Journal of Engineering and Applied Sciences 2 (10): 1509-1515, 2007
ISSN: 1816-949X

Improvement of Ridge Regression Using Differential Evolution

[1]Sung-Hae Jun and [2]Im-Geol Oh
[1]Department of Bioinformatics and Statistics, Cheongju University, Chungbuk, Korea
[2]Department of Internet Engineering, Hanseo University, Chungnam, Korea

Abstract: Multicollinearity problem in learning machines occurs when there are high dependencies among the input variables. The problem increases the variance of predictive model to cause unstable results. In regression models, the multicollinearity is also a problem to be solved. Ridge regression is a good method to settle the problem of regression. In general, the shrinkage parameter of ridge regression is determined by the arts of researchers. But, the selections are not always good. So, in this study, we propose an improvement of ridge regression using differential evolution. This is an evolutionary ridge regression to find better shrinkage parameter. To verify performance of our research, we make experiments using objective data sets.

Key words: Improvement, ride regression, differential evolution, regression model

INTRODUCTION

Regression is a predictive method for data mining. Generally, this has one or more explanatory variables which predict or explain a response variable with quantitative or qualitative values. Using regression, we have got good performances in supervised learning of data mining. But, multicollinearity decreases the efficiency of regression model. That is, when high correlation within input variables is in regression data, multicollinearity raises the serious problem in the inference of regression coefficients. So, we need a settlement the problem of multicollinearity in regression. A solution of the problem is ridge regression. This is one of biased estimation techniques. Also, we overcome the problem of multicollinearity by ridge regression. Ridge regression was proposed by Hoerl and Kennard (1970) researched the ridge regression for applying to nonorthogonal problems. The nonorthogonal problems raise unsatisfactory least square results (Hines *et al*., 2002; Hoerl and Kennard, 1970). The ridge regression is able to lead to good estimator estimates in nearly rank deficient problems (Ismail and Principe, 1996). The ridge estimator diminishes the mean square error by reducing the variance compared with ordinary least squares estimators. So, the main goal of the ridge regression is to solve the problems of the least squares method which are encountered whenever the input vectors are highly correlated, that is, multicollinearity (Olague *et al*., 2003). In ridge regression, we have to determine a shrinkage parameter which is a small positive quantity. For example, when this parameter is 0, the regression model is to be linear regression. Generally the shrinkage parameter of ridge regression is determined by the arts of researchers using prior knowledge. But, the selections are not always good and efficient. In this study, we propose an improvement of ridge regression using Differential Evolution (DE-ridge) to settle the problem of ridge regression. Our proposed is an evolutionary ridge regression to find better shrinkage parameter than previous determined parameters. To verify performance of DE-ridge, we make experiments using objective data sets from objective machine learning data and simulation data.

Differential evolution: Differential Evolution (DE) is a parallel direct and a population based search methods (Engelbrecht, 2002; Price *et al*., 2005; Ronkkonen *et al*., 2005; Storn and Price, 1997). DE is not depended on a mutation operator based on probability distribution. A new operator is applied to DE. The operator of uses the differences between randomly selected individuals. Also, DE supports good genetic method to solve real valued problems. In general, DE combines two vectors with another vector. Compared with other evolutionary algorithms, DE has fast computing speed to get the solutions (Storn and Price, 1997). We show the general view of DE in the following.

Corresponding Author: Sung-Hae Jun, Department of Bioinformatics and Statistics, Cheongju University, Chungbuk, Korea

(s1) Initialize the population.
(s2) Select randomly target vectors, (x_1, x_2, x_3).
(s3) Build weighted difference vector, $F(x_1 - x_2)$.
(s4) Add x_3 to $F(x_1 - x_2)$ for generating a trial vector, v
$v = x_3 + F(x_1 - x_2)$.
(s5) Do crossover the trial vector and the current vector, x_i with Crossover Rate (CR).
(s6) Replace or retain x_i.

