LINEAR AND SUPPORT VECTOR REGRESSIONS BASED ON GEOMETRICAL CORRELATION OF DATA

Kaijun Wang[1]*, Junying Zhang[1], Lixin Guo[2] and Chongyang Tu[1]

*[1]School of Computer Science and Engineering, Xidian University, Xian 710071, P. R. China.
Email: sunice9@yahoo.com
[2]Dept of Computer Science, Xian Institute of Post-telecommunications, Xian 710061, P. R. China.

ABSTRACT

Linear regression (LR) and support vector regression (SVR) are widely used in data analysis. Geometrical correlation learning (GcLearn) was proposed recently to improve the predictive ability of LR and SVR through mining and using correlations between data of a variable (inner correlation). This paper theoretically analyzes prediction performance of the GcLearn method and proves that GcLearn LR and SVR will have better prediction performance than traditional LR and SVR for prediction tasks when good inner correlations are obtained and predictions by traditional LR and SVR are far away from their neighbor training data under inner correlation. This gives the applicable condition of GcLearn method.

Keywords: Geometrical correlation learning, Geometrical correlation of data, Regression analysis

1   INTRODUCTION

In multivariate data analysis, linear regression (LR) and support vector regression (SVR) (Lin, 2001; Meyer et al., 2003; Zhou et al., 2006; Chang & Lin, 2006) are widely used to analyze the relationship between a dependent variable and a set of independent variables, and a regression model between variables by LR and SVR methods may be used for prediction tasks (Chen et al., 2004; Wagner et al., 2005; Sinnakaudan et al., 2006). When a regression model is constructed for 1-dimensional continuous variables, dependency relations among variables are the focus, but correlations between data of a variable are usually neglected. It is expected to improve predictive ability of LR and SVR methods by using this correlation information.

Let *Y* be the response variable of some 1-dimensional dependent variables and have *n* data. The space relation of the *n* data points in 1-dimensional coordinates *y*, i.e., T=$\{y_1, y_2, y_3, \ldots, y_n\}$ where $y_1 \leq y_2 \leq y_3 \leq \ldots \leq y_n$, is called *value correlation* of the data, while the neighbor relation between the *n* data points with some varying trends is called *trend correlation*. The value correlation indicates which data points are close each other in 1-dimensional coordinates, and the trend correlation is usually from additional information and prior knowledge about correlations, e.g., the trend correlation is time relation if the *n* data points are time series and vary with time. Both value correlation and trend correlation are called *inner correlations*, and an inner correlation represented by a geometric entity is called a *geometrical correlation*.

Recently, geometrical correlation learning (GcLearn) (Wang et al., 2007) has been proposed to mine and use the inner correlations inherent in data for LR and SVR methods. GcLearn method can improve the predictive ability of LR and SVR methods, as it makes use of additional information - inner correlations of data, while traditional LR and SVR methods do not. First, GcLearn finds inner correlation T from data of the response variable. According to T, GcLearn projects the 1-dimensional data of each variable to a 2-dimensional smooth curve (called a *curve manifold*), which is approximated by piecewise quadratic polynomial curves. Thus, a curve manifold represents data of a variable and shows their geometrical correlation. Then, a regression model for variables is found by the designed geometric regression method and by minimizing the fitting error of the model to curve manifolds. Finally, the optimal regression model *F* is found through an optimization process for piecewise curve approximation. When we predict with model *F*, test datum *x* of predictor variable *X* is also projected to *x'* on the curve manifold of *X* (it is similar for other predictor variables), and then the *x'* is used as the input to model *F* for predictions.

In this paper, we theoretically analyze the prediction performance of GcLearn LR and SVR methods and prove that the GcLearn LR/SVR method gives better prediction performance than the traditional LR/SVR method when good inner correlations are obtained and predictions from a LR/SVR model are bad. The GcLearn method from our previous work is introduced in Section 2, and the theoretical analysis of its prediction performance is in Section 3. Finally, the conclusion is given in Section 4.

