Advances in Adaptive Data Analysis
Vol. 2, No. 4 (2010) 451–462

DOI: 10.1142/S1793536910000574

MEANINGFUL REGRESSION COEFFICIENTS BUILT BY DATA GRADIENTS

STAN LIPOVETSKY
GfK Custom Research North America,
8401 Golden Valley Road,
Minneapolis, MN 55427, USA
stan.lipovetsky@gfk.com

Multiple regression's coefficients define change in the dependent variable due to a predictor's change while all other predictors are constant. Rearranging data to paired differences of observations and keeping only biggest changes yield a matrix of a single variable change, which is close to orthogonal design, so there is no impact of multicollinearity on the regression. A similar approach is used for meaningful coefficients of nonlinear regressions with coefficients of half-elasticity, elasticity, and odds' elasticity due the gradients in each predictor. In contrast to regular linear and nonlinear regressions, the suggested technique produces interpretable coefficients not prone to multicollinearity effects.

Keywords: Regressions; gradients; multicollinearity.

1. Introduction

Regression modeling presents one of the main tools of multivariate statistical methods. The ordinary least squares (OLS) multiple regression is the best linear aggregate of the predictors to fit and predict the data. However, such a model is not always useful for the analysis of the individual predictor's impact on the dependent variable. Because of high correlations, or multicollinearity among the predictors, the correlation matrix can be ill-conditioned; therefore, its inversion yields OLS coefficients that are inflated and shifted to a wide range of values of both signs. It is not immediately clear how to analyze such a model, i.e. to estimate which of its regressors are more important, and to use this regression to predict improvement in the dependent variable by changing the values of the independent variables. For instance, if a presumably beneficial variable received a negative coefficient in multiple regression, should we use its higher or lower value to get a lift in the outcome?

The problem is that OLS had never been designed to find meaningful individual coefficients, only their best linear combination to fit the data. Therefore, the individual coefficients in OLS are often hardly interpretable, with the exception of models obtained by data in active experiments by orthogonal plans. To reduce the effects of multicollinearity on the regression coefficients various approaches have been suggested. Among those are ridge regression and its modifications [Hoerl and Kennard, 1970, 1976; Jensen and Ramirez, 2008; Lipovetsky, 2010], lasso, least angle, and sparse regressions [Tibshirani, 1996; Efron *et al.*, 2004; Meier *et al.*, 2008; Fraley and Hesterberg, 2009; James and Radchenko, 2009; Lipovetsky, 2009], and Shapley value regression [Lipovetsky and Conklin, 2001a].

This paper presents a new approach to regression modeling based on the meaning and interpretation of the regression coefficients themselves. For a multiple linear regression, its coefficients are defined as a change in the dependent variable due to the change in each predictor, subject to all other predictors fixed at constant levels. If we rearrange the data into paired differences among the observations and clean it of smaller changes, it is possible to produce a matrix corresponding to change in only one variable in each paired observation. Due to the nearly orthogonal design, there is no multicollinearity effect when constructing the multiple regression. To adjust the model to the best quality of fit, an additional procedure developed earlier for adjusting ridge models [Lipovetsky, 2010] is applied. The obtained regression yields meaningful coefficients constructed in accordance with their definition as slopes by each variable subject to all other regressors fixed constant.

A similar approach is suggested for constructing meaningful coefficients for nonlinear models as well, including exponential, multiplicative Cobb–Douglas, and logistic models. Their coefficients correspond to elasticity, half-elasticity, and elasticity of odds due to a change in each predictor and constant levels for all the other variables. In contrast to regular linear and nonlinear regressions, the considered techniques produce clear-cut and interpretable solutions for the coefficients and good quality characteristics overall.

The paper is organized as follows: Secs. 2 and 3 describe the new technique for the multiple linear and linear-link regressions, respectively. Section 4 presents numerical results of the suggested approach in comparison with other techniques, and Sec. 5 summarizes.