Shrinkage parameter of ridge regression: Ridge regression is one of the methods for solving multicollinearity (Myers, 1990). Multicollinearity caused by near-linear dependencies among the input variables may produce large variances. So, it decreases the performances of predic-tive models. Linear regression is a estimation procedure by Ordinary Least Squares (OLS) (Hastie *et al*., 2001). OLS gives unbiased estimate and the minimum variance of all unbiased estimators. But, OLS does not give upper bound on the variance of the estimators. In addition, it may produce large variance when multicollinearity presents. Ridge regression is one of biased estimation techniques to solve the problems of linear regression by OLS. The ridge regression estimator of the coefficient β is computed by solving for b_{ridge} in the following equation.

$$(X'X + kI)b_{ridge} = X'y \quad (1)$$

Where X is the design matrix and displays the input data. Y represents known output data and I is identity matrix. K ($\geq$ 0) is a shrinkage parameter. b_{ridge} is given by the following.

$$b_{ridge} = (X'X + kI)^{-1}X'y \quad (2)$$

In ridge regression, the role of k is to moderate the variance of the estimators. So, it is important to determine the optimal k. But, the selection of k is performed subjectively using different measures which are ridge trace, repeated method to convergence, Prediction sum of Squares (PRESS), Generalized Crossvalidation (GCV), VIF, Hoerl-Kennard-Baldwin (HKB) estimator (Myers, 1990; Zhang and Horvath, 2006). Also, λ has been determined by a tuning and validation set (Embrechts, 2004). These approaches demand quite computing time consuming for large data sets (Huet *et al*., 2003). So, we propose an efficient method for determining optimal k using an evolutionary ridge regression.

DIFFERENTIAL EVOLUTION BASED RIDGE REGRESSION

In this study, an efficient method for improving ridge regression using differential evolution is proposed. We call this differential evolution based ridge regression (DE-ridge). Differential evolution is used to determine the shrinkage parameter of the ridge regression in our research. We use PRESS (prediction sum of square) statistic for constructing the fitness function of DE-ridge. The statistic is based on the following residual sets.

$$R_j = \hat{f}(x_j) - y_j \quad (3)$$

$\hat{f}(x_j)$, which is jth output, is predictive values of output variable by input x_j. Also, y_j is practical jth observation of output values. In this study, we define fitness function of DE-ridge as the following.

$$Fit(x_i) = \sum_{i=1}^{n}\left(y_i - f(x_{i,-i})\right)^2 \quad (4)$$

Where, $f(x_{i,-i})$ is the predictive value evaluated at $x = x_i$ and y_i is not used in obtaining the regression coefficients. Our DE-ridge algorithm is shown in the following four steps.

(1: Initial step)
Let g = 0;
initialize p,
initialize k
initialize population L_g with n individuals

In this step, g, p and k in the initial step are generation, reproduction probability and scaling factor respectively. The range of p is between 0 and 1. Also k has positive values. L_g is initial population which has candidate solutions of shrinkage parameters for ridge regression. We reproduce a candidate parameter of population using next DE reproduction step.

(2: Reproduction step)
For a $l_{g,i}$ in $\{l_{g,1}, l_{g,2}, l_{g,n}\}$
select n_1, n_2, n_3 from U (1, … ,n) ($n_1 \# n_2 \# n_3$)
select r~U(1,…,R)
for s=1,…R

if (U(0,1) < p or s=r)

$$O_{g,ni} = l_{g,n_3 i} + k(l_{g,n_1 i} - l_{g,n_2 i})$$

Else

$$O_{g,ni} = l_{g,ni}$$

l_{gj} is an offspring of generation g from current population $\{l_{g,1}, l_{g,2},..., l_{g,n}\}$. We randomly select three candidates, which are $l_{g,n1}$,$l_{g,n2}$ and $l_{g,n3}$ with n_1 # n_2 # n_3 from the current population. Also, n_1, n_2 and n_3 are generated from uniform distribution U(1, …, n). Continuously r, which is the number of genes of a chromosome, is selected from U(1, … , R). $O_{g,ni}$ and $l_{g,ni}$ are the jth parameters of the offspring and the parent.

(3: New population step)

Select new population L_{g+1}

$$l_{g+1,n} = \begin{cases} o_{g,n} & \text{if } Fit(o_{g,n}) \le Fit(l_{g,n}) \\ l_{g,n} & \text{otherwise} \end{cases}$$

We are able to update better population for objective shrinkage parameter of ridge regression using fitness function Fit (x_i) which is defined above. For all candidates, the parameter with better fitness is selected for new population. That is, the current parent is replaced with its offspring if the fitness of the offspring is better, otherwise the parent is gone to the next generation.

(4: Convergence step)

Repeat until convergence

The convergence criteria of our DE-ridge are similar to other evolutionary algorithms (Eiben and Smith, 2003). DE-ridge convergence is reached when the maximum number of generation is over and the fitness value does not change significantly.

