Variable selection in Gamma regression model using binary gray Wolf optimization algorithm

1) Niam Abdulmunim Al-Thanoon, 2) Omar Saber Qasim. 3) Zakariya Yahya Algamal

1) Department of Operations Research and Artificial Intelligence, University of Mosul, Mosul, Iraq
2) Department of Mathematics, University of Mosul, Mosul, Iraq
3) Department of Statistics and Informatics, University of Mosul, Iraq.

E-mail: zakariya.algamal@uomosul.edu.iq

Abstract. In the real life applications, large amounts of variables have been accumulated quickly. Selection of variables is a very useful tool for improving the prediction accuracy by identifying the most relative variables that related to the study. Gamma regression model is one of the most models that applied in several science fields. Gray Wolf optimization algorithm (GWO) is one of the proposed nature-inspired algorithms that can efficiently be employed for variable selection. In this paper, chaotic GWO is proposed to perform variable selection for gamma regression model. The simulation studies and a real data application are used to evaluate the performance of our proposed procedure in terms of prediction accuracy and variable selection criteria. The obtained results demonstrated the efficiency of our proposed methods comparing with other popular methods.

1. Introduction

Gamma regression model is widely applied method for studying automobile insurance claims and medical science [1-3]. “Specifically, when the response variable under the study is distributed as gamma distribution [4,5].

In many real applications, recent developments in technologies have made the possibility to measure a large number of variables. In the regression modeling, the existence of huge number has a negative effect by overfitting the regression model. Therefore, identification of a small subset of important variables from a large number of variables set for accurate prediction is an important role for building predictive regression models [6].

When the number of variables increases, the traditional variable selection methods, such as stepwise selection, forward selection, and backward elimination computationally become an exhaustive search and require a long time for computing. Penalization methods, (lasso) [7], (scad) [8], elastic net [9], and adaptive lasso [10], are become an attractive methods for simultaneously performing variable selection and model estimation.
Recently, the naturally inspired algorithms, such as genetic algorithm, particle swarm optimization algorithm, firefly algorithm, and Gray Wolf optimization algorithm, have a great attraction and proved their efficiency as variable selection methods [11]. This is because that the main target in variable selection is to minimize the number of selected variables while maintaining the maximum accuracy of prediction, and, therefore, they can be considered as optimization problems [12].

Several researchers have employed the naturally inspired algorithms for variable selection in regression models. Broadhurst, et al. [13] employed the genetic algorithm for variable selection in linear regression models, with application in chemometrics. Drezner, et al. [14] proposed to use tabu search algorithm in model selection in the linear regression model. On the other hand, a hybrid algorithm of genetic algorithm and simulated annealing was proposed as a subset selection method in linear regression model by Örkcü [15]. Brusco [16] performed a comparison of simulated annealing algorithms for variable selection in principal component analysis and discriminant analysis. Besides, the differential evolution algorithm was used as a variable selection in linear regression model by Dünder, et al. [17]. In generalized linear models, the natural inspired algorithms for variable selection are also used, such as, logistic regression model [18,19], Poisson regression model [20,21], and gamma regression model [22].

The purpose of this paper is to propose chaotic GWO, which is a swarm intelligence technique, as an alternative variable selection method for use in gamma regression model. The proposed algorithm will efficiently help in identifying the most relevant variables in the count data regression model with a high prediction. The superiority of the proposed algorithm is proved through different simulation settings and a real data application”.

2. Gamma regression model

In epidemiology, social, and economic studies, positively skewed data are often arisen. Gamma distribution is a well-known distribution that fits such type of data. “Gamma regression model (GRM) is used to model the relationship between the non-negative skewed response variable and potentially variables [23].

Assume $y_i$ is the response variable which is following a gamma distribution with shape parameter $\nu$ and scale parameter $\gamma$, i.e. $y_i \sim \text{Gamma}(\nu, \gamma)$, then the probability density function is defined as

$$f(y_i) = \frac{\gamma}{\Gamma(\nu)} y_i^{\nu-1} e^{-\gamma y_i}, \quad y_i \geq 0,$$

with $E(y) = \nu / \gamma = \theta$ and $\text{var}(y) = \nu / \gamma^2 = \theta^2 / \nu$. Given that $\gamma = \nu / \theta$, Eq. (3) can re-parameterized as a function of the mean ($\theta$) and the shape ($\nu$) parameters and written depending on the exponential function as

$$f(y_i) = \text{EXP} \left\{ y_i (-1/\theta) - \frac{\log(-1/\theta)}{1/\nu} + c(y_i, \nu) \right\},$$

where the canonical link function is $-1/\theta$, the dispersion parameter is $\phi = 1/\nu$ and $c(y_i, \nu) = \nu \log(\nu) + \nu \log(y_i) - \log(y_i) - \log(\Gamma(\nu))$. 