2. Linear Regression by Data Gradients

Let all the variables be standardized, so centered and normalized by standard deviations. The theoretical multiple linear regression model can be presented as follows:

$$y_i = b_1 x_{i1} + b_2 x_{i2} + \cdots + b_n x_{in}, \tag{1}$$

where y_i and x_{ij} are the standardized i-th observations ($i = 1, \ldots, N$) by the dependent variable y and by each j-th independent variable $x_j (j = 1, 2, \ldots, n)$, and b_j are the so-called beta-coefficients of the standardized regression. Consider two observations in i-th and h-th points (1), and subtract one of them from another:

$$y_i - y_h = b_1(x_{i1} - x_{h1}) + b_2(x_{i2} - x_{h2}) + \cdots + b_n(x_{in} - x_{hn}). \tag{2}$$

It is clear that each individual coefficient b_j of the regression has a meaning of the change in y due to the change in the predictor x_j while all the other predictors are held constant. Therefore, if with the exception of only x_j all the other predictors have zero difference, then Eq. (2) yields:

$$y_i - y_h = b_j(x_{ij} - x_{hj}). \tag{3}$$

Thus, each coefficient b_j of linear regression equals an absolute change in y due to a change in the j-th predictor x_j while all other predictors are unchanged, so $x_{ik} - x_{hk} = 0$ for all $k \neq j$. In explicit form, the coefficient can be represented from Eq. (3) as:

$$b_j = \frac{(y_i - y_h)}{(x_{ij} - x_{hj})}. \tag{4}$$

The expression (4) shows a paired coefficient built by a data gradient of one predictor in two points of observations.

With total data, each coefficient can be estimated as the mean of all the gradients (4):

$$b_j = \frac{1}{N(N-1)} \sum_{i \neq h}^{N} \frac{y_i - y_h}{x_{ij} - x_{hj}}. \tag{5}$$

Of course, the points with the same values in two observations, $x_{ij} - x_{hj} = 0$, are not used in Eq. (5). Another estimation of b_j can be made using the paired regression for model (3) of the differences in y by the differences in a predictor x_j. As it is well known, the coefficient of paired regression equals the quotient of the sample covariance between the two variables to the variance of the independent variable, which in case of Eq. (3) yields:

$$b_j = \frac{\sum_{i \neq h}^{N} (y_i - y_h)(x_{ij} - x_{hj})}{\sum_{i \neq h}^{N} (x_{ij} - x_{hj})^2}, \tag{6}$$

On the other hand, the second moments calculated by paired differences are proportional to the second moments calculated by the centered data [Puri and Sen, 1971; Lipovetsky and Conklin, 2001b]. With bar denoting the mean value of a variable, the sample covariance can be calculated by either of two ways:

$$\mathrm{cov}(y, x_j) = \frac{1}{N-1} \sum_{i=1}^{N} (y_i - \bar{y})(x_{ij} - \bar{x}_j) = \frac{1}{2N(N-1)} \sum_{i \neq h}^{N} (y_i - y_h)(x_{ij} - x_{hj}) \tag{7}$$

and similar for the variance. Therefore, each coefficient b_j estimated by all the data (6) can be calculated simply by the original data as a coefficient of the paired regression of y by the j-th variable. By the same reason, the OLS being applied to the data in models (1) or (2) yields the same coefficients of regression.

Consideration in Eqs. (3)–(7) is valuable only if one predictor changes and all the others are constant in each paired difference of i-th and h-th observations (2).