2 GCLEARN LINEAR AND SUPPORT VECTOR REGRESSIONS

The GcLearn method (Wang et al., 2007) is briefly introduced in this section. The GcLearn method includes finding inner correlations, projecting data to curve manifolds as per inner correlations, finding a regression model from curve manifolds and predicting with the found model.

The first work is to find inner correlations and project the data of each variable to a 2-dimensional smooth curve *M* so that the *M* represents the data and their geometrical correlations. The value correlations are mined from data while the trend correlations of the data depend on or are gained by prior knowledge about correlations, e.g., the trend correlation is time relation if the *n* data points are time series and vary with time. Here we introduce the whole procedure by an example of value correlations, and it is similar for trend correlations. It is hard to make one curve for all the data of a variable reasonably and accurately in all cases, so we approximate the *M* via piecewise curves; each fit to partial data, including data division, making of a piecewise curve fitting to a data group, and connections of piecewise curves. The construction of such *M* is discussed within the framework of manifold theory (Chen, 2001).

Let variable *Y* be the response variable of *q* 1-dimensional continuous dependent variables and have *n* data points. The value correlation is obtained when the *n* data points are placed in 1-dimensional coordinates *y*, i.e., T=$\{y_1, y_2, y_3, \ldots, y_n\}$ where $y_1 \leq y_2 \leq y_3 \leq \ldots \leq y_n$. Then, natural numbers are adopted to represent T for constructing *M*, i.e., T=$\{1,2,3, \ldots, n\}$ instead of and corresponding to $\{y_1, y_2, y_3, \ldots, y_n\}$. Let the *n* data be divided equally into *k* groups each with $m=n/k$ data along T (e.g., second group $\{y_{m+1}, y_{m+2}, \ldots, y_{m+m}\}$) and each data group be called *local data*. To simplify discussion, the notation $y_j \in R$ ($j=1,2,\ldots,m$) is adopted for each data group.

Then, GcLearn projects the *n* data points onto a 2-dimensional smooth curve *M* (called *curve manifold*) (Chen, 2001) in space *yOt* (coordinates *t* is used to denote T):

$$P: \{y_j\}_{j=1 \sim n} \rightarrow M: M = \cup U_i, \tag{1}$$

where local region U_i of *M* is described by a piece of quadratic polynomial curve C_i: $y=f_i(u)$ ($u \in [1,m]$) under local coordinate system *yOu* (coordinates *u* are used to denote a part of T for local data). The piecewise curve $y=f_i(u)$ fits to the data of *i*th group (*m* pairs of $\{y_j,u_j\}$ where $u_j=j$ and $j=1,2,\ldots,m$) and is found by a least-squares fit (Mo & Liu, 2003; Wolfram, 2007). The united local regions need to be connected smoothly so that their joints are continuous and smooth, but in real applications, we do not do this difficult work on account of slight influence on outcomes.

The next step is to design a geometric regression method to construct a regression model with curve manifolds, which is discussed by example of the linear regression between two variables. The least-squares fit method combined with integral computation of curve manifolds for solving a regression equation is called as *geometric regression* method. Let there be a linear regression equation for predictor variable *X* and response variable *Y*, and let same coordinate system *xOu* be used for local regions of curve manifold M_x and the same *yOu* for M_Y. Then, any point of a local region of M_x is (*x,u*) and any point of M_Y is (*y,u*).

Theorem 2.1 (Geometric regression). Let $y=\phi_0+\phi x$ be the linear regression equation between variables *X* and *Y* (their curve manifolds are M_x and M_Y). Let every curve manifold comprise *k* local regions and the local region of M_x be $x=g_l(u)$ and the one of M_Y be $y=f_l(u)$, where $u \in [u_1, u_2]$ and $l=1 \sim k$. And the means of *x* values of M_x and the means of *y* values of M_Y are set as follows:

$$\overline{x} = \frac{1}{k} \frac{1}{u_2-u_1} \sum_{l=1}^{k} \int_{u_1}^{u_2} g_l(u)du, \tag{2}$$

$$\overline{y} = \frac{1}{k} \frac{1}{u_2-u_1} \sum_{l=1}^{k} \int_{u_1}^{u_2} f_l(u)du. \tag{3}$$