GRM is usually modeled using the canonical link function (reciprocal), \( \theta_i = -1/x_i^T \beta \) which is expressed as a linear combination of covariates \( x_i = (x_{i1}, ..., x_{ip})^T \). The log link function, \( \theta_i = \exp(x_i^T \beta) \), is alternatively used rather than the reciprocal link function because it ensures that \( \theta_i > 0 \).

The maximum likelihood method of Eq. (4) is the most common method of estimating the coefficients of GRM. Assuming that the observations are independent and \( \theta_i = -1/x_i^T \beta \), the log-likelihood function is given by

\[
\ell(\beta) = \sum_{i=1}^{n} \left\{ y_i x_i^T \beta - \log(x_i^T \beta) \right\} + c(y_i, v),
\]

the ML estimator is then obtained by computing the first derivative of the Eq. (3) and setting it equal to zero, as

\[
\frac{\partial \ell(\beta)}{\partial \beta} = \frac{1}{v} \sum_{i=1}^{n} \left[ y_i - \frac{1}{x_i^T \beta} \right] x_i = 0.
\]

Depending on the iteratively weighted least squares (IWLS) algorithm, in each iteration, the parameters are updated by

\[
\beta^{(r+1)} = \beta^{(r)} + I^{-1}(\beta^{(r)}) S(\beta^{(r)}),
\]

where \( S(\beta) = \frac{\partial \ell(\beta)}{\partial \beta} \) and \( I^{-1}(\beta) = -E \left( \frac{\partial^2 \ell(\beta)}{\partial \beta \partial \beta^T} \right)^{-1} \). The final step of the estimated coefficients is defined as

\[
\hat{\beta}_{GR} = (X^T \hat{W} X)^{-1} X^T \hat{W} \hat{u},
\]

where \( \hat{W} = \text{diag}(\hat{\theta}_i^2) \) and \( \hat{u} \) is a vector where \( i^{th} \) element equals” to \( \hat{u}_i = \hat{\theta}_i + (y_i - \hat{\theta}_i) / \hat{\theta}_i^2 \).

### 3. Chaotic grey wolf optimization algorithm

Mirjalili, et al. [24] presented a new metaheuristics algorithms as a swarm intelligence, “which is known as the grey wolf optimizer (GWO) algorithm. The GWO simulate the behavior of leadership and hunting in organisms of grey wolf. The GWO simulates the driving hierarchy in the environment and this distinguishes it from the rest of the swarm algorithms. The simulation of hunting in the GWO algorithm is done through the hierarchy of leadership, where the crowd is divided into different groups and levels such as alpha, beta, and omega [24].”

Gray wolves belong to the Canidae family and are classified as top predators because they belong to the top of the food chain. The first level of the leadership hierarchy is the alpha (\( \alpha \)) type and they represent the leaders, they may be female or male, and they are responsible for making all the
decisions related to hunting, sleep, time to wake and so on. “The second level in the hierarchy of leadership is the beta ($\beta$), where these wolves are helping wolves in the first level of the alpha in making decisions. Wolves in the second level ($\beta$) respect wolves in the first level ($\alpha$) and reinforce decision-making and act as their consultant. In the third level, there is a type of omega ($\gamma$) and plays the role of scapegoat for the flock. All wolves from other levels are submitted to wolves of the omega type. It may seem that wolves in the third level are not an important person, but it is observed that the group without them face fighting and internal problems. This is due to the venting of vehemence and frustration of all wolves by the omega ($\gamma$). This helps in fulfilling the whole pack and preserve the dominance structure [25]. Wolves, which are not alpha ($\alpha$), beta ($\beta$), or omega ($\gamma$), are called the subordinate or delta ($\delta$), and wolves in this species must be subjugated to alpha ($\alpha$) and beta ($\beta$), but they dominate the omega ($\gamma$) wolves.