It corresponds to the absence of correlations among the predictors, which is hardly possible for a real observed or elicited data, with the exception of active experimental designs with orthogonal plans. If two or more variables are highly correlated, it leads to poorly defined coefficients in the model (1) due to multicollinearity. Indeed, suppose the first two variables are highly correlated, so their values are changing simultaneously in a similar way, so differences of any two i-th and h-th observations are almost equal, $x_{i1} - x_{h1} \approx x_{i2} - x_{h2}$. With all the other variables of zero differences, the change in y (2) is:

$$y_i - y_h = b_1(x_{i1} - x_{h1}) + b_2(x_{i2} - x_{h2}) \approx (b_1 + b_2)(x_{i1} - x_{h1}), \tag{8}$$

which means that the two coefficients b_1 and b_2 cannot be identified separately, but only as a total coefficient $b = b_1 + b_2$. Then, any numerical estimation would produce arbitrary individual values of b_1 and b_2 such that only their total is defined uniquely. Thus, a solution for identification of each coefficient of regression can be found in the following procedure.

Suppose, the paired differences of y are stacked in the vector Δy of the $N(N-1)$ order, and all the x differences are stacked in the rows of the matrix ΔX consisting of $N(N-1)$ rows and n columns (those are the same differences used in Eqs. (2)–(5)). The number $N(N-1)$ corresponds to all the differences without those between the same points of observations (it is possible to use only half of them because those of any i-th and h-th observations and vice versa would be of the opposite signs). In each row of the matrix ΔX, let us find the element with the maximum absolute value $\max_j(|x_{ij} - x_{hj}|)$ and keep the value $x_{ij} - x_{hj}$ in its place while all the other values in this row are set to zero. The cleaned matrix can then be denoted by tilde, $\Delta\tilde{X}$, and its elements are:

$$\Delta\tilde{x}_{tj} \equiv \Delta\tilde{x}_{ih,j} = \begin{cases} x_{ij} - x_{hj}, & \text{if } |x_{ij} - x_{hj}| = \max\limits_j(|x_{ij} - x_{hj}|) \\ 0, & \text{if } |x_{ij} - x_{hj}| < \max\limits_j(|x_{ij} - x_{hj}|), \end{cases} \tag{9}$$

where t denotes an index changing from 1 to total $N(N-1)$ rows by all paired i-th and h-th observations' differences. If there are two or more coinciding maximum values in any t-th row, it is totally excluded from the matrix (9) and from the vector Δy because such a case is a source for the problem described in Eq. (8). The obtained matrix $\Delta\tilde{X}$ is close to orthogonal, so the corresponding correlation matrix has nondiagonal elements close to zero in comparison with the diagonal ones.

The OLS multiple regression can be found by the stacked differences' data as follows. The model with the theoretical dependence (2) can be presented via Eq. (9) in matrix form as:

$$\Delta y = \Delta\tilde{X}\beta + \varepsilon, \tag{10}$$

where Δy is the stacked vector of differences in the dependent variable, β is a vector of the regression coefficients, and ε is a vector of the deviations in the model. The least squares (LS) objective of minimizing the sum of squared deviations is:

$$S^2 = ||\varepsilon||^2 = (\Delta y - \Delta\tilde{X}\beta)'(\Delta y - \Delta\tilde{X}\beta) = \Delta y'\Delta y - 2\beta'\Delta\tilde{X}'\Delta y + \beta'\Delta\tilde{X}'\Delta\tilde{X}\beta, \tag{11}$$

where prime denotes transposition. Note that after transformation to the stacked and cleaned matrix and vector, the cross-products for the second-moments in Eq. (11) correspond to the covariance rather than correlation. The first-order condition $\partial S^2/\partial\beta = 0$ of minimization yields a normal system of equations with the corresponding OLS solution:

$$\beta = (\Delta\tilde{X}'\Delta\tilde{X})^{-1}\Delta\tilde{X}'\Delta y. \tag{12}$$

The matrix $\Delta\tilde{X}'\Delta\tilde{X}$ is close to an identity matrix, so there are no vastly inflated values in its inversion (12), or in the coefficients of regression. Thus, in contrast to the regular regression, there is no problem with multicollinearity in the solution (12), which approximately equals $\Delta\tilde{X}'\Delta y$, so it consists of the values proportional to the coefficients of paired regressions (6).

