014) and (Bhat & Raju, 2017). In this paper, we adapted the following method which is defined as

$$D = \frac{\alpha_i^2 - 1}{\lambda_i + \alpha_i^2} \quad (10)$$

3. The proposed methods

The efficiency of Liu estimator strongly depends on appropriately choosing the shrinkage parameter. A choice of shrinkage parameter that is too small leads to overfitting the GLE, while shrinkage parameter that is too large shrinks β by too much, making a bias-variance tradeoff.

Particle swarm optimization (PSO) is a nature-inspired metaheuristic algorithm that was originally proposed by Kennedy and Eberhart (1995) for solving continuous optimization problems.

PSO inspires the social or collective behavior of animals such as bird flocking and fish schooling. PSO compares with the other computation intelligence-based algorithms has several advantages, such as simple implementation, computationally higher efficiency, fewer parameters to tune, scalability and flexibility, robustness. For instance, comparing with genetic algorithm, there is no crossover and mutation genetic operation (Chen et al., 2014; Kiran, 2017; Lin et al., 2008; Lu et al., 2009; Zhou & Dickerson, 2014).

PSO performs the searching using a population, which is called swarm, of particles. Each particle has three features: (1) position, (2) velocity, and (3) fitness value. In PSO, each particle can be represented as a candidate solution (position) in the search space. The particles fly through the search space by their own efforts and in cooperation with other particles and they follow the best solutions they have achieved (local best solutions), as well as tracking the best solutions that they found (the best global solution) (Cervantes et al., 2017; Lai et al., 2016; Mirjalili & Lewis, 2013; Wen et al., 2011).

Mathematically, the search space is assumed to be Q -dimensional and there are m of particles in the swarm where $q = 1, 2, \dots, Q$. During the movement, each particle has a position vector $\mathbf{x}_i = \{x_{i1}, x_{i2}, \dots, x_{iq}\}$ with a velocity vector $\mathbf{v}_i = \{v_{i1}, v_{i2}, \dots, v_{iq}\}$. In the PSO algorithm, the best position, which gives the best fitness value for the particle i , is called best previous position denoted as $Pbest_i = \{Pbest_{i1}, Pbest_{i2}, \dots, Pbest_{iq}\}$. The best position found by all particles is called the global best, which is denoted as $Gbest_i = \{Gbest_{i1}, Gbest_{i2}, \dots, Gbest_{iq}\}$. In each iteration, the PSO algorithm searches for the optimal solution by updating the position and the velocity of the i^{th} particle according to the following two equations:

$$v_{iq}^{t+1} = z \times v_{iq}^t + c_1 \times r_1 \times (Pbest_{iq}^t - x_{iq}^t) + c_2 \times r_2 \times (Gbest_{iq}^t - x_{iq}^t), \quad (11)$$

$$x_{iq}^{t+1} = x_{iq}^t + v_{iq}^{t+1}, \quad (12)$$

Where t denotes the iteration in the algorithm, z is the inertia weight which is used to balance between the global search and the local search. In addition, C_1 (the cognition learning factor) and C_2 (social learning factor) are the acceleration coefficients. While, r_1 and r_2 are random values selected from a uniform distribution with (0,1).

In this paper, a PSO algorithm is proposed to determine the shrinkage parameter matrix. The proposed method will efficiently help to reduce the MSE. The parameter configurations for our proposed method are presented as follows.

- 1) The number of particles, m , is set to 50 and the number of iterations is $t_{max} = 100$. The acceleration coefficients c_1 and c_2 are set within the range [2, 4]. The c_1 and c_2 are updating during the iteration according to the following equations:

$$c_1 = c_{1,\min} + \frac{t}{t_{\max}} (c_{1,\max} - c_{1,\min}), \quad (13)$$

$$c_2 = c_{2,\min} + \frac{t}{t_{\max}} (c_{2,\max} - c_{2,\min}), \quad (14)$$

Further, the minimum and the maximum values for the inertial weight are: $z_{\min} = 0.2$ and $z_{\max} = 0.9$. The inertial weight is updating according to the following equation:

$$z = z_{\max} - \frac{t}{t_{\max}} (z_{\max} - z_{\min}). \quad (15)$$

- 2) The positions of each particle are randomly determined. The position of a particle represents the shrinkage parameters, k_i . Here the dimension of each particle is the number of explanatory variables. The initial positions of the particles are generated from a uniform distribution within the range [0, 1].
- 3) The initial velocities of each particle are generated from a uniform distribution within the range [0, 4].
- 4) The fitness function is the MSE
- 5) The velocities and positions are updated using Eq. (11) and Eq. (12), respectively.
- 6) Steps 4 and 5 are repeated until a t_{max} is reached.