Mathematical models for each level of the leadership pyramid of the GWO are calculated through the following:

$$\vec{D} = \vec{c}_x p(t) - \vec{x}(t),$$

(7)

$$\vec{x}(t+1) = \vec{x}_p(t) - \vec{a} \vec{d},$$

(8)

where $t$ shows the current iteration, $\vec{x}_p$ indicates the position vector of the prey, $\vec{x}$ represent the position vector of a grey wolf. The vectors $\vec{a}$ and $\vec{c}$ are defined mathematically as follows:

$$\vec{a} = 2\vec{L} \vec{r}_1 - \vec{L},$$

$$\vec{c} = 2\vec{r}_2,$$

(9)

where the components of $\vec{L}$ are linearly reduced from 2 to 0 over the course of iterations and $\vec{r}_1, \vec{r}_2$ are random vectors in $[0, 1]$.}

### 3.1.1 Hunting

There are three main steps that are applied during hunting prey. There are: (1) the search for prey, (2) encircling, and, (3) attacking. The mathematical behavior of the gray wolf algorithm is simulated by assuming that alpha ($\alpha$), beta ($\beta$), and delta ($\delta$) have potential knowledge of the prey location. Mathematical equations in this regard are developed by

$$\vec{D}_{\alpha} = \vec{c}_{\alpha} \vec{x}_a - \vec{x},$$

$$\vec{D}_{\beta} = \vec{c}_{\beta} \vec{x}_\beta - \vec{x},$$

$$\vec{D}_{\delta} = \vec{c}_{\delta} \vec{x}_\delta - \vec{x},$$

(10)

$$\vec{x}_1 = \vec{x}_a - \vec{a}_1 (\vec{D}_{\alpha}),$$

$$\vec{x}_2 = \vec{x}_\beta - \vec{a}_2 (\vec{D}_{\beta}),$$

$$\vec{x}_3 = \vec{x}_\delta - \vec{a}_3 (\vec{D}_{\delta}),$$

(11)

$$\vec{x} = \frac{\vec{x}_1 + \vec{x}_2 + \vec{x}_3}{3}.$$  

(12)
\[ \tilde{a}_i = 2\bar{L} \tilde{r}_1 - \bar{L}, \]
\[ \tilde{c}_i = 2\tilde{r}_2, \] (13)

where \( \tilde{a} \) is a random value in the interval \([-\bar{L}, \bar{L}]\). The gray wolves are compelled to attack the prey when random value \( a < 1 \). The prey is searched through exploration ability and attack prey the ability to exploit. The arbitrary values of \( L \) are utilized to force the search to move away from the prey [27]. The arbitrary values of \( L \) are applied to force the search to move away from the prey.

The positions of gray wolves are continuously changing in space to whatever point. In some problems such as feature selection, solutions are limited to binary 0 or 1 values. In such case, BGWO is proposed by Emary, et al. [28]. The wolves update equation is a function of three position vectors namely \( x_\alpha, x_\beta, x_\delta \) which can attract each wolf of the flock towards the first three best solutions. In any given time, the aggregation of solutions is in binary form and all the solutions are on the corner of a hypercube. To update the positions of the given wolf based on the basic GWO algorithm, while keeping the binary restriction according to the Eq. (14).

The main updating equation in the bGWO algorithm can be formulated, in this approach as follows [28]:

\[ x_{i+1} = \text{crossover}(x_1, x_2, x_3), \] (14)

where \( \text{crossover}(x, y, z) \) is a suitable crossover between solutions \( x, y, z \) and \( x, x_2, x_3 \) are binary vectors representing the effect of a wolf in bGWO, which move towards the alpha; beta; delta gray wolves in order. \( x_1, x_2, x_3 \) are calculated using Eqs. (15), (18), and (21), respectively, as

\[ x_i^d = \begin{cases} 
1 & \text{if } (x_i^d + bstep_i^d) \geq 1, \\
0 & \text{OW} 
\end{cases}, \] (15)

where \( x_i^d \) represents the position vector of the alpha (\( \alpha \)) wolf in the dimension \( d \), and \( bstep_i^d \) is a binary step in the dimension \( d \) which is calculated by the following equation:

\[ bstep_i^d = \begin{cases} 
1 & \text{if } cstep_i^d \geq \text{rand} \\
0 & \text{OW} 
\end{cases}, \] (16)

where \( \text{rand} \) is a random number derived from the uniform distribution in the closed period \([0,1]\), and \( cstep_i^d \) is the continuous-valued step size for dimension \( d \) and can be calculated using the sigmoidal function through Eq. (17).