When the solution (12) by the stacked data (9) for the model (10) is found, it can be adjusted to the original model (1) using the following procedure developed earlier for ridge regression models [Lipovetsky, 2010]. A quality of fit for the model (1) is estimated by the residual sum of squares similar to Eq. (11), or by the coefficient of multiple determination:

$$R^2 = 1 - S^2 = 2b'X'y - b'X'Xb, \tag{13}$$

where the original standardized data are used, so the relation $y'y = 1$ holds. Multiplying the vector-row b' by the OLS normal system of equations $X'Xb = X'y$ yields the equality $b'X'Xb = b'X'y$ with which the characteristic (13) reaches its maximum and reduces to the following forms:

$$R^2 = b'X'y = b'X'Xb. \tag{14}$$

For solution (12), consider a proportionally modified vector of the original model (1):

$$b = q\beta, \tag{15}$$

and substitute it into the general expression for the multiple determination coefficient (13):

$$R^2 = 2q\beta'X'y - q^2\beta'X'X\beta. \tag{16}$$

This concave quadratic function by the fit parameter q reaches its maximum at the value:

$$q = \frac{\beta'X'y}{\beta'X'X\beta}, \tag{17}$$

with the vector β defined in Eq. (12). Substituting Eq. (17) into Eqs. (15)–(16) yields the adjusted solution and the coefficient of multiple determination maximum attained:

$$b = \frac{\beta' X' y}{\beta' X' X \beta} \beta, \quad R^2 = \frac{(\beta' X' y)^2}{\beta' X' X \beta}. \tag{18}$$

The adjusted solution b (18) is easy to find by Eq. (12), and the maximum R^2 (18) can be represented in any of two equivalent forms (14).

For the linear model (1) in the original units

$$y = a_0 + a_1 x_1 + a_2 x_2 + \cdots + a_n x_n, \tag{19}$$

its coefficients can be constructed from the standard coefficients b (18) by the relations:

$$a_j = \frac{b_j \sigma_y}{\sigma_j}, \quad j = 1, \ldots, n; \quad a_0 = \bar{y} - a_1 \bar{x}_1 - \cdots - a_n \bar{x}_n, \tag{20}$$

where σ_y and σ_j denote standard deviations for y and each x_j. The intercept a_0 (19) is defined using the mean values of the variables denoted in Eq. (20) by bar. Similarly to Eq. (4), each coefficient $a_j = \Delta y / \Delta x_j$ of regression (19) corresponds to the change in y due to the change in one x with others being fixed.

3. Nonlinear and Linear Link Regressions

Many other models can be linearized and constructed with the meaningful coefficients by the approach described above. For instance, the exponential model

$$e^y = e^{a_0} x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n}, \tag{21}$$

after linearization becomes

$$y = a_0 + a_1 \ln x_1 + a_2 \ln x_2 + \cdots + a_n \ln x_n. \tag{22}$$

Therefore, each coefficient a_j of regression can be presented via a partial derivative, and approximately by the finite changes Δy and Δx_j, subject to all other variables fixed:

$$a_j = \frac{\partial y}{\partial \ln x_j} = \frac{\partial y}{\frac{\partial x_j}{x_j}} \approx \frac{\Delta y}{\frac{\Delta x_j}{x_j}}. \tag{23}$$

This is the half-elasticity of absolute change in y due to the relative (percentage) change of x_j. The above-described approach (1)–(20) of constructing coefficients not prone to multicollinearity can be used by the predictors transformed to the relative changes.

