Applied Medical Informatics
Original Research Vol. 30, No. 2 /2012, pp: 1-6

Exploring the 3D Surfaces with Modified Method of Steepest Descent

Wioletta GRZENDA[1], Maciej GORKIEWICZ[2]

[1] Warsaw School of Economics, Institute of Statistics and Demography, ul. Madalińskiego 6/8, 02-513 Warszawa, Poland.
[2] Higher School of Social and Technical Sciences in Radom, ul. Wodna 13/21, 26-600 Radom, Poland.
E-mails: Wioletta.grzenda@sgh.waw.pl; gorkiewicz@poczta.fm

* Author to whom correspondence should be addressed; Tel.: +48 22 564 92 70; Fax: (22) 564 86 43.

Received: 6 February 2012 / Accepted: 20 May 2012 / Published online: 13 June 2012

Abstract
Aim: To prove expediency of the steepest descent method to divide a given cloud of (Y, X1, X2) points into the spatial clusters with purpose to estimate a simple regression model Y = f(Z|X1,X2) at each cluster. *Material and Method*: The exemplary data sets {Y, X1, X2} were drawn randomly from assumed 3D surface: Y = f(X1,X2), and then a random noise was added to variable Y. A polynomial model Y = f(X1,X2) and a set of models Y = f(Z|X1,X2) were estimated separately, both under Akaike information criterion (AIC), and then compared with respect to their determination coefficients R-square, and the residuals' distributions. *Results*: In the artificial data set studied, the both compared methods after several iterations can provide regression models of the quite similar quality. *Conclusions*: Because the proposed novel method seems to be more robust to outliers, and easier to graphical presentations and to intuitive understanding than the conventional way of building a regression model, the proposed novel method can be recommended to use by non-statisticians, especially in situation when, besides usual moderate noise, the sporadic but influential measurement errors can occur.

Keywords: Steepest descent method; k-NN; Spatial clusters; Regression; Interactive estimation.

Introduction

The various models of 3D surfaces are in everyday use at almost all practical domains, usually in the form of ready to use maps, like disease maps for epidemiologists and the other public health professionals [1]. Therefore, numerous analytic tools suitable to create a map by an experienced statistician, basing on a set of an available data, have recently been developed [1,2].

Alas, in situation when some non-statistician is looking for a rapid answer to some practical question, these tools have some worrying inadequacies, associated mainly with insufficient emphasis given by designers to practical troubleshooting, that is to data validation before analyses, recognizing routine pitfalls and critical traps during processing, and validation of results after processing. Even a very clear main idea of an estimation procedure, and a perfect methodological guide, both can be perceived as incomprehensible because of specific statistic terminology, and due to lack of framework to parallel comparative computations with exemplary data [3]. Then, the non-statisticians ordinarily perceived all process of estimation as generally unmanageable, not only due

to troubles with statistical terminology, but also as a result of the frequent silent assumptions [3,4], and inadequate use of graphics [5,6].

This paper provides an alternative strategy for the analysis of a set of $\Omega = \{yi, x1.i, x2.i\}$ data, i = 1, 2, …, N, when the primary focus of interest is not only on predicting the mean values mean(Y|(X1, X2)) for freely chosen point (X1, X2), as at [1] and [2], but also on predicting the extreme gradients max(ΔY|D1.2=1) and min(ΔY|D1.2=1); where: Y is a variable under interest; X1 and X2 can be interpreted as geographical coordinates, and D1.2 can be interpreted as Euclidean distance from point (X1, X2) at the X1X2 plane.

Our approach can be summarized with three basic ideas: (i) divide compound surface into some simple parts, all with comprehensible features; (ii) apply interactive method with adequate visual output that enables researcher to select on-screen specific subsets of data and identify their characteristics, or to isolate outliers; (iii) allow researcher to rely only on his own experience and intuition, but give him opportunity to apply recognized formal criterions and statistical procedures at each step of analyses.

The first phase of the proposed procedure was aimed to separate at a data set Ω the clusters of three kinds: first kind of the flat clusters ΩF, included triplets of neighbouring points with small gradients: (ΔY|D1.2=1) ≈ 0; second kind of the sloped clusters ΩS, included points of sensible positive gradients (ΔY|(ΔD)), under restriction that every sloped cluster can be considered as a connected graph, what correspond to known steepest descent method, [7]; and the third kind of the abnormal clusters ΩA, included isolated points and points of anomalous gradients.