4. Monte Carlo simulation results

In this section, a comprehensive simulation study was conducted to evaluate the performance of the proposed method. Following McDonald and Galarneau (1975), the explanatory variables with different degree of multicollinearity are generated by

$$x_{ij} = (1 - \rho^2)^{1/2} w_{ij} + \rho w_{ip}, \quad i = 1, 2, \dots, n, \quad p = 1, 2, \dots, p, \quad (27)$$

Where ρ^2 represents the correlation between the explanatory variables and w_{ij} 's are independent standard normal pseudo-random numbers. The response variable is generated by

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \varepsilon_i, \quad (28)$$

Where ε_i is independent and identically normal distributed pseudo-random numbers with zero mean and variance σ^2 . Because the sample size has direct impact on the prediction accuracy, three representative values of the sample size are considered: 30, 50 and 150. In addition, the number of the explanatory variables are considered as $p \in \{4, 8, 12\}$. Further, because we are interested in the effect of multicollinearity, in which the degrees of correlation considered more important, three values of the pairwise correlation are considered with $\rho = \{0.90, 0.95, 0.99\}$. Besides, the values of σ^2 is 1.

For a combination of these different values of n, p, ρ , the generated data is repeated 5000 times and the averaged mean squared errors (MSE) is calculated as

$$MSE(\hat{\beta}) = \frac{1}{5000} \sum_{i=1}^{5000} (\hat{\beta} - \beta)^T (\hat{\beta} - \beta), \tag{29}$$

Where $\hat{\beta}$ is the obtained Liu estimator by the methods. The MSE values from the Monte Carlo simulation study are reported in Tables 1 – 3. Several observations can be obtained as follows:

- 1) The simulation results indicate that the PSO method of selecting D is superior to the other used selection methods for all combinations of n, p , and ρ in terms of MSE. We can see that PSO method has smaller MSE and significantly lower MSE than others.
- 2) It is seen from Tables 1 – 3 that $\hat{\beta}_{PSO}$ estimator using PSO method is usually more efficient than the OLS estimator for all values of n, p and when multicollinearity is high or severe.
- 3) In terms of ρ values, there is increasing in the MSE values when the correlation degree increases regardless the value of n and p .
- 4) Regarding the number of explanatory variables, it is easily seen that there is a negative impact MSE, where there are increasing in their values when the p increasing from four explanatory variables to twelve explanatory variables.
- 5) With respect to the value of n , the MSE values decrease when n increases, regardless the value of ρ and p .
- 6) All the selection methods of D are superior to the OLS estimator in terms of MSE.

Table 1: Average MSE when n=30

		OLS	PSO	D
p=4	r=0.90	1.829	0.823	1.231
	r=0.95	1.908	0.906	1.119
	r=0.99	1.888	0.876	1.303
p=8	r=0.90	2.702	0.695	1.712
	r=0.95	2.722	0.713	1.739
	r=0.99	2.751	0.745	1.820
p=12	r=0.90	3.593	1.477	1.606
	r=0.95	3.635	1.597	2.636
	r=0.99	3.615	1.613	2.721

Table 2: Average MSE when n=50

		OLS	PSO	D
p=4	r=0.90	1.926	0.924	1.227
	r=0.95	1.934	0.932	1.238
	r=0.99	1.918	0.917	1.829
p=8	r=0.90	2.826	0.824	1.828
	r=0.95	2.835	0.831	1.840
	r=0.99	2.819	0.817	1.851
p=12	r=0.90	3.765	1.759	2.767
	r=0.95	3.767	1.758	2.772
	r=0.99	3.755	1.735	2.801

Table 3: Average MSE when n=150

		OLS	PSO	D
p=4	r=0.90	1.942	0.942	1.342
	r=0.95	1.982	0.981	1.382
	r=0.99	1.967	0.967	1.569
p=8	r=0.90	2.948	0.947	1.947
	r=0.95	2.939	0.938	1.939
	r=0.99	2.946	0.946	1.950
p=12	r=0.90	3.922	1.921	2.921
	r=0.95	3.925	1.924	2.925
	r=0.99	3.913	1.913	2.918

5. Real application results

To evaluate the predictive performance of the proposed method and to compare its performance with the other used methods in a real data application, the Portland cement dataset is employed. Portland cement dataset became a standard dataset to examine and to remedy the multicollinearity. It was widely used by numerous researchers. This dataset comes from an experimental investigation of heat evolved during the setting and hardening of Portland cements of varied composition and the dependence of this heat on the percentages of four compounds in the clinkers from which the cement was produced. There are 13 observations of heat evolved in calories per gram of cement (y), tricalcium aluminate (x_1), tetracalcium silicate (x_2), tetracalcium aluminoferrite (x_3), and dicalcium silicate (x_4).