Another widely used model is a multiplicative analog of the Cobb–Douglas function

$$y = e^{a_0} x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n}, \tag{24}$$

which in linearized form is:

$$\ln y = a_0 + a_1 \ln x_1 + a_2 \ln x_2 + \cdots + a_n \ln x_n. \tag{25}$$

Each coefficient a_j can be presented via the partial derivative and approximated by the finite changes as follows:

$$a_j = \frac{\partial \ln y}{\partial \ln x_j} = \frac{\frac{\partial y}{y}}{\frac{\partial x_j}{x_j}} \approx \frac{\frac{\Delta y}{y}}{\frac{\Delta x_j}{x_j}}. \tag{26}$$

Therefore, a coefficient of regression corresponds to the elasticity, or the percentage gradient in y by x_j. The same approach (1)–(20) for constructing meaningful coefficients can be used by the dependent variable and all the predictors transformed into the relative changes, or to the differences of logarithms for the observed values.

For any objective used for constructing a regression, such as OLS for linear and linearized models, or maximum likelihood for nonlinear models, the numerical estimation of the parameters would use a covariance or correlation matrix, or its analog — a Hessian matrix of the second derivatives of the objective by the estimated parameters. In linear modeling, it leads to inverting such a matrix to find an OLS solution. In nonlinear estimation, the Newton–Raphson procedure is applied for iterative solving, which includes inversion of the Hessian. If such a matrix has two or more similar variables, its determinant is close to zero, and inversion yields highly inflated estimates belonging to a wide range of values. Therefore, more complicated models could also easily produce pointless parameters, and special consideration is needed to obtain meaningful individual coefficients of regression. Let us describe how to solve such a problem for a model with binary output in logistic regression.

A binary 0-1 dependent variable p can be considered as a logit model of the predictors:

$$p = \frac{1}{1 + \exp(-(a_0 + a_1 x_1 + \cdots + a_n x_n))}. \tag{27}$$

Estimation of the coefficients in Eq. (27) is performed using maximum likelihood objectives [Long, 1997; Lloid, 1999; Lipovetsky and Conklin, 2000; Lipovetsky, 2006]. When the model (27) is constructed, it can be used for prediction of the values p by the independent variables for each observation. The predicted values present a continuous variable in the range $p \in [0, 1]$, which estimates the probability of the binary event.

By the predicted values p, the model (27) can be represented as a linear link:

$$y \equiv \ln \frac{p}{1 - p} = a_0 + a_1 x_1 + a_2 x_2 + \cdots + a_n x_n, \tag{28}$$

where the left-hand side corresponds to the new dependent variable z defined by the logarithm of the probability's odds. A coefficient of the logit model can be obtained as a partial derivative of the logarithm of odds by the predictor when all the others are held fixed:

$$a_j = \frac{\partial y}{\partial x_j} \approx \frac{\Delta y}{\Delta x_j} = \frac{\Delta \ln\left(\frac{p}{(1-p)}\right)}{\Delta x_j} = \frac{\Delta\left(\frac{p}{(1-p)}\right)}{\left(\frac{p}{(1-p)}\right)\Delta x_j}. \tag{29}$$

Thus, a coefficient of the logit model can be interpreted as the change in the logarithm of odds, or in other words as the relative change of the odds, due to the change in the corresponding predictor.

As with any predictors' aggregate in a statistical model, the individual coefficients (27) may not have sense. However, with the help of definition (29), it is possible to use the following procedure to obtain the meaningful logit coefficients. First, construct the model (27) and find the predicted values p. Second, construct the new variable z of the logarithm of the probability's odds (28). Third, apply the above-given approaches (1)–(20) to reconstruct the parameters of the linear link model (28). And forth, substitute the newly obtained meaningful parameters of (28) into the logit model (27).

4. Numerical Example

For a numerical example, consider data from a real research project on customer satisfaction with a bank. The dependent variable is overall satisfaction, and 28 independent variables are briefly listed in Table 1.