At the start an all data set Ω is considered as subset of candidates. During process of defining clusters the candidates are step-by-step removed from this subset to active clusters ΩF or ΩS or ΩA. Candidates are enrolled into consideration with use of known k-NN procedure [8], as nearest neighbours to active clusters ΩF or ΩS. In case of need, candidates can be enrolled randomly.

During clustering some neighbouring clusters can join one to other naturally, accordingly to logic of the procedure. In case of need, the procedures described at [1,2] can be applied to force a further aggregation of the flat clusters, and procedures like [9] to force joining of the sloped clusters.

The second phase of the proposed procedure was aimed to characterize every sloped cluster with its own only bivariate regression model Y = Yj + bj*fj(Zj); where: j – index of a cluster, j = 1, 2, …, J; Yj - estimated maximal value of Y at j-th cluster; Zj – a measure of the distance at a j-th cluster from point (X1,X2)|Yj; Yj, bj – coefficient of regressions; fj(Zj) – monotone, at least quasi-monotone function. Finally, it should be make an attempt aimed to join separate regression models into a single common model for all ΩF and ΩS clusters.

There are several known methods aimed to define a measure of the distances Zj. The simple way, at the frames of the rank-based methods, put Zj = 0 for a (X1,X2)|Yj; then, for any pair of two adjacent knots at an ordered graph, a ΔZj = 1 [10]. The other simple way to define distance at an ordered graph is to sum up all distances Dk.k-1 along a path from considered k-th knot to knot of Zj = 0. We suggest other somewhat more complicated method, corresponding to ideas of principal components approach and to joint nonparametric transformation to linearity, [11]. The proposed method based on assumption that each sloped cluster under consideration can be reduced on a X1X2 plane to a single line of a general local descent; with adequate arrangement of the projections of this cluster points on this line. Subsequently, distances Z are measured as length of a path along this line. Finally, the additive regression models can be easy transformed to simple linear regression models, due to monotonicity of the fj(Zj) function, [12].

In the present preliminary study the all problems with linearization were omitted because the exemplary data sets of N = 100 points (Y, X1, X2) were drawn randomly from assumed a single flat cluster ΩF: Y = 0, and from four sloped clusters of the linear shapes either Y ~ -|X1 – 20| or Y ~ -|X2 – 20|, and then a random noise was added to values of variable Y.

The illustrative data sets were explored twice, first time with classical procedures for selecting a structure of the polynomial regression, and then independently, with the proposed method. This way of preliminary confirmation follows to known Popper's falsification principle, [13], that suggests a simple empirical way to reveal an unsuitability of a given model by generating the falsifiable predictions with it.

The method has the advantages of being conceptually simple and much easier to graphical presentations and to intuitive understanding than the conventional way of building a regression model. Therefore, it provides better opportunity for a researcher aimed to utilize his experience and professional knowledge at process of constructing a model, then other know procedures [14].

Material and Method

In the present paper the illustrative artificial data set of N = 100 random data (Y, X1, X2) was investigated twice, first time with classical procedures for selecting a structure of the polynomial regression, and then independently, with the proposed method.

With aim to obtain the illustrative data set the exemplary 3D surface was constructed as follows. The five squares of the size 40*40 were arranged on X1X2 plane in such a way, that four squares named A, B, C and D, stick closely to a central square named flat square. Then the N = 20 points (X1, X2) were drawn randomly from each of above squares separately, assuming uniform distribution, and ideal values of variable Y at the flat square all were put Y = 0; but at the other squares were computed as follows, at the A: Y = -0,5*(X1 – 20); at the B: Y = -(X2 – 20); at the C: Y = +(X2 – 20); and at the D: Y = +2*(X1 – 20). Finally, some random noise was added to above ideal data. The illustrative data set is presented in figure 1.