\[ cstep_i^d = \frac{1}{1 + e^{-10(\text{a}_i^d D_d^a - 0.5)}}, \] (17)
where \( x'_a \) and \( b'_a \) are calculated using Eqs. (9) and (10) in the dimension \( d \).

\[
x'_a^d = \begin{cases} 
1 & \text{if } (x^d_\beta + bstep^d_\beta) \geq 1 \\
0 & \text{OW}
\end{cases},
\]

(18)

where \( x^d_\beta \) represents the position vector of the beta (\( \beta \)) wolf in the dimension \( d \), and \( bstep^d_\beta \) is a binary step in the dimension \( d \) which is calculated by the following equation:

\[
bstep^d_\beta = \begin{cases} 
1 & \text{if } cstep^d_\beta \geq \text{rand} \\
0 & \text{OW}
\end{cases},
\]

(19)

where \( \text{rand} \) is a random number derived from the uniform distribution in the closed period \( [0,1] \), and \( cstep^d_\beta \) is the continuous-valued step size for dimension \( d \) and can be calculated using the sigmoidal function through Eq. (20)

\[
cstep^d_\beta = \frac{1}{1+e^{-10((a'_d b'_d)-0.5)}},
\]

(20)

where \( a'_d \) and \( b'_d \) are calculated using equations (9), and (10) in the dimension \( d \).

\[
x'_\delta^d = \begin{cases} 
1 & \text{if } (x^d_\delta + bstep^d_\delta) \geq 1 \\
0 & \text{OW}
\end{cases},
\]

(21)

where \( x^d_\delta \) represents the position vector of the delta (\( \delta \)) wolf in the dimension \( d \), and \( bstep^d_\delta \) is a binary step in the dimension \( d \) which is calculated by the following equation:

\[
bstep^d_\delta = \begin{cases} 
1 & \text{if } cstep^d_\delta \geq \text{rand} \\
0 & \text{OW}
\end{cases},
\]

(22)

where \( \text{rand} \) is a random number derived from the uniform distribution in the closed period \( [0,1] \), and \( cstep^d_\delta \) is the continuous-valued step size for dimension \( d \) and can be calculated using the sigmoidal function through Eq. (23)

\[
cstep^d_\delta = \frac{1}{1+e^{-10((a'_d b'_d)-0.5)}},
\]

(23)

where \( a'_d \) and \( b'_d \) are calculated using Eqs. (9) and (10) in the dimension \( d \). The crossover process is then applied to each of the solutions \( a,b,c \) as shown in the following equation:
where $a_d, b_d$, and $c_d$ represent the binary values for the first, second and third parameter in dimension $d$, $x_d$ is the crossover process output at dimension $d$, and $\text{rand}$ is a random number derived from the uniform distribution in the closed period [0, 1]. The fitness function is defined as

$$\text{fitness} = \min \left[ \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \right].$$

(25)

### 4. Computational results

In this section, the performance of our proposed variable selection method, CGWO is tested. Further, the performance of CGWO is compared with the GWO, Bayesian information criteria (BIC), and Akaike information criteria (AIC) that are defined as, respectively,

$$\text{AIC} = -2\ell(\hat{\beta}) + 2 \times q,$$

(26)

$$\text{BIC} = -2\ell(\hat{\beta}) + \log(n) \times q,$$

(27)

where $\ell(\hat{\beta})$ is the log-likelihood for PRM and $q$ is the number of selected variables.

### 5. Simulation results

In this section, the same simulation settings of Algamal and Lee [29] and Wang, et al. [30] are used. The sample size is considered with $n \in \{50, 100, 200\}$.

**Simulation 1:** “In this simulation, 20 explanatory variables are generated from multivariate normal distributions with mean vector 0 and covariance matrix $\Sigma$ which elements $\rho(x_i, x_j) = \rho^{i-j}$ with $\rho = 0.5$. The true vector of parameters is given by $\beta = (1.5, -1.5, 1.5, -1.6, 1.5, 0, ..., 0, 1.5, -1.5, 1.5, -1.6, 1.5, 0, ..., 0)^T$ with 10 true explanatory variables and the rest in non-true variables.

**Simulation 2:** Here, The true vector of parameters is given by $\beta = (1.6, -0.88, 0.95, -1.10, 0.70, 0, ..., 0)^T$ with 5 true explanatory variables and 15 non-true variables. The explanatory variables are generated as same as simulation 1 with $\rho(x_i, x_j) = 0.5$.