RESULTS

We verify improved performances of our DE-ridge using data sets form objective machine learning data and syn-thetic data. By the evaluative measures of ridge trace, repeated method to convergence, Prediction sum of Squares (PRESS), Generalized Crossvalidation (GCV), VIF (variance inflation factor), Hoerl-Kennard-Baldwin (HKB) estimator (Myers, 1990; Zhang and Harrath, 2006). The following table shows summary information of our experimental data sets.

California is California housing data set from the StatLib repository in the Table 1 (http://lib.stat.cmu.edu/datasets/). From the DELVE reposi-tory, kinematics data is concerned with the forward kinematics of 8 link robot arm (http://www.cs.toronto.edu/~delve/). Last objective data set, machine is machine cpu which is relative cpu performance data from UCI machine learning repository (http://mlearn.ics.uci.edu/MLRepository.html). We know the multicollinearity information of the data by VIF value in the above table. So, we find that California and syn1 data sets have multicollinearity. In our experiment, syn1, syn2 and syn3 are simulation data sets with different correlation coefficient as the following figure.

In the above Fig. 1 a and b represent high and low correlated. Fig. 1c shows independency between input variables. We know visual information of experimental data sets in the following figure.

Figure 2 a-c scatter plot matrix of california, kinematics and machine data respectively. Also, the plots of synthetic data sets according to correlation coefficients are shown in Fig. 2 d-f . Our experimental results are shown in Table 2. We compare DE-ridge with original ridge regression by MSE (Mean Squared Error) (Hajkins, 1999; Mitchell, 1997; Vapnik, 1998).

According to the results, we know the improved performance of DE-ridge. Because MSE values of DE-

Table 1: Summary of data sets

Data	# of points	# of variables	VIF
California	20640	8	36.0890
Kinematics	8192	8	1.0013
Machine	209	6	3.2740
Syn1	1000	3	7.1688
Syn2	1000	3	1.4605
Syn3	1000	3	1.0015

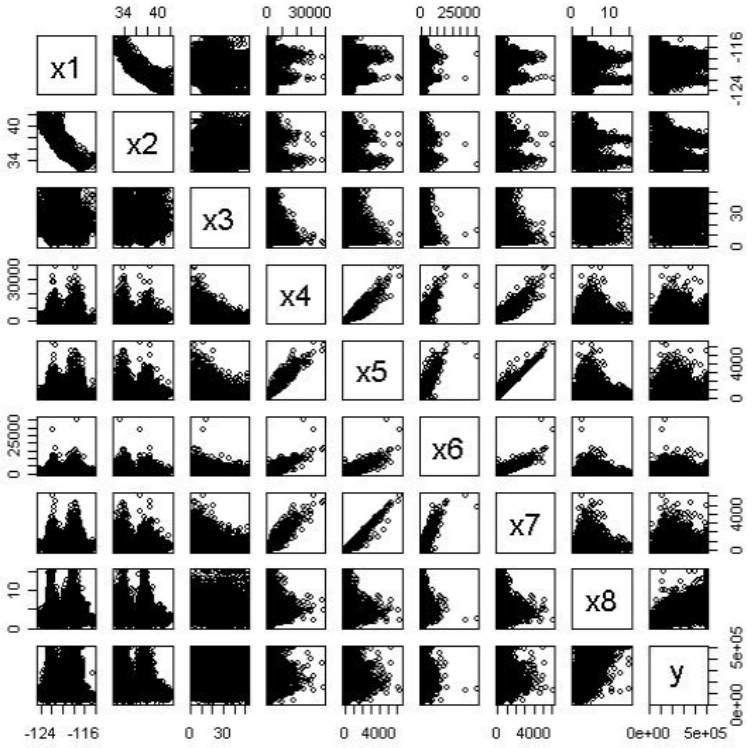
Table 2: Comparison results ridge and DE-ridge regressions

Data	Regression type	Parameter	MSE
California	Ridge	2.20	0.398317
	DE-ridge	19.87	0.398229
Kinematics	Ridge	10.90	0.040900
	DE-ridge	0.05	0.040894
Machine	Ridge	2.47	0.368413
	DE-ridge	99.98	0.303627
Syn1	Ridge	20.06	1.151916
	DE-ridge	0.02	1.151513
Syn2	Ridge	13.38	0.973950
	DE-ridge	61.20	0.969694
Syn2	Ridge	7.38	1.015627
	DE-ridge	28.50	1.015298