The polynomial regression models were estimated with classical LSE (least-square-error) procedure and with MLL procedure, both available at SAS software, [15]. The analyses started with polynomial of the forth order (1), and step-by-step chose properly, [16], a single regression coefficient to put it equal to zero up to obtain minimum of Akaike's informative criterion AIC.

$$Y = a + \sum bi*(X1)i + \sum ci*(X2)i \; ; \; \Pi bi*ci \neq 0; \; i = 1, 2, 3, 4; \quad (1)$$

For LSE procedure AIC was defined with formula (2), and for MLL procedure with (3).

$$AIC = N*\ln(SS.res/N) + 2*df.regr. \quad (2)$$

$$AIC = LL + 2*df.regr. \quad (3)$$

where: SS.res – sum of squares of residuals; df.regr – degree of freedom of regression model.

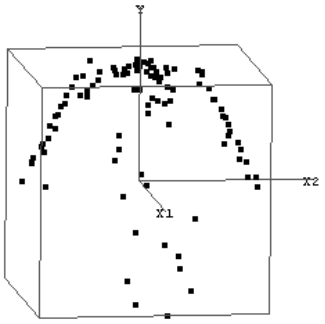

Figure 1. The axonometric view of the investigated data set

The alternative method for the regression estimation was organized as follows. At the first phase the points of the illustrative random points (Y, X1, X2) set were connected into spatial 3D clusters of three kinds, flat clusters, sloped clusters, and anomalies, under general rule, that any cluster can attach only either some separated still point or a point from other cluster of the same kind. At the next phase the spatial 3D clusters were transformed into 2D clusters at a XY space. Then, at each sloped cluster separately, the simple regression Y = f(X) was estimated, and at each flat cluster variable Y was estimated with its mean value at this cluster. At the last phase the common regression model for all flat and sloped clusters were constructed under criterion of minimum AIC (2).

At the first phase the connecting was based on the steepest descent, [7], and k-NN, [8], approaches. The distance D from any point to any other one was calculated as Euclidean distance

at the X1X2 plane. At each a flat or a sloped cluster the active and inactive points were differentiated. In case of need, if anyone cluster with active points didn't exist, then a data point to consideration should be freely chosen from the unconnected points. A chosen point should be included to anomalies either it was more distant from any other admissible point than assumed maximal permissible distance, or if all nearer admissible points had too great estimated gradients G = ΔY/D. Otherwise, it initiated with its neighbouring point a flat or a sloped cluster. At the flat clusters the active points were connected into pairs of neighbouring active points, but at the sloped clusters weren't connected. For any, freely chosen pair of neighbouring active points at a flat cluster as the candidate to attachment the nearest point from outside is considered. If Y at this point was placed inside assumed interval from actual mean value of Y at this flat cluster, say mean(Y)± deltaY, than this candidate point should be attached as a new active point, otherwise a chosen pair of active points should be classified as inactive pair. At the sloped clusters for any freely chosen active point a set of candidates included no more than assumed number of K nearest admissible points from outside, no more distant than assumed maximal permissible distance. For each of candidate a value of a gradient is estimated as G = ΔY/D, and from consideration were excluded candidates with either too great G: |G| > Gmax; or too little G: |G| < Gmin. From the remaining candidates, if any, the candidates with extreme gradients were attached to a cluster as the new active points. In any case, a previously active point turn into inactive one. During this phase a researcher can freely change all parameters: number of nearer neighbours, limits deltaY, Gmax and Gmin.

At the next phase the mean direction tg(μ) of descent on Y at the X1X2 plane was estimated for each sloped cluster separately, under criterion of maximal sum of projection of gradients. With this aim for each pair of the connected points at a considered cluster the direction tg(φ) = ΔX2/ΔX1 was calculated, and projection Gp of its gradient G = ΔY/D was estimated as Gp = G*cos(φ - μ). The Akaike's AIC (2) can be used to estimate common direction tg(μ) for several clusters. A new axis Z for each estimated direction tg(μ) was located through origin (X1 = 0; X2 = 0), and each point of a sloped cluster was projected on its axis Z. The points of a flat cluster can be projected on a freely chosen axis Z. The points classified as anomalies were excluded from further analyses. Finally, the initial data sample (Y, X1, X2) took form (Y, Z).