**Simulation 3:** In this simulation, 8 explanatory variables are generated as same as simulation 1 with $\rho(x_i, x_j) = 0.5^{i-j}$. The true parameter vector is given by $\beta = (0.25, ..., 0.25)^T$. 

7
For all the simulation examples 1 – 3, the response variable is generated according to PRM as \( y_i \sim \text{Gamma}(\exp(x_i^T \beta), 0.5) \). For performance evaluation of the CGWO, the mean squared error (MSE) is used as a prediction accuracy criteria, which is defined as \( \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 / n \). In terms of variable selection performance, the number of the truly nonzero coefficients which are incorrectly set to zero (I), and the number of the true zero coefficients which are correctly set to zero (C). The higher the values of C, and the lower the values of I, the better the variable selection performance is. All computations of this paper were conducted using R. Based on 300 times of repeating simulation, the averaged MSE, I, and C are listed in Tables 1 – 3, respectively.

It shows from these tables that the CGWO method there has a significant improvement where it has a much better average of MSE than those GWO, AIC, and BIC methods. For instance, in Table 1 when \( n = 50 \), the MSE reduction by CGWO was about 42.35%, 34.90%, and 30.79% comparing with AIC, BIC, and GWO respectively. Further, regardless of the value of \( n \), the CGWO often shows the smallest MSE among the competitor methods.

In terms of variable selection performance, our proposed method obviously selects a very few irrelevant variables comparing with GWO, AIC, and BIC, where the number of the true zero coefficients which are correctly set to zero is high comparing with others. For example, in Table 3 when \( n = 200 \), CGWO does not select, on average, about 8 irrelevant variables out of 10 irrelevant variables. While PSO, AIC, and BIC select more than 4 irrelevant variables. On the other hand, CGWO performs very well with the smallest I (the number of the truly nonzero coefficients which are incorrectly set to zero) among all the used methods. This indicates that CGWO misses a very few important variables.

From the results of simulation 3 (Table 3), the model is dense, and, therefore, all the methods have zero values for the criterion C. On the other hand, CGWO is the best because the number of nonzero variables that have been identified as irrelevant variables is smaller compared with GWO, AIC, and BIC. It is worth noting that AIC has inferior performance in all simulation examples comparing with GWO, BIC, and CGWO methods”.

Table 1: Simulation 1 results, on average.

| Methods | MSE  | C   | I   |
|---------|------|-----|-----|
|         | \( n = 50 \) |     |     |
| CGWO    | 5.933 | 8.462 | 0.711 |
| AIC     | 10.293 | 5.373 | 4.188 |
| BIC     | 9.114  | 5.112 | 3.919 |
| GWO     | 8.573  | 6.915 | 3.143 |
|         | \( n = 100 \) |     |     |
| CGWO    | 5.746  | 8.586 | 0.767 |
| AIC     | 10.006 | 5.497 | 4.244 |
| BIC     | 8.917  | 5.236 | 3.975 |
| GWO     | 8.389  | 7.039 | 3.199 |
|         | \( n = 200 \) |     |     |
| CGWO    | 5.697  | 8.623 | 1.342 |
| AIC     | 10.057 | 5.534 | 4.219 |
| BIC     | 8.878  | 5.273 | 4.55  |
| GWO     | 8.337  | 7.076 | 2.959 |
Table 2: Simulation 2 results, on average.

| Methods | MSE  | C     | I     |
|---------|------|-------|-------|
|         | $n = 50$ |       |       |
| CGWO    | 7.405 | 13.237| 1.214 |
| AIC     | 11.765| 7.927 | 3.52  |
| BIC     | 10.586| 7.142 | 3.131 |
| GWO     | 10.045| 9.513 | 2.762 |
|         | $n = 100$ |     |       |
| CGWO    | 7.218 | 13.301| 1.246 |
| AIC     | 11.478| 7.991 | 3.552 |
| BIC     | 10.389| 7.206 | 3.163 |
| GWO     | 9.861 | 9.577 | 2.794 |
|         | $n = 200$ |     |       |
| CGWO    | 7.169 | 13.312| 1.254 |
| AIC     | 11.529| 8.002 | 3.56  |
| BIC     | 10.35 | 7.217 | 3.171 |
| GWO     | 9.809 | 9.588 | 2.802 |