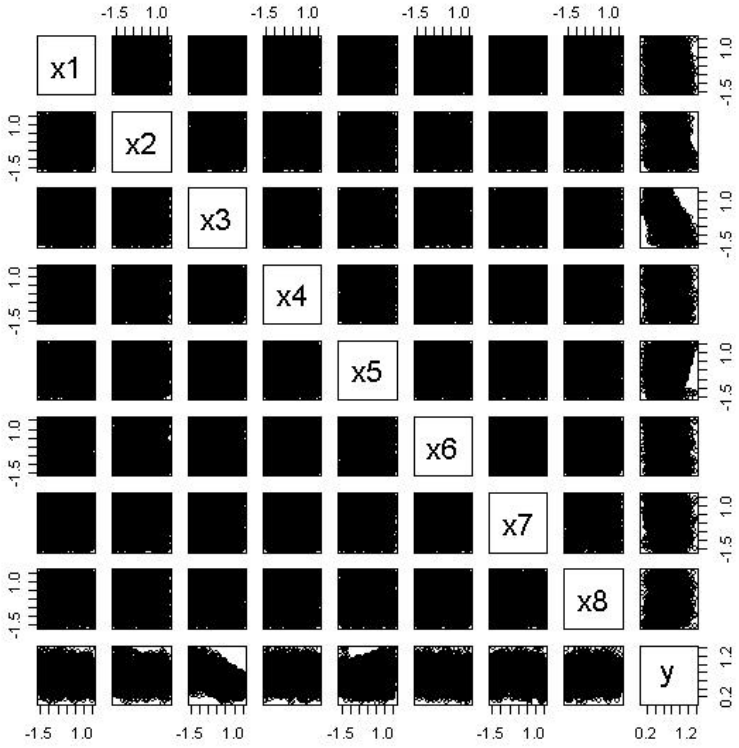
$$\begin{pmatrix} 1 & 0.9 & 0.7 \\ 0.9 & 1 & 0.8 \\ 0.7 & 0.8 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 0.3 & 0.4 \\ 0.3 & 1 & 0.5 \\ 0.4 & 0.5 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

(a) (b) (c)

Fig. 1: Correlation coefficient of synthetic data

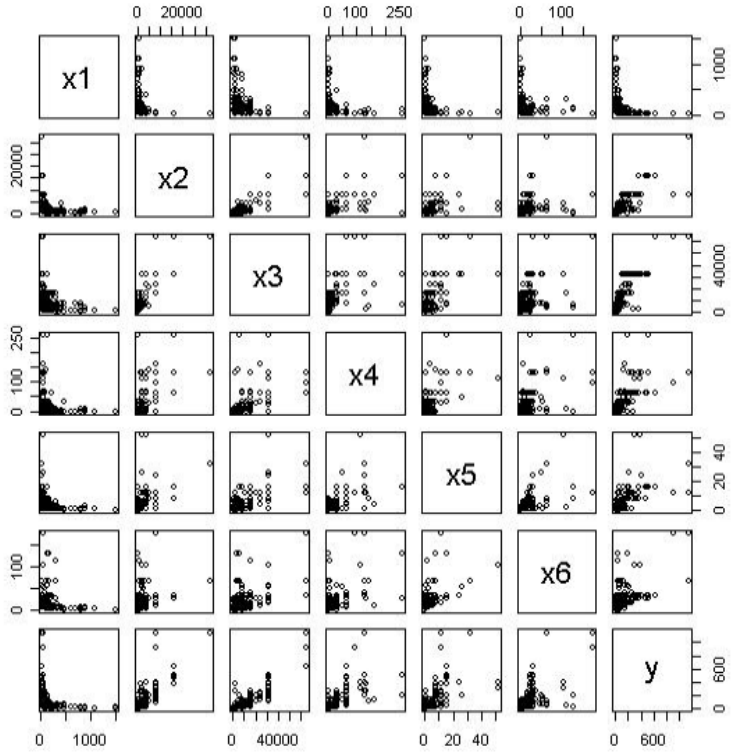


(a)