At the next phase a regression model Y = f(Z) was estimated as usual for each sloped cluster separately. The Akaike's AIC (2) can be used to estimate common regression model Y = f(Z) for several clusters. With this aim, in case of linear models of the form: Y = a + b*Z, it first it should be proved if it is admissible to put equal intercepts and equal slopes at some clusters; then, to put equal slopes; and finely, to put intercepts equal to zero for several clusters, with respect to proper new origin of Z and Y axis.

Results

Results of classical LSE and MLL polynomial regression, see Table.1

Table 1. Results of classical LSE and MLL polynomial regression

df.reg	AIC|LSE	AIC|MLL	P(next=true)	Estimated model
8	94.52	380.31	0.71	$Y = a + b_{11}X_1+b_{12}X_1^2+b_{13}X_1^3+b_{14}X_1^4 + b_{21}X_2+b_{22}X_2^2+b_{23}X_2^3+b_{24}X_2^4$
7	92.71	378.50	0.71	$Y = a + 0 +b_{12}X_1^2+b_{13}X_1^3+b_{14}X_1^4 + b_{21}X_2+b_{22}X_2^2+b_{23}X_2^3+b_{24}X_2^4$
6	90.96	376.74	0.73	$Y = a + 0 +b_{12}X_1^2+b_{13}X_1^3+b_{14}X_1^4 + 0 +b_{22}X_2^2+b_{23}X_2^3+b_{24}X_2^4$
5	88.98	374.77	0.10	$Y = a + 0 +b_{12}X_1^2+b_{13}X_1^3+b_{14}X_1^4 + 0 +b_{22}X_2^2+ 0 +b_{24}X_2^4$
4	93.36	379.15	#	$Y = a + 0 +b_{12}X_1^2+b_{13}X_1^3+ 0 + 0 +b_{22}X_2^2+ 0 +b_{24}X_2^4$

For the initial model with 9 regression parameters we received $R^2 = 0.9760$ and adjusted $R^2 = 0.9739$

The most informative model has only 6 regression parameters $R^2 = 0.9759 < 0.9760$ but adjusted $R^2 = 0.9746 > 0.9739$.

Table 2. Descriptive statistics for data (Y, Z) divided into five clusters

Parameter	Flat	A	B	C	D	all
N	17	20	21	22	20	100
mean Y	0.02	-20.86	-9.89	-9.42	-5.15	-9.35
SD	0.15	11.82	6.20	6.40	2.98	9.51
R(Y,Z)	-0.19	-0.98	-0.98	-0.97	-0.98	-0.73
p(R(Y,Z))	0.46	0.0000	0.0000	0.0000	0.0000	0.0000

N = number of observations; SD = standard deviation; R(Y,Z) = correlation coefficient; p(R(Y,Z)) = significance of hypothesis: R(Y,Z) = 0.

Table 3. Results of building common regression model under Akaike's informative criterion AIC

Df.reg	Clusters	AIC	P(next=true)	Model
9	5	76.66	65%	$Y = Y_{flat} + DV_i*(a_i + b_i*Z)$; i = A, B, C, D
7	4	75.39	60%	$Y = Y_{flat} + DV_i*(a_i + b_i*Z)$; i = A, BC, D
6	4	74.55	0.0%	$Y = Y_{flat} - Y_0 + DV_i*b_i*(Z - Z_0)$; i = A, BC, D
5	3	111.4	#	$Y = Y_{flat} - Y_0 + DV_i*b_i*(Z - Z_0)$; i = AD, BC

where: Y0, Z0 – coordinates on new origin of Y and Z axis

Discussion

The illustrative data sets were investigated twice, first time with classical procedures for selecting a structure of the polynomial regression, and then independently, with the proposed method. This way of preliminary confirmation follows to known Popper's falsification principle [13], that suggests a simple empirical way to reveal an unsuitability of a given model by generating the falsifiable predictions with it.

The method has the advantages of being conceptually simple and much easier to graphical presentations and to intuitive understanding than the conventional way of building a regression model. Therefore, it provides better opportunity for a researcher aimed to utilize his experience and professional knowledge at process of constructing a model, then other know procedures [14].

It should be noted that in case if the spatial clusters objectively exist, then conventional brushing technique provides good opportunity to detect them on the base of simple visual inspection of the plots of residuals (i.e. regression errors) versus considered variables, for good example with proper explanations one can see [17].

