Exploring the 3D Surfaces with Modified Method of Steepest Descent

Wioletta GRZENDA¹, Maciej GORKIEWICZ²

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 \((Y, X_1, X_2)\) points into the spatial clusters with purpose to estimate a simple regression model \(Y = f(Z|X_1,X_2)\) at each cluster. Material and Method: The exemplary data sets \({Y, X_1, X_2}\) were drawn randomly from assumed 3D surface: \(Y = f(X_1,X_2)\), and then a random noise was added to variable \(Y\). A polynomial model \(Y = f(X_1,X_2)\) and a set of models \(Y = f(Z|X_1,X_2)\) were estimated separately, both under Akaike information criterion (AIC), and then compared with respect to their determination coefficients \(R^2\), 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 $\Omega = \{y_i, x1.i, x2.i\}$ data, $i = 1, 2, \ldots, N$, when the primary focus of interest is not only on predicting the mean values $\text{mean}(Y|(X1, X2))$ for freely chosen point $(X1, X2)$, as at [1] and [2], but also on predicting the extreme gradients $\max(\Delta Y|D1.2=1)$ and $\min(\Delta Y|D1.2=1)$; 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 $\Omega$ the clusters of three kinds: first kind of the flat clusters $\Omega_F$, included triplets of neighbouring points with small gradients: $(\Delta Y|D1.2=1) \approx 0$; second kind of the sloped clusters $\Omega_S$, included points of sensible positive gradients $(\Delta Y|\Delta D)$, 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 $\Omega_A$, included isolated points and points of anomalous gradients.

At the start an all data set $\Omega$ is considered as subset of candidates. During process of defining clusters the candidates are step-by-step removed from this subset to active clusters $\Omega_F$ or $\Omega_S$ or $\Omega_A$. Candidates are enrolled into consideration with use of known k-NN procedure [8], as nearest neighbours to active clusters $\Omega_F$ or $\Omega_S$. 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 $Y = Y_j + bj*fj(Zj)$; where: $j$ – index of a cluster, $j = 1, 2, \ldots, J$; $Y_j$ – estimated maximal value of $Y$ at $j$-th cluster; $Zj$ – a measure of the distance at a $j$-th cluster from point $(X1,X2)|Y_j$; $Y_j$, $bj$ – 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 $\Omega_F$ and $\Omega_S$ 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 $(X1,X2)|Y_j$; then, for any pair of two adjacent knots at an ordered graph, a $\Delta Zj = 1$ [10]. The other 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 $\Omega_F$: $Y = 0$, and from four sloped clusters of the linear shapes either $Y \sim |X1 – 20|$ or $Y \sim |X2 – 20|$, 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.
Exploring the 3D Surfaces with Modified Method of Steepest Descent

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.

\[
Y = a + \sum b_i (X1)^i + \sum c_i (X2)^i ; \quad i = 1, 2, 3, 4; \quad \Pi b_i c_i \neq 0;
\]  
(1)

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

\[
AIC = N*ln(SS.res/N) + 2*df.regr. \quad \text{(2)}
\]

\[
AIC = LL + 2*df.regr. \quad \text{(3)}
\]

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](image)

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 \( G = \Delta Y/D \). 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 included to anomalies either it was more distant from any other admissible point than assumed predefined maximal permissible distance, or if all nearer admissible points had too great estimated gradients \( G \leq \). In any case, a previously active point turn into inactive one. During this phase a remaining candidates, if any, the candidates with extreme gradients were attached to a cluster as the candidate a value of a gradient is estimated as \( G = \Delta Y/D \), and from consideration were excluded points from outside, no more distant than assumed maximal permissible distance. For each of 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 \( G = \Delta Y/D \), and from consideration were excluded candidates with either too great \( G \geq G_{\text{max}} \) or too little \( G \leq G_{\text{min}} \). 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 \( \tan(\mu) \) 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 \( \tan(\mu) = \Delta X_2/\Delta X_1 \) was calculated, and projection \( G_p \) of its gradient \( G = \Delta Y/D \) was estimated as \( G_p = G \cdot \cos(\phi - \mu) \). The Akaike’s AIC (2) can be used to estimate common direction \( \tan(\mu) \) for several clusters. A new axis \( Z \) for each estimated direction \( \tan(\mu) \) was located through origin \((X_1 = 0; X_2 = 0)\), 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 \((Y, X_1, X_2)\) 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: \( Y = a + b_1 Z + b_2 Z^2 \), 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

<table>
<thead>
<tr>
<th>df.reg</th>
<th>AIC</th>
<th>LSE</th>
<th>AIC</th>
<th>MLL</th>
<th>P(next=true)</th>
<th>Estimated model</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>94.52</td>
<td>380.31</td>
<td>0.71</td>
<td>( Y = a + b_1 X_1 + b_2 X_1^2 + b_3 X_1^3 + b_4 X_1^4 + b_5 X_2 + b_6 X_2^2 + b_7 X_2^3 + b_8 X_2^4 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>92.71</td>
<td>378.50</td>
<td>0.71</td>
<td>( Y = a + 0 + b_1 X_1^2 + b_2 X_1^3 + b_3 X_1^4 + b_4 X_2 + b_5 X_2^2 + b_6 X_2^3 + b_7 X_2^4 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>90.96</td>
<td>376.74</td>
<td>0.73</td>
<td>( Y = a + 0 + b_1 X_1^2 + b_2 X_1^3 + b_3 X_1^4 + 0 + b_4 X_2 + b_5 X_2^2 + b_6 X_2^3 + b_7 X_2^4 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>88.98</td>
<td>374.77</td>
<td>0.10</td>
<td>( Y = a + 0 + b_1 X_1^2 + b_2 X_1^3 + b_3 X_1^4 + 0 + b_4 X_2 + 0 + b_5 X_2^4 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>93.36</td>
<td>379.15</td>
<td>#</td>
<td>( Y = a + 0 + b_1 X_1^2 + b_2 X_1^3 + 0 + 0 + b_4 X_2 + 0 + b_5 X_2^4 )</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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

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

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

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

<table>
<thead>
<tr>
<th>Df.reg</th>
<th>Clusters</th>
<th>AIC</th>
<th>$P(\text{next=true})$</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>5</td>
<td>76.66</td>
<td>65%</td>
<td>$Y = Y_{\text{flat}} + DV_i^*(a_i + b_i^*Z); i = A, B, C, D$</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>75.39</td>
<td>60%</td>
<td>$Y = Y_{\text{flat}} + DV_i^*(a_i + b_i^*Z); i = A, B, C, D$</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>74.55</td>
<td>0%</td>
<td>$Y = Y_{\text{flat}} - Y_0 + DV_i^<em>b_i^</em>(Z - Z_0); i = A, B, C, D$</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>111.4</td>
<td>#</td>
<td>$Y = Y_{\text{flat}} - Y_0 + DV_i^<em>b_i^</em>(Z - Z_0); i = AD, BC$</td>
</tr>
</tbody>
</table>

where: $Y_0, Z_0$ – 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