Research Article

ISSN 1112-9867

Available online at http://www.jfas.info

COMPARISION OF THE ESTIMATION OF THE LEAST SQUARE AND GENETIC ALGORITHM THROUTH SERVAL FUNCTION IN R

N. Abbasi*, L. Nazari Sabooki

Department of Statistics, Payame Noor University, I. R. Iran

Published online: 25 June 2016

ABSTRACT

R software is considered software in which various available functions make it possible to conduct extensive statistical calculations. Genetic algorithm is a method through which we search for approximate solution in optimization. This article aims at evaluating the functions existed in the R software which are employed for approximate solution in optimization. As a result, there has been produced least squares by usual methods for linear and non-linear models through genetic algorithm in this research.

Keywords: Least Squares, Genetic Algorithm, Linear Model, Nonlinear Model, Squares Average.

Author Correspondence, e-mail: c.author@mail.com
doi: http://dx.doi.org/10.4314/jfas.8vi2s.88

1. INTRODUCTION

The R software and its package conducts a broad range of statistical and graph models such as linear and non-linear models, classic statistical tests, analysis of time series, clustering, etc. The R software has been simply developed through the expansion of functions presented as packages. Most of the standard functions of the R software have been written according to a certain algorithm in the software which makes it easier to employ for its users.

1.1. Least Squares

Least squares are defined as a set which minimizes the erroneous squares produced in every equation to aim at fixing the parameters of a function to an optimal model to fit the data. The data consist of $n$points $(x_i\ ,y_i)$, $i = 1, 2, \ldots , n$ in which x_i is the independent variable and y_i is the dependent variable which their quantities are obtained through observation. The function model is in the form of $f(x,\beta)$ in which the adjustable m parameters are located in the β vector. The aim is to find the best parameter quantities to fit the data for the model. In the least ideal squares, the sum of the remaining squares is the least.

$$S = \sum_{i=1}^{n} r_i^2$$

The meaning of the remaining squares is the difference between the observed data with the quantity obtained from the model.

$$r_i = y_i - f(x_i, \beta)$$

This is an example of straight linear model in two dimensions, with the width from the origin β_0 and the slope β_i. Its function appears to be as $f(x,\beta) = \beta_0 + \beta_1 x$.

The data point may consist of more than one independent variable. In more general cases, there might be one or more independent variables or one or more dependent variables in every data point.

1.2. Genetic Algorithm

In genetic algorithm, the population of candidate solution (individual names, creatures or phenotype) related to optimization has evolved toward better solutions. Each individual from the population has sets of attribute (chromosomes or the effect of genotype) than can mutate and alternate. Tradihonally, an individual in binary is shown a string of zeros and one, however, other coding is also possible. Usually evolution is derived from a population randomly and the process of recurrence in the population in every repetition is called generation. In every generation the ability to live of each individual is usually the extent of the target function in the solvation of optimization. Suitable people are randomly selected from the present population and the genome of that modified person (combination and probably accidental mutation) and the new generation is thus produced this new generation is the candidate solution which is used in the next repetition of this algorithm normally, the algorithm is finished when it has reproduced the maximum number of generation with the ability to live has reached a satisfactory level for the population

1.3. The purpose of the article

The problem of finding the least estimator for erroneous square is directly calculated by the available formulas and the technical texts.

However these models are such that it is impossible to indicate the final formula for estimators so, we have recourse to number analysis. This will lead to find the optimized estimator. To examine the power of this algorithm, first we study the linear models in which linear estimator are defined. This comparison is done through two ga and lm function that appeared in details in part 2. According to that we will see in part 3, we will consider a non-linear model and we conduct the comparison by the two $ga$and nls.

2. SIMPEL LINEAR MODELS

Consider the following linear model:

$$y_i = 4 + 8x_i + e_i; \qquad i = 1, 2, \dots, n.$$

where for $i = 1, 2, \dots, n, e_i \sim N(0, 25)$. Now, by changing the size of sample, we compare the accuracy of the estimation of the linear model through the two least methods of erroneous squares and genetic algorithm. Certainly, every time that the program runs, the result related to the number changed. In some cases, the genetic algorithm operates better than the least erroneous squares.

Table 1. Results of fitting the simulated model of $y = 4 + 8x$ on the basis of the two methods of least erroneous squares and genetic algorithm

	n	*100*	*90*	*80*	*70*	*60*	*50*	*40*	*30*	*20*	*10*
GA	β_0	6.709	1.203	8.170	8.267	3.561	9.333	3.460	4.384	7.794	3.722
	β_1	7.977	8.067	7.961	8.034	8.044	7.941	7.980	8.037	7.968	8.183
	MSE	**613.1**	**584.6**	**741.8**	**647.0**	**772.7**	**826.8**	**433.8**	**517.2**	**346.1**	**452.3**
		93	**06**	**69**	**52**	**38**	**39**	**84**	**61**	**00**	**30**
LS	β_0	6.728	1.130	8.138	8.423	3.547	9.736	3.462	4.396	7.552	3.712
	β_1	7.977	8.067	7.962	8.032	8.044	7.934	7.980	8.036	7.971	8.183
	MSE	**613.1**	**584.5**	**741.8**	**647.0**	**772.7**	**826.7**	**433.8**	**517.2**	**346.0**	**452.3**
		93	**95**	**66**	**46**	**38**	**96**	**84**	**61**	**83**	**30**

In this part, five outlayer data are added to the data set of table 1. The results will be considered again in table 2.