It should be noted that all process of regression model estimation can be totally automated [18], even with respect to including into process some experts opinions [14], and procedures of defining spatial clusters with k-NN technique [19].

Conclusions

Because the proposed novel method seems to be more robust to outliers, and easier to graphical presentations and to intuitive understanding than the conventional way of building a regression model, the proposed novel method can be recommended to use by non-statisticians, especially in situation when, besides usual moderate noise, the sporadic but influential measurement errors can occur.

Ethical Issues

No clinical or animal research has been conducted as a part of this research.

Conflict of Interest

The authors declare that they have no conflict of interest.

Acknowledgements

Cross-border Biostatistics Meeting (4-5.11. 2011), Szeged (Hungary).

References

1. Osnes K. Iterative Random Aggregation of Small Units Using Regional Measures of Spatial Autocorrelation for Cluster Localization. Statistics in Medicine 1999;18(6):707-725.
2. Lawson AB. Statistical Methods in Spatial Epidemiology. New York: John Wiley; 2001.
3. Kaplan DT. Statistical Modeling, a Fresh Approach. Scotts Valley: CreateSpace; 2009.
4. Hélie S. An Introduction to Model Selection: Tools and Algorithms. Tutorials in Quantitative Methods for Psychology 2006;2(1):1-10.
5. Hamilton LC. Regression with graphics. Belmont: Wadsworth; 1992.
6. Gelman A. Why Tables Are Really Much Better Than Graphs. Journal of Computational and Graphical Statistics 2011;20(1):3-7.
7. Kuznetsov YA. Steepest descent. Encyclopedia of Mathematics [Internet]. 2011 [updated 2011 Feb 07; cited 2011 Dec 08]. Available from: URL: http://www.encyclopediaofmath.org/index.php?title=Steepest_descent,_method_of&oldid=13529
8. Bolboaca SD. Assessment of Random Assignment in Training and Test Sets using Generalized Cluster Analysis Technique. Applied Medical Informatics 2011;2:9-14.
9. Gaffney S, Smyth P. Joint Probabilistic Curve Clustering and Alignment. New York: Advances in Neural Information Processing Systems; 2004.
10. Kovac A, Smith A. Nonparametric Regression on a Graph. Journal of Computational and Graphical Statistics 2011;20(2):432-447.
11. Górkiewicz M. Estimation of measurement error using local sampling and joint nonparametric linearisation. Acta Un. Lodziensis. Folia Oeconomica 2002;162:99-108.
12. Linton OB, Chen R, Wang N, Härdle W. An Analysis of Transformations for Additive Nonparametric Regression. J Amer Statist Assoc 1997;92:1512-1521.
13. Popper K. The Logic of Scientific Discovery. New York: Routledge, 1959/2004.
14. Onishi H. A Knowledge-Based Variable Selection Method for Box-Cox Transformation. J Japan Statist Soc 2002;32(1):15-42.
15. SAS Product Documentation [Internet]. 2011 [updated 2011 Apr 07; cited 2011 Dec 08]. Available from: URL: http://support.sas.com/documentation/index.html
16. Motulsky H, Christopoulos A. (2004). Fitting Models to Biological Data Using Linear and Nonlinear Regression. A practical guide to curve fitting. Oxford: Oxford University Press, 2004. Also available on-line: URL: http://www.graphpad.com/manuals/prism4/RegressionBook.pdf
17. Kuiper S. Introduction to Multiple Regression: How Much Is Your Car Worth? Journal of Statistics Education, 2008, 16(3) [Internet]. 2008 [updated 2008 May 30; cited 2011 Dec 08]. Available from: URL: www.amstat.org/publications/jse/v16n3/kuiper.html
18. AutoFit. The Web's Multiple Regression Home Page [Internet]. 2011 [updated 2011 Feb 01; cited 2011 Dec 08]. Available from: URL: http://www.eskimo.com/~brainy/
19. Karegowda AG, Jayaram MA, A.S. Manjunath AS. Combining Akaike's Information Criterion (AIC) and the Golden-Section Search Technique to find Optimal Numbers of K-Nearest Neighbors. International Journal of Computer Applications 2010;2(1):80-87.