The least-squares fit method is used to solve for the coefficients of $y=\phi_0+\phi x$ with M_x and M_Y, and then the regression coefficients are:

$$\phi = Lyx/Lxx, \tag{4}$$

$$\phi_0 = \overline{y} - \phi \overline{x}, \tag{5}$$

where

$$Lyx = \sum_{l=1}^{k} \int_{u_1}^{u_2} f_l(u)g_l(u)du - k(u_2 - u_1)\overline{y}\,\overline{x} \quad, \tag{6}$$

$$Lxx = \sum_{l=1}^{k} \int_{u_1}^{u_2} g_l(u)^2 du - k(u_2 - u_1)\overline{x}^2, \tag{7}$$

$$Lyy = \sum_{l=1}^{k} \int_{u_1}^{u_2} f_l(u)^2 du - k(u_2 - u_1)\overline{y}^2. \tag{8}$$

This theorem tells how to solve for unknown coefficients of a linear regression equation with curve manifolds. The geometric regression of SVR has a similar principle, and the integral value of every local region instead of original data is used as input data for traditional SVR.

The final step is to find an optimal regression model F by the goodness of model fitting to curve manifolds. A geometrization parameter v ($v=1,2,\ldots,n$) is defined as the unified length of every local region, and then $k=n/v$ is the number of divided data groups. The length of every local curve is set to be $3v$, and the central 1/3 part of a local curve is taken as the local region of a curve manifold (called *central cutting*). The central cutting makes neighbor data groups overlapping, but there is no overlapping between neighbor local regions. This optimization problem can be solved by finding an optimal v ($v=5\sim14$ is the proper searching scope as per our experiences) to minimize the residual sums of squares of model fitting:

$$RSS(v)=RSS(k)=Lyy-2\phi Lyx+\phi^2 Lxx, \tag{9}$$

$$v_o = \arg\min RSS(v). \tag{10}$$

With the optimal parameter v_0, the optimal curve manifolds are constructed, and then the optimal regression model F is found with the optimal curve manifolds.

When we do a prediction, the above work is done under a uniform value correlation T for both test and training data. The uniform T is established by using prediction values of Y (other than test and training data of Y), which are derived from the regression model by traditional LR/SVR. Then, under the T, all the data are projected to curve manifolds, and the optimal regression model F is found with the parts of curve manifolds corresponding to training data. For example, the curve manifold Mx is constructed for both test and training data of X according to T. Thus, the test datum x is also projected to x' on the Mx (we say that the x' is on Mx instead of (x', u) on Mx for simplification), and then the x' is used as input of model F and the output (e.g., $y'=\phi_0+\phi x'$) is its prediction.

The experiments on artificial and real data are performed to evaluate prediction performance of GcLearn LR and ε-SVR (SVR with radial basis kernels) methods based on 10-fold cross validation and mean squared errors. For the linear and nonlinear artificial data sets with different Gaussian noise variances from 0.2 to 2.2, the experimental results show that GcLearn LR and ε-SVR reduce prediction errors by about 45% - 80% compared with traditional LR and ε-SVR. For benchmark real-world data sets (the pyrim, servo, auto-price, cpu, and auto-mpg) from UCI database (Blake & Merz, 1998), the experimental results show that GcLearn LR and ε-SVR reduce prediction errors by about 4% - 46% compared with traditional LR and ε-SVR.

3   PERFORMANCE ANALYSIS

For a performance analysis of GcLearn LR and SVR methods compared with traditional LR and SVR, we have theorem 3.1 for the lowest performance and theorem 3.2 for better performance of GcLearn LR and SVR methods when prediction errors are used for performance evaluation. Proofs for both theorems are given in Appendices 7.1 and 7.2 respectively.