All the variables are measured as a Likert scale from 1 to 10, from the worst to the best estimate, and there are 407 respondents. In its first numerical column, Table 1 presents the vector of pair correlations r_{yx} of y with x — all the correlations are positive, as is expected by the beneficial meaning of the predictors with overall satisfaction with the bank. However, because of multicollinearity among the 28 variables, 12 of them (43%, almost a half of the predictors!) receive negative signs of the coefficients in the linear regression — see beta coefficients (1) in the column of OLS in Table 1. Therefore, this linear regression is useless for analysis of the predictors' impact on overall satisfaction. It is also hardly possible to use such a model for prediction because it is not clear which regressors should be increased to reach a lift in the output. For instance, should we increase the values of presumably useful variables if in the model they have negative signs?

The next three columns of linear models in Table 1 present the improved estimations of the regression coefficients that are all positive: SV, or Shapley value regression constructed from the game theory approach [Lipovetsky and Conklin, 2001a]; Slope and OLS Grad linear models constructed by slopes (5), and by OLS gradient data (12), respectively. The last two regressions are estimated by the paired cleaned data (9) and adjusted (18) by the fit parameter (17). The q parameter, coefficient of multiple determination R^2, and the quotients of the multiple determination to its OLS maximum value $R^2/R^2_{\max}$ are given in the last three rows of Table 1, under the

Table 1. Linear and logit regular regressions, SV regressions, and models by data gradients.

Predictors	r_{xy}	Linear models				Logit and linear link models			
		OLS	SV	Slope	OLS Grad	Logit regular	SV link	Slope link	OLS Grad link
Bank is the leader	0.599	−0.007	0.060	0.053	0.056	0.035	0.058	0.075	0.075
New products	0.571	−0.025	0.052	0.035	0.034	−0.088	0.037	0.015	0.018
Prefer do business	0.631	−0.031	0.064	0.048	0.050	−0.265	0.054	0.032	0.038
Better value	0.594	0.067	0.060	0.063	0.059	0.196	0.073	0.095	0.091
Relationship	0.582	0.022	0.057	0.042	0.039	0.175	0.082	0.080	0.073
Easy do business	0.694	0.242	0.109	0.077	0.070	0.234	0.093	0.080	0.073
Solutions	0.671	0.126	0.082	0.079	0.079	0.096	0.065	0.074	0.075
Knows your needs	0.649	0.105	0.075	0.071	0.068	0.105	0.057	0.083	0.080
Satisfied w. checking	0.659	0.250	0.103	0.091	0.087	0.439	0.108	0.115	0.109
Satisfied w. deposits	0.332	−0.035	0.023	0.035	0.044	−0.080	0.019	0.049	0.056
Satisfied w. mortgage	0.283	0.083	0.024	0.040	0.047	0.452	0.055	0.068	0.064
Sat. w. other loans	0.388	0.078	0.037	0.068	0.070	0.020	0.027	0.057	0.054
Satisfied w. investment	0.207	0.044	0.017	0.036	0.037	−0.003	0.025	0.047	0.044
Sat. w. visa/master	0.368	0.012	0.032	0.037	0.037	0.220	0.035	0.041	0.043
Sat. w. check/debit	0.210	−0.026	0.012	0.019	0.020	0.628	0.036	0.042	0.044
Fee checking	0.610	0.024	0.071	0.056	0.055	−0.071	0.056	0.071	0.069
Fee deposits	0.319	−0.004	0.017	0.027	0.027	−0.016	0.008	0.024	0.032
Fee mortgage	0.224	−0.073	0.006	0.024	0.016	−0.098	0.019	0.039	0.026
Fee other loans	0.336	−0.025	0.018	0.044	0.047	0.130	0.025	0.052	0.049
Fee investments	0.184	−0.066	0.007	0.021	0.018	0.289	0.025	0.054	0.058
Fee visa/master	0.331	0.051	0.026	0.013	0.015	−0.098	0.013	0.011	0.015
Fee check/debit	0.251	−0.003	0.001	0.033	0.053	−0.621	0.006	0.005	0.020
Sat w. call center	0.611	0.153	0.087	0.066	0.054	0.246	0.087	0.091	0.088
Sat. w. call rep.	0.311	0.022	0.024	0.038	0.038	−0.018	0.000	0.037	0.038
Sat. w. call auto	0.214	−0.047	0.002	0.022	0.022	−0.129	0.014	0.019	0.023
Sat. online	0.163	0.040	0.000	0.023	0.020	0.621	0.079	0.058	0.050
Problem managed	0.239	0.149	0.040	0.044	0.044	1.171	0.074	0.090	0.078
Problem resolved	0.189	−0.101	0.017	0.021	0.026	−1.087	0.013	0.022	0.030
q		1	1	0.182	0.158	1	0.853	0.205	0.184
R^2		0.656	0.613	0.586	0.577	0.590	0.580	0.505	0.497
$R^2/R^2_{\max}$		1	0.934	0.893	0.880	1	0.983	0.856	0.842