Table 3: Simulation 3 results, on average.

| Methods | MSE  | C | I |
|---------|------|---|---|
|         | $n = 50$ | | |
| CGWO    | 7.079 | 0 | 1.254 |
| AIC     | 11.439| 0 | 3.56 |
| BIC     | 10.26 | 0 | 3.171 |
| GWO     | 9.719 | 0 | 2.802 |
|         | $n = 100$ | | |
| CGWO    | 6.892 | 0 | 0.356 |
| AIC     | 11.152| 0 | 2.579 |
| BIC     | 10.063| 0 | 2.371 |
| GWO     | 9.535 | 0 | 0.99 |
|         | $n = 200$ | | |
| CGWO    | 6.843 | 0 | 0.33 |
| AIC     | 11.203| 0 | 2.551 |
| BIC     | 10.024| 0 | 2.222 |
| GWO     | 9.483 | 0 | 0.914 |

6. Real application result

To make the benefit of the our proposed method in the real application, “a chemistry dataset with $(n, p) = (65,15)$, of imidazo[4,5-b]pyridine derivatives [31]. The response of interest is the biological activities (IC$_{50}$) [32]. A Chi-square test as a goodness of fit is used to check whether the biological activities variables has the gamma distribution. The result of the test equals to 9.3657 with
p-value equals to 0.9534. This indicating that the gamma distribution fits very well to this response variable. The estimation of the dispersion parameter is 0.0066”.

Table 4 summarizes the MSE and the selected variables for each used method for the real data application.

As seen from the result of Table 4, CGWO can remarkably reduce the MSE comparing with GWO, AIC, and BIC. In terms of selected variables, on the other hand, it clearly seen from Table 4 that CGWO only select 6 variables out of 15 variables when the gamma model is assumed. CGWO selected the explanatory variables $x_1, x_2, x_7, x_8, x_{11}$, and $x_{15}$. These selected variables are identified as relevant variables to the study. Comparing with GWO and BIC, CGWO includes few variables with the MSE is less than them”.

| Methods | Selected variables | MSE  |
|---------|--------------------|------|
| CGWO   | $x_1, x_7, x_8, x_{11}, x_{15}, x_2$ | 1592.21 |
| AIC    | $x_1, x_2, x_3, x_6, x_8, x_{10}, x_{11}, x_{14}$ | 1862.86 |
| BIC    | $x_1, x_2, x_3, x_6, x_7, x_9, x_{10}, x_{11}, x_{14}, x_{15}$ | 1831.81 |
| GWO    | $x_1, x_2, x_5, x_7, x_8, x_{11}, x_{15}$ | 1625.05 |

7. Conclusion

In this paper, the problem of selecting variables in gamma regression model is considered. A chaotic grey wolf optimization algorithm was proposed as a variable selection method. The results obtained from simulation examples and real data applications demonstrated the superiority of the CGWO in terms of MSE, I, and C comparing with GWO, AIC, and BIC methods.