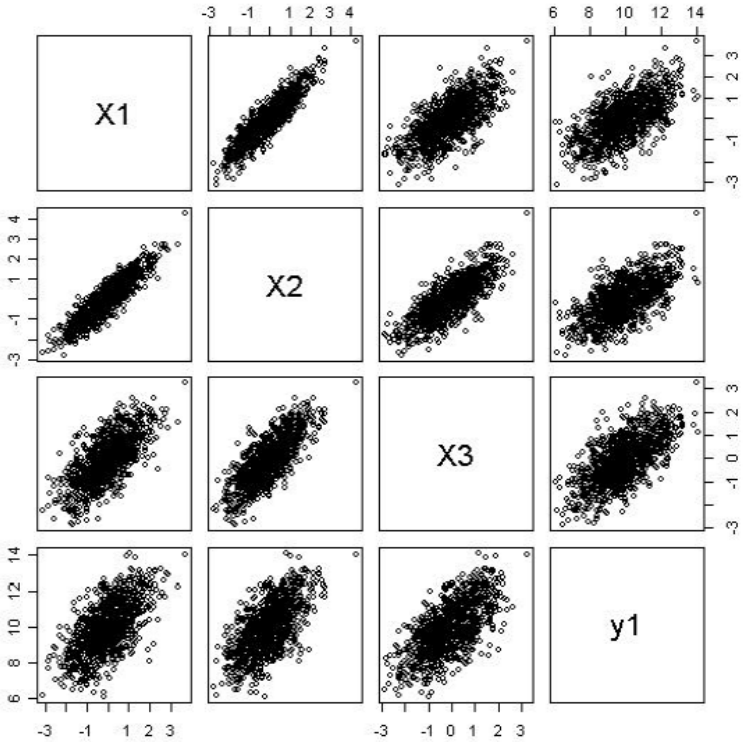


(b)

Fig. 1: Continue

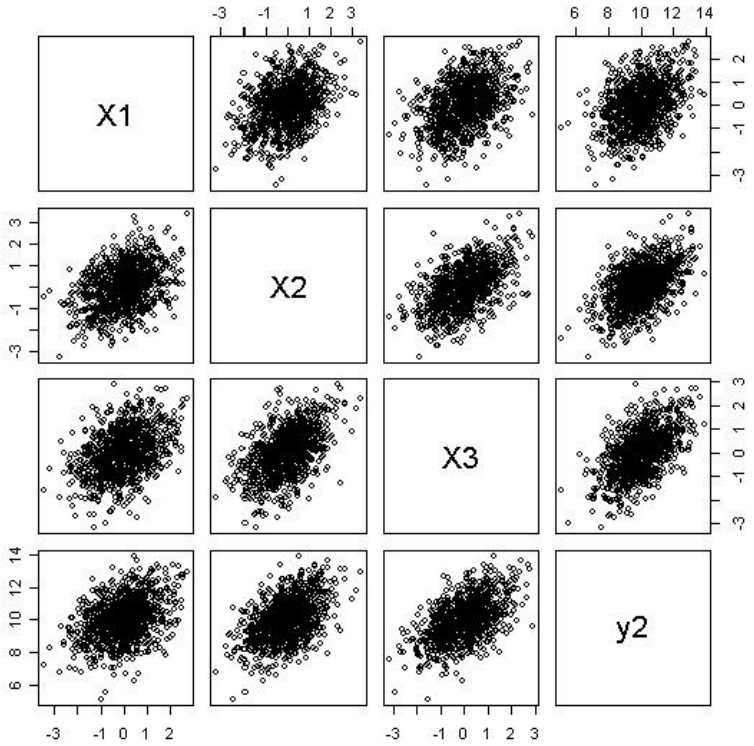


(c)