coefficients of corresponding regressions. The multiple OLS regression has a good overall $R^2 = 0.656$, while the quality of the SV, Slope, and OLS Grad models are slightly less at about 93%, 89%, and 88%, respectively. It is a trade-off for obtaining the models with meaningful coefficients useful for both analysis and prediction.

The right-hand side of Table 1 presents Logit modeling, as it is described after the formulae (27)–(29). Using the same bank data, the problem consists of determining the predictors' impact on top two box reach so the higher level of overall satisfaction is considered. For this aim, the dependent 10-point scale was transformed to a binary variable with 1 for top two and 0 for all other boxes, respectively. Logistic regression by all the original predictors is presented in Table 1, the column titled Logit regular. This model also contains 12 variables with negative signs in spite of the presumably positive influence of the independent variables on attaining the top two levels of overall satisfaction, although many of these variables of negative impact differ from those in the OLS model. Anyway, it is not clear how to interpret the negative coefficients with useful predictors. The quality of fit $R^2 = 0.590$ in this model is given by a quasi coefficient of determination, the analog of the coefficient of multiple determination estimated as one minus the quotient of the residual deviance to null deviance.

However, the constructed Logit model can be successfully used for finding the predicted values for the probability of the event by Eq. (27). Then, the output y of the logarithm of the probability's odds is calculated as the dependent variable in the linear link model (28). With the data for the linear model (28), SV regression is constructed for reestimating the coefficients of the logit. In addition, the algorithm of paired cleaned differences (9) is used to obtain the data for finding the Slope (5) and OLS Grad (12) linear link models (28) constructed as the percentage increments in odds by the change in each individual predictor (29), with adjustment (18) by fit parameter (17). These linear link models are presented in the last three columns of Table 1. Fitting parameter q, the coefficient of multiple determination R^2, and the quotients of the multiple determination to its Logit maximum value $R^2/R^2_{\max}$ are given in the last three rows of Table 1, under the coefficients of Logit and linear link regressions. The quality of the linear link SV, Slope, and OLS Grad models is about 98%, 86%, and 84% of the Logit, respectively. Again, it is a small price to pay for getting meaningful coefficients for the logit model. The sets of positive coefficients of any of the three last models from Table 1 can be used in the logit regression (27). Such a model presents meaningful and interpretable parameters of the predictor's incremental influence on lifting the relative odds of probability (29) of the event under consideration.

5. Summary

Using the definition of a coefficient of linear regression as the change in the dependent variable due to the change in the predictor, subject to all other predictors being held constant, it is shown how to rearrange the data into paired differences among the observations and clean them of smaller changes. The obtained data matrix corresponds to a one-variable change in each paired observation, so such a design is close to orthogonal, and there cannot be any impact from multicollinearity on the constructed coefficients of multiple regression. An additional procedure for adjusting the model to the best quality of fit is applied. This approach can also be used for constructing meaningful coefficients of nonlinear regressions such as exponential, Cobb–Douglas, and logistic models with coefficients interpreted as half-elasticity, elasticity, and odds' elasticity due to gradients in each predictor.