REFERENCES

1. De Jong P, Heller GZ. Generalized linear models for insurance data. Cambridge University Press Cambridge, 2008.
2. Dunder E, Gumustekin S, Cengiz MA. Variable selection in gamma regression models via artificial bee colony algorithm. Journal of Applied Statistics. 2016; 1-9.
3. Malehi AS, Pourmotahari F, Angali KA. Statistical models for the analysis of skewed healthcare cost data: A simulation study. Health Economics Review. 2015; 5: 1-11.
4. Al-Abood AM, Young DH. Improved deviance goodness of fit statistics for a gamma regression model. Communications in Statistics - Theory and Methods. 1986; 15: 1865-1874.
5. Hattab WM. A derivation of prediction intervals for gamma regression. Journal of Statistical Computation and Simulation. 2016; 86: 3512-3526.
6. Algamal ZY, Lee MH. Penalized logistic regression with the adaptive LASSO for gene selection in high-dimensional cancer classification. Expert Systems with Applications. 2015; 42: 9326-9332.
7. Tibshirani R. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 1996; 58: 267-288.
8. Fan J, Li R. Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American Statistical Association. 2001; 96: 1348-1360.
9. Zou H, Hastie T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2005; 67: 301-320.
10. Zou H. The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*. 2006; 101: 1418-1429.
11. Sayed GI, Hassanien AE, Azar AT. Feature selection via a novel chaotic crow search algorithm. *Neural Computing and Applications*. 2017.
12. Sindhu R, Ngadiran R, Yacob YM, Zahri NAH, Hariharan M. Sine–cosine algorithm for feature selection with elitism strategy and new updating mechanism. *Neural Computing and Applications*. 2017; 28: 2947-2958.
13. Broadhurst D, Goodacre R, Jones A, Rowland JJ, Kell DB. Genetic algorithms as a method for variable selection in multiple linear regression and partial least squares regression, with applications to pyrolysis mass spectrometry. *Analytica Chimica Acta*. 1997; 348: 71-86.
14. Drezner Z, Marcoulides GA, Salhi S. Tabu search model selection in multiple regression analysis. *Communications in Statistics - Simulation and Computation*. 1999; 28: 349-367.
15. Örkü H. Subset selection in multiple linear regression models: A hybrid of genetic and simulated annealing algorithms. *Applied Mathematics and Computation*. 2013; 219: 11018-11028.
16. Brusco MJ. A comparison of simulated annealing algorithms for variable selection in principal component analysis and discriminant analysis. *Computational Statistics & Data Analysis*. 2014; 77: 38-53.
17. Dünder E, Gümüştekin S, Murat N, Cengiz MA. Variable selection in linear regression analysis with alternative Bayesian information criteria using differential evaluation algorithm. *Communications in Statistics - Simulation and Computation*. 2017; 47: 605-614.
18. Pacheco J, Casado S, Núñez L. A variable selection method based on Tabu search for logistic regression models. *European Journal of Operational Research*. 2009; 199: 506-511.
19. Unler A, Murat A. A discrete particle swarm optimization method for feature selection in binary classification problems. *European Journal of Operational Research*. 2010; 206: 528-539.
20. Koç H, Dünder E, Gümüştekin S, Koç T, Cengiz MA. Particle swarm optimization-based variable selection in Poisson regression analysis via information complexity-type criteria. *Communications in Statistics - Theory and Methods*. 2017: 1-9.
21. Massaro TJ, Bozdogan H. Variable subset selection via GA and information complexity in mixtures of Poisson and negative binomial regression models. *arXiv preprint arXiv:1505.05229*. 2015.
22. Dunder E, Gumustekin S, Cengiz MA. Variable selection in gamma regression models via artificial bee colony algorithm. *Journal of Applied Statistics*. 2016; 45: 8-16.
23. Uusipaikka E. *Confidence intervals in generalized regression models*. Chapman & Hall/CRC Press: NW, 2009.
24. Mirjalili S, Mirjalili SM, Lewis A. Grey wolf optimizer. *Advances in engineering software*. 2014; 69: 46-61.
25. Singh N, Hachimi H. A New Hybrid Whale Optimizer Algorithm with Mean Strategy of Grey Wolf Optimizer for Global Optimization. *Mathematical and Computational Applications*. 2018; 23: 14.
26. Li Q, Chen H, Huang H, Zhao X, Cai Z, Tong C, Liu W, Tian X. An Enhanced Grey Wolf Optimization Based Feature Selection Wrapped Kernel Extreme Learning Machine for Medical Diagnosis. *Computational and mathematical methods in medicine*. 2017; 2017: 9512741.
27. Emary E, Zawbaa HM, Grosan C. Experienced gray wolf optimization through reinforcement learning and neural networks. *IEEE transactions on neural networks and learning systems*. 2018; 29: 681-694.
28. Emary E, Zawbaa HM, Hassanien AE. Binary grey wolf optimization approaches for feature selection. *Neurocomputing*. 2016; **172**: 371-381.

29. Algamal ZY, Lee MH. Adjusted adaptive lasso in high-dimensional Poisson regression model. *Modern Applied Science*. 2015; **9**: 170-176.

30. Wang Z, Ma S, Zappitelli M, Parikh C, Wang CY, Devarajan P. Penalized count data regression with application to hospital stay after pediatric cardiac surgery. *Stat. Meth. Med. Res*. 2014; In press.

31. Algamal ZY, Lee MH, Al-Fakih AM, Aziz M. High-dimensional QSAR prediction of anticancer potency of imidazo[4,5-b]pyridine derivatives using adjusted adaptive LASSO. *Journal of Chemometrics*. 2015; **29**: 547-556.

32. Algamal ZY, Lee MH. A novel molecular descriptor selection method in QSAR classification model based on weighted penalized logistic regression. *Journal of Chemometrics*. 2017; **31**: e2915.