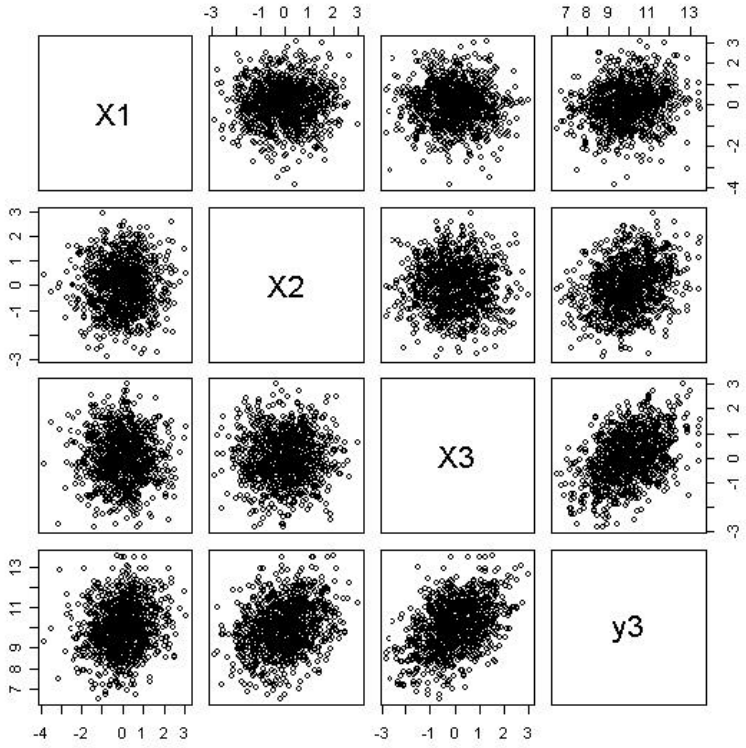


(d)

Fig. 1: Continue



(e)



(f)

Fig. 2: Visual information of data sets

ridge in above six data sets are smaller than original ridge regression. So, we are able to verify our research.

CONCLUSION

In this study, we proposed DE-ridge model to settle the problem of ridge regression. Using DE, the shrinkage parameter of ridge regression was able to be determined objectively. In addition, we knew improved performance of DE-ridge by experimental results using six data sets. In future works, we will apply DE for objective determination of the parameters in diverse machine learning algorithms.

REFERENCES

Data for Evaluating Learning in Valid Experiments, http://www.cs.toronto.edu/~delve/

Eiben, A.E. and J.E. Smith, 2003. Introduction to Evolutionary Computing, Springer.

Embrechts, M.J., 2004. Direct kernel least-squares support vector machines with heuristic regularization. Proc. IEEE. Int. Joint Conf. Neural Networks, 1: 687-692.

Engelbrecht, A.P., 2002. Computational Intelligence An Introduction, Wiley.

Hastie, T., R. Tibshirani and J. Friedman, 2001. The Elements of Statistical Learning, Springer.

Haykin, S., 1999. Neural Networks, Prentice Hall.

Hines, J.W., A.V. Gribok, A.M. Urmanov and M.A. Buckner, 2002. Selection of Multiple Regularization Parameters in Local Ridge Regression Using Evolutionary Algorithms and Prediction Risk Optimization. Proceedings of 4th Interna-tional Conference of Inverse Problems In Engineering.

Hoerl, A.E. and R.W. Kennard, 1970a. Ridge Regression: biased estimation for nonorthogonal problems, Technometrics, 12: 55-67.

Hoerl, A.E. and R.W. Kennard, 1970b. Ridge Regression: application to nonorthogonal problems. Technometrics, 12: 69-82.

Huet, S., A. Bouvier, M.A. Poursat and E. Jolivet, 2003. Statistical Tools for Nonlinear Regression, Springer Series in Statistics, Springer.

Ismail, M.Y. and J.C. Principe, 1996. Equivalence between RLS algorithms and the ridge regression technique, Signals. Proc. IEEE. Conf. Rec. Thirtieth Asilomar, 2: 1083-1087.

Mitchell, T.M., 1997. Machine Learning, McGraw-Hill.

Myers, R.H., 1990. Classical and Modern Regression with Applications, Duxbury Press.

Olague, G., B. Hernandez and E. Dunn, 2003. Hybrid Evolutionary Ridge Regression Approach for High-Accurate Corner Extraction. Proc. IEEE. Compu. Soc. Conf. Computer Vision and Pattern Recog., pp: 1-6.

Price, K.V., R. Storn and J. Lampinen, 2005. Differential Evolution-a practical approach to global optimization, Springer.

Ronkkonen, J., S. Kukkonen and K.V. Price, 2005. Real-Parameter Optimization with Differential Evolution. Proc. IEEE. Congress on Evolut. Comp., 1: 506-513.

StatLib---Datasets Archive, http://lib.stat.cmu.edu/datasets/

Storn, R. and K.V. Price, 1997. Differential Evolution-a fast and efficient heuristic for global optimization over continuous spaces. J. Global Optimiz., 11: 341-359.

UCI Machine Learning Repository, http://mlearn.ics.uci.edu/MLRepository.html.

Vapnik, V.Z., 1998. Statistical Learning Theory, John Wiley and Sons, Inc.

Zhang, B. and S. Horvath, 2006. Ridge regression based hybrid genetic algorithms for multi-locus quantitative trait map-ping. International Journal of Bioinformatics Research and Application, Vol. 1.