Before fitting the linear regression model, the explanatory variables and the response variable are standardized. Then, eigenvalues of XX matrix are calculated with $\lambda_1 = 26.8284$, $\lambda_2 = 18.9127$, $\lambda_3 = 2.2392$, and $\lambda_4 = 0.0194$ resulting in a condition number $\sqrt{\lambda_1 / \lambda_2} = 1376.8810$. Thus, the multicollinearity is existed. As a result, using RE and GRE will be more suitable than the OLS. The predictive performance for each used method is computed using the MSE ($MSE = (1/n) \sum_{i=1}^n (y_i - \hat{y}_i)^2$) and the results are given in Table 4.

It is apparent from Table 4 that there is an improvement of the predictive capability of the PSO comparing with the other used methods, where PSO significantly reduced the MSE. The reduction of MSE using PSO was 14.282% and 11.916% compared with OLS and D, respectively.

Table 4: Real Application Results for the Used Methods

Method	MSE
OLS	9303.049
PSO	8117.127
D	9215.278

6. Conclusion

In this paper, a new shrinkage parameter selection of the generalized Liu estimator, which is depending on employing the particle swarm optimization algorithm, was proposed. This proposed method allows us to handle multicollinearity with decreasing the variability of shrinkage parameter selection. Simulation and results demonstrate that the proposed method is outperformed several classical methods.