The performance of (GcLearn) LR/SVR method is defined to be predictive ability of a regression model by (GcLearn) LR/SVR method. The predictive ability is assessed by prediction errors, the mean squared error between the prediction values of a regression model on test data, and the true values of test data. Thus, the performance analysis refers to the comparison of predictive ability between the GcLearn LR/SVR method and the traditional LR/SVR method.

It is known that a local region U_i of M corresponds to a piece of quadratic polynomial curve C_i: $y=f_i(u)$

($u \in [1,m]$) in yOu, and the local curve $y=f_i(u)$ fits to ith data group with m pairs of $\{y_j,u_j\}$. The projection of a data group to a local region of M is called *local projection*. Corresponding to the minimum parameter $v=1$, the minimal local projection is that every data group with $3v=3$ data is fitted to a local curve C_i, whose central 1/3 part with a length of 1 is taken as local region U_i of M.

Under the minimal local projection, three cases appear as follows. First, the corresponding projection points (y_j,u_j) on U_i/C_i are the same as the original data pair $\{y_j,u_j\}$ (j=1,2,3) of ith data group on account of the principle that a curve is determined by three points in a plane, so the function of the geometrical correlation disappears. Second, the integral value of the central part of curve C_i is approximately equal to original data y_2, so the geometric regression method is almost the same as traditional regression method. Third, the following theorem 3.1 states that GcLearn LR and traditional LR methods give approximately same regression equation.

Theorem 3.1 (Minimal geometric regression) Let there be a linear dependent relation for q 1-dimensional continuous variables. If the geometric regression is performed based on minimal local projection, the resulting regression equation for q variables is approximately the same as that found by the traditional linear regression method.

Therefore, the GcLearn LR and traditional LR methods have almost the same performance under the minimal local projection owing to above three cases or reasons, which indicates the lowest performance of the GcLearn LR and SVR methods. The conclusion is the same for GcLearn SVR under the minimal local projection, as there are same reasons mentioned above, and both the GcLearn SVR and traditional SVR use the same SVR procedure and method to infer a regression model.

The better performance of GcLearn under the larger local projection (v>1) will be discussed as follows. It is known that GcLearn uses inner correlations between test and training data. Both the test and training data are projected to the same curve manifolds, so the test and training data are correlated through the geometrical correlation or curve manifolds. Under larger local projection, the geometrical correlation will take effect, and the test datum x of predictor variable X are projected to x' on the M_x (we say that the x' is on M_x instead of (x', u) on M_x for simplification), and then the x' is used as input of a regression model.

Let there be good inner correlation T between test data and training data (test data are close to some training data under T). It is expected for GcLearn LR that the prediction by GcLearn LR model on a test datum is on the M_z of response variable Z (this is the first condition), which means that this prediction is close to its neighbor training data under T so that this prediction tends to (nearby) its true value. The second condition is that traditional LR method gives the prediction of the same test datum far away from its neighbor training data under T, which means that this prediction is far away from its true value. When the two conditions are satisfied, we can conclude that GcLearn LR method has better prediction performance than traditional LR method.

The following theorem 3.2 states that GcLearn will give its prediction z'_k on M_z under the good inner correlation T even if test datum x_k of X is far away from M_x, which indicates that GcLearn LR satisfies the first condition. Therefore, by adding theorem 3.2 to the second condition, we conclude that GcLearn LR method will have better prediction performance than traditional LR method when good inner correlations are obtained and the predictions by traditional LR method are far away from their neighbor training data under correlation T.

Theorem 3.2 (Geometric projection). Let a linear dependent relation be given for q 1-dimensional continuous variables, each with n+1 data points and a good inner correlation T. Let curve manifolds M_x and M_z be constructed under T for predictor variable X and response variable Z respectively. Let there be a test datum x_k and n training data points $\{x_i\}$ (i=1~n+1, $i \neq k$, n>3) for variable X. If the x_k far away from M_x makes the prediction z_k by traditional linear regression far away from M_z, GcLearn linear regression will give its prediction z'_k on M_z.