In contrast to regular linear and nonlinear regressions, the suggested technique produces interpretable coefficients not prone to multicollinearity effects.

The numerical results considered above for linear, logistic, and linear link regressions by the paired differences data are very typical and have been observed with various data sets. Shapley value regression and slope-gradient models are developed in rather different approaches but they yield similar sets of coefficients, supporting each other results. It is interesting to note that in the work [Lipovetsky, 2009], the SV approach was also compared with ridge regressions and other models of special parameterization techniques — they all produce close results. In future research, it is important to establish conditions for this similarity of the different techniques giving solutions not prone to the effects of multicollinearity.

Although the SV models could have a better quality of fit than the gradient data models, they are much more complicated to implement and need special software, so are less available in comparison with the easily attainable paired difference data (9) and the slope (5) or regular OLS (12) estimations, including those by the linear link (28) for logistic regressions. Correlations by the transformed data (9) are close to zero, so the variance inflation factor for the model (12) approximately equals one, which indicates the absence of multicollinearity distortion of the coefficients of regression. The same is true for the linear link model (28), which produces coefficients for the logistic model not prone to multicollinearity. Future research should also consider a weighting procedure for keeping more data in Eq. (9) while holding its orthogonal design.

The coefficients of regressions obtained in the suggested approach of data gradients are stable, easily interpretable, useful in practical applications, and provide a meaningful analysis of the individual predictors' impact on the outcome variable.

References

Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. (2004). Least angle regression. *Ann. Stat.*, **32**: 407–489.

Fraley, C. and Hesterberg, T. (2009). Least angle regression and LASSO for large datasets. *Stat. Anal. Data Min.*, **1**: 251–259.

Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*, **12**: 55–67.

Hoerl, A. E. and Kennard, R. W. (1976). Ridge regression: iterative estimation of the biasing parameter. *Commun. Stat.*, **A5**: 77–88.

James, G. M., Radchenko, P. and Lv, J. (2009). DASSO: connections between the Dantzig selector and lasso. *J. R. Stat. Soc.*, **B71**: 127–142.

Jensen, D. R. and Ramirez, D. E. (2008). Anomalies in the foundations of ridge regression. *Int. Stat. Rev.*, **76**: 89–105.

Lipovetsky, S. (2006). Entropy criterion in logistic regression and Shapley value of predictors. *J. Mod. Appl. Stat. Methods*, **5**: 121–132.

Lipovetsky, S. (2009). Linear regression with special coefficient features attained via parameterization in exponential, logistic, and multinomial-logit forms. *Math. Comput. Model.*, **49**: 1427–1435.

Lipovetsky, S. (2010). Enhanced ridge regressions. *Math. Comput. Model.*, **51**: 338–348.

Lipovetsky, S. and Conklin, M. (2000). Box-Cox generalization of logistic and algebraic binary response models. *Int. J. Oper. Quant. Manage.*, **6**: 276–285.

Lipovetsky, S. and Conklin, M. (2001a). Analysis of regression in game theory approach. *Appl. Stochastic Models Bus. Ind.*, **17**: 319–330.

Lipovetsky, S. and Conklin, M. (2001b). Regression as weighted mean of partial lines: interpretation, properties, and extensions. *Int. J. Math. Educ. Sci. Technol.*, **32**: 697–706.

Lloid, C. J. (1999). *Statistical Analysis of Categorical Data*. Wiley, New York.

Long, J. S. (1997). *Regression Models for Categorical and Limited Dependent Variables*. Sage Publication, London.

Meier, L., van de Geer, S. and Buhlmann, P. (2008). The group lasso for logistic regression. *J. R. Stat. Soc.*, **B70**: 53–71.

Puri, M. L. and Sen, P. K. (1971). *Nonparametric Methods in Multivariate Analysis*. Wiley, New York.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Stat. Soc.*, **B58**: 267–288.