References

- [1] Al-Hassan, Y. M. (2010). Performance of a new ridge regression estimator. *Journal of the Association of Arab Universities for Basic and Applied Sciences*, 9(1), 23-26. <https://doi.org/10.1016/j.jaubas.2010.12.006>.
- [2] Alheety, M., & Kibria, B. G. (2009). On the Liu and almost unbiased Liu estimators in the presence of multicollinearity with heteroscedastic or correlated errors. *Surveys in Mathematics and its Applications*, 4, 155-167.
- [3] Alheety, M., & Kibria, B. G. (2014). A generalized stochastic restricted ridge regression estimator. *Communications in Statistics-Theory and Methods*, 43(20), 4415-4427. <https://doi.org/10.1080/03610926.2012.724506>.
- [4] Alkhamisi, M. A., & Shukur, G. (2007). A Monte Carlo study of recent ridge parameters. *Communications in Statistics—Simulation and Computation*, 36(3), 535-547. <https://doi.org/10.1080/03610910701208619>.
- [5] Asar, Y., & Genç, A. (2015). New shrinkage parameters for the Liu-type logistic estimators. *Communications in Statistics - Simulation and Computation*, 45(3), 1094-1103. <https://doi.org/10.1080/03610918.2014.995815>.
- [6] Asar, Y., Karabrahimoğlu, A., & Genç, A. (2014). Modified ridge regression parameters: A comparative Monte Carlo study. *Hacettepe Journal of Mathematics and Statistics*, 43(5), 827-841.
- [7] Batah, F. S. M., Ramanathan, T. V., & Gore, S. D. (2008). The efficiency of modified jackknife and ridge type regression estimators: a comparison. *Surveys in Mathematics & its Applications*, 3.
- [8] Bhat, S., & Raju, V. (2017). A class of generalized ridge estimators. *Communications in Statistics-Simulation and Computation*, 46(7), 5105-5112. <https://doi.org/10.1080/03610918.2016.1144765>.
- [9] Cervantes, J., Garcia-Lamont, F., Rodriguez, L., López, A., Castilla, J. R., & Trueba, A. (2017). PSO-based method for SVM classification on skewed data sets. *Neurocomputing*, 228, 187-197. <https://doi.org/10.1016/j.neucom.2016.10.041>.
- [10] Chen, K.-H., Wang, K.-J., Wang, K.-M., & Angelia, M.-A. (2014). Applying particle swarm optimization-based decision tree classifier for cancer classification on gene expression data. *Applied Soft Computing*, 24(0), 773-780. <https://doi.org/10.1016/j.asoc.2014.08.032>.
- [11] Dorugade, A. (2014). New ridge parameters for ridge regression. *Journal of the Association of Arab Universities for Basic and Applied Sciences*, 15(1), 94-99. <https://doi.org/10.1016/j.jaubas.2013.03.005>.
- [12] Dorugade, A., & Kashid, D. (2010). Alternative method for choosing ridge parameter for regression. *Applied Mathematical Sciences*, 4(9), 447-456.
- [13] Firinguetti, L. (1999). A generalized ridge regression estimator and its finite sample properties: A generalized ridge regression estimator. *Communications in Statistics-Theory and Methods*, 28(5), 1217-1229. <https://doi.org/10.1080/03610929908832353>.
- [14] Hocking, R. R., Speed, F., & Lynn, M. (1976). A class of biased estimators in linear regression. *Technometrics*, 18(4), 425-437. <https://doi.org/10.1080/00401706.1976.10489474>.
- [15] Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55-67. <https://doi.org/10.1080/00401706.1970.10488634>.
- [16] Kennedy, J., & Eberhart, R. C. (1995). Particle swarm optimization. *Proceedings of IEEE Conference on Neural Network*, 4, 1942-1948. <https://doi.org/10.1109/ICNN.1995.488968>.
- [17] Kiran, M. S. (2017). Particle swarm optimization with a new update mechanism. *Applied Soft Computing*, 60, 670-678. <https://doi.org/10.1016/j.asoc.2017.07.050>.
- [18] Lai, C.-M., Yeh, W.-C., & Chang, C.-Y. (2016). Gene selection using information gain and improved simplified swarm optimization. *Neurocomputing*, 218, 331-338. <https://doi.org/10.1016/j.neucom.2016.08.089>.
- [19] Lin, S.-W., Ying, K.-C., Chen, S.-C., & Lee, Z.-J. (2008). Particle swarm optimization for parameter determination and feature selection of support vector machines. *Expert Systems with Applications*, 35(4), 1817-1824. <https://doi.org/10.1016/j.eswa.2007.08.088>.
- [20] Liu, K. (1993). A new class of biased estimate in linear regression. *Communications in Statistics-Theory and Methods*, 22(2), 393-402. <https://doi.org/10.1080/03610929308831027>.
- [21] Lu, Y., Wang, S., Li, S., & Zhou, C. (2009). Particle swarm optimizer for variable weighting in clustering high-dimensional data. *Machine Learning*, 82(1), 43-70. <https://doi.org/10.1007/s10994-009-5154-2>.
- [22] Månsson, K., Shukur, G., & Golam Kibria, B. (2010). A simulation study of some ridge regression estimators under different distributional assumptions. *Communications in Statistics-Simulation and Computation*, 39(8), 1639-1670. <https://doi.org/10.1080/03610918.2010.508862>.
- [23] McDonald, G. C., & Galarneau, D. I. (1975). A Monte Carlo evaluation of some ridge-type estimators. *Journal of the American Statistical Association*, 70(350), 407-416. <https://doi.org/10.1080/01621459.1975.10479882>.
- [24] Mirjalili, S., & Lewis, A. (2013). S-shaped versus V-shaped transfer functions for binary Particle Swarm Optimization. *Swarm and Evolutionary Computation*, 9, 1-14. <https://doi.org/10.1016/j.swevo.2012.09.002>.
- [25] Nomura, M. (1988). On the almost unbiased ridge regression estimator. *Communications in Statistics-Simulation and Computation*, 17(3), 729-743. <https://doi.org/10.1080/03610918808812690>.
- [26] Qasim, M., Amin, M., & Amanullah, M. (2018). On the performance of some new Liu parameters for the gamma regression model. *Journal of Statistical Computation and Simulation*, 88(16), 3065-3080. <https://doi.org/10.1080/00949655.2018.1498502>.

- [27] Troskie, C., & Chalton, D. (1996). *Detection of outliers in the presence of multicollinearity*. Paper presented at the Multidimensional statistical analysis and theory of random matrices, Proceedings of the Sixth Lukacs Symposium, eds. Gupta, AK and VL Girko. <https://doi.org/10.1515/9783110916690-022>.
- [28] Wen, J. H., Zhong, K. J., Tang, L. J., Jiang, J. H., Wu, H. L., Shen, G. L., & Yu, R. Q. (2011). Adaptive variable-weighted support vector machine as optimized by particle swarm optimization algorithm with application of QSAR studies. *Talanta*, 84(1), 13-18. <https://doi.org/10.1016/j.talanta.2010.11.039>.
- [29] Yang, S.-P., & Emura, T. (2017). A Bayesian approach with generalized ridge estimation for high-dimensional regression and testing. *Communications in Statistics-Simulation and Computation*, 46(8), 6083-6105. <https://doi.org/10.1080/03610918.2016.1193195>.
- [30] Zhou, W., & Dickerson, J. A. (2014). A novel class dependent feature selection method for cancer biomarker discovery. *Computers in Biology and Medicine*, 47, 66-75. <https://doi.org/10.1016/j.compbiomed.2014.01.014>.