The conclusion is the same for GcLearn SVR, as GcLearn projects the curve manifold M_x to M_z by the nonlinear regression model $z=f(x)$ learned from curve manifolds of training data and transforms the x_k deviating from M_x to the x'_k on M_x, so the prediction z'_k by $z=f(x)$ is on M_z.

4   CONCLUSION

The theoretical analysis demonstrates that the GcLearn LR and SVR methods will have better prediction

performance than traditional LR and SVR methods for prediction tasks when good inner correlations are obtained and the prediction results by traditional LR and SVR methods are bad or far away from their neighbor training data under the inner correlations. This conclusion also indicates the application condition of the GcLearn method.

5   ACKNOWLEDGEMENTS

6   REFERENCES

Blake, C. L. & Merz, C. J. (1998) UCI repository of machine learning databases. University of California, Irvine. Retrieved from the WWW, September 24, 2007: http://mlearn.ics.uci.edu/MLRepository.html

Chang, C. C. & Lin, C. J. (2006) LIBSVM: a library for support vector machines. Retrieved from the WWW, September 24, 2007: http://www.csie.ntu.edu.tw/~cjlin/libsvm/

Chen, B. J. Chang, M. W. & Lin, C. J. (2004) Load Forecasting Using Support Vector Machines: A Study on EUNITE Competition. *IEEE Transactions on Power Systems 19,* 1821-1830.

Chen, W. (2001) An introduction to Differential Manifold. Beijing, High Education Press.

Lin, C.J., 2001. Formulations of support vector machines: a note from an optimization point of view. *Neural Computation 13(2)*, 307-317.

Meyer, D., Leisch, F., & Hornik, K. (2003) The Support Vector Machine Under Test, *Neurocomputing, 55*:169-186.

Mo, G. & Liu, K. (2003) *Methodology of function approximation*. Beijing, Science Press.

Sinnakaudan, S. K., Ab-Ghani, A., Ahmad, M. S. S., & Zakaria, N. A. (2006) Multiple Linear Regression Model for Total Bed Material Load Prediction. *Journal of Hydraulic Engineering 132 (5)*, 521-528.

Wagner, M., Adamczak, R., Porollo, A., & Meller, J. (2005) Linear Regression Models for Solvent Accessibility Prediction in Proteins. *Journal of Computational Biology 12(3),* 355-369.

Wang, K., Zhang, J., & Guo, L. (2007) Geometrical correlation learning: a new method to mine geometrical correlation between data for regressions. http://mywebpage.netscape.com/sunice8/sup.htm.

Wolfram Research (2007) Retrieved from the WWW, September 2, 2007: http://mathworld.wolfram.com/LeastSquaresFittingPolynomial.html.

Zhou, W., Zhang, L. Jiao, L., & Jin, P. (2006) Support Vector Regression Based on Unconstrained Convex Quadratic Programming. *ICNC (1)*: 167-174.

7   APPENDICES

7.1 Proof of Theorem 3.1

Proof. Let n groups of data $D_i = \{x_1^i, ..., x_q^i\}$ (i=1,…, n) corresponding to q variable X_1, …, X_q be given and correspond to their inner correlations T={1,2,3,…,n}. Using these data, we can find a linear regression equation of the q variables under least square regression error (or traditional linear regression method):

$$f(X_1, ..., X_q) = 0.$$ (11)

Without losing generality, let $q=2$ and the linear regression equation between response variable Y and variable X be

$$L_1: y = \phi_0 + \phi x.$$ (12)

By the method of least-squares fit, the coefficients of regression equation L_1 may be found with n pairs of data $D_i = \{y_i, x_i\}$ ($i=1, ..., n$):

$$\phi = \frac{\sum_{i=1}^{n} y_i x_i - n\overline{y}\,\overline{x}}{\sum_{i=1}^{n} x_i x_i - n\overline{x}\,\overline{x}},$$ (13)

$$\phi_0 = \overline{y} - \phi\overline{x},$$ (14)

where $\overline{y}$ denotes the mean of y_i and $\overline{x}$ the mean of x_i. Then, we have

$