Table 2. Results of fitting the simulated model of $y = 4 + 8x$ in the presence of five outlayer data based on the two methods of least erroneous square and genetic algorithm

	n	*100*	*90*	*80*	*70*	*60*	*50*	*40*	*30*	*20*	*10*
GA	β_0	9.864	9.744	9.928	9.924	9.991	9.901	9.761	9.662	9.989	9.944
	β_1	7.863	7.942	7.949	7.926	7.900	7.984	7.935	7.825	8.060	8.265
	MSE	**1091**	**1187**	**13274.**	**1517**	**1799**	**2117**	**2654**	**3595**	**5537**	**12313**
		3.662	**5.803**	**14658**	**0.862**	**3.078**	**4.646**	**5.801**	**6.949**	**8.418**	**1.100**
LS	β_0	38.29	37.56	47.056	57.11	69.04	63.70	91.93	92.53	128.7	209.30
		0	8		1	9	3	3	5	90	1
	β_1	7.427	7.518	7.386	7.226	7.044	7.176	6.710	6.585	6.254	5.126
	MSE	**1069**	**1166**	**12895.**	**1454**	**1699**	**2032**	**2446**	**3368**	**5001**	**10120**
		6.870	**6.050**	**898**	**9.477**	**4.368**	**3.321**	**7.317**	**3.221**	**6.420**	**4.200**

3. NON-LINEAR MODEL

Consider the following nonlinear model:

$$y_i = \frac{25}{1 + \exp(-2\, x_i)} + e_i; \qquad i = 1, 2, \dots, n$$

where for $i = 1, 2, \ldots, n,\ e_i \sim N(0{,}25)$. We will change the size of the sample as in table 3. In all the cause, the least square method has operated better in non-linear method than the genetic method. Table 4 is the study of non-linear model with five outlayer data.

Table 3: The result of fitting of the simulated model $25/\mathrm{y} = 1 + \exp(-2\mathrm{x}) + \mathrm{e}$ on the basis of the two, least erroneous square and genetic algorithm methods.

	n	*100*	*90*	*80*	*70*	*60*	*50*	*40*	*30*	*20*	*10*
GA	β_0	24.146	25.007	24.893	24.331	23.249	24.434	23.940	23.579	23.479	23.891
	β_1	-2.490	-2.0363	-2.085	-3.135	-3.668	-2.261	-2.494	-4.173	-3.150	-2.123
	MSE	**1.619**	**0.821**	**1.219**	**2.426**	**2.942**	**1.134**	**1.565**	**3.371**	**1.837**	**1.736**
LS	β_0	25.187	25.291	24.996	25.497	25.107	24.696	25.082	25.151	25.094	24.366
	β_1	-1.900	-1.882	-2.024	-1.829	-1.981	-2.118	-2.040	-1.950	-1.998	-2.149
	MSE	**0.069**	**0.0003**	**0.026**	**5.e-05**	**0.017**	**0.008**	**0.010**	**0.0133**	**0.025**	**0.005**

Table 4: The result of fitting of the simulated model $\mathbf{25/y = 1 + exp(-2x) + e}$ on the basis of the two, least erroneous squares and genetic algorithm methods, with the presence of five outlayer data.

	n	*100*	*90*	*80*	*70*	*60*	*50*	*40*	*30*	*20*	*10*
GA	β_0	23.934	23.197	24.491	21.872	22.373	24.272	24.861	23.178	21.170	25.604
	β_1	-2.546	-2.985	-1.850	-3.380	-2.742	-1.879	-1.355	-1.920	-2.269	-0.701
	MSE	**7.039**	**7.944**	**7.932**	**11.585**	**11.107**	**11.653**	**13.057**	**16.910**	**22.092**	**29.483**
LS	β_0	25.003	24.640	24.611	24.635	24.040	25.065	25.070	24.248	25.195	25.409
	β_1	-1.676	-1.810	-1.777	-1.685	-1.938	-1.413	-1.311	-1.447	-0.964	-0.705
	MSE	**0.043**	**0.077**	**0.006**	**0.067**	**0.019**	**0.013**	**0.073**	**0.087**	**0.016**	**0.044**

4. CONCLUSION

As we have observed, in the linear model of erroneous square, in the least squares and in genetic algorithm in the linear method, almost the same results have been achieved without the presence of outlayer points. With the presence of 5% of data of outlayer points, however, there is no significant difference between the mean of erroneous square in the method of the least squares and the genetic algorithm, the mean of erroneous squares in the least squares is less which by increasing the data these amounts of means will become less. In non-linear model, without the presence of outlayer data, the mean for erroneous squares in the least squares is less, but with the 5% out layer data, the mean for erroneous squares is more different in the least squares and operates better than genetic algorithm.

5. REFERENCES

[1]. Bretscher, Otto (1995). *Linear Algebra With Applications* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

[2]. Tseng, L. Y. and Yang, S., "Genetic algorithms for clustering, feature selection and classification" ,IEEE Int. Conference on Neural Networks, pp.1612-1616,1997.

[3]. Vafaie, H. and Imam, I., "Feature selection methods: genetic algorithms vs. greedy-like search". Proc. of the Int. conference on fuzzy and intelligent control systems, 1994.

[4]. Fogel, D.B. " What is Evolutionary Computation?" IEEE Spectrum, Feb 2000, pp. 26-32

[5]. Bala, J., Huary, J., Vafaie, H., De Jong, K. and Wechslev, H. "Hybrid learning using genetic algorithms and decision trees for pattern classification" , IJCAI conference , Montreal , August 19-25,1995

[6]. Aggarwal, C. C., Yu, S. P., "An effective and efficient algorithm for high-dimensional outlier detection, The VLDB Journal, 2005, vol. 14, pp. 211–221.

How to cite this article:

Abbasi N and Nazari Sabooki L. Comparision of the estimation of the least square and genetic algorithm throuth serval function in R. J. Fundam. Appl. Sci., 2016, *8(2S), 1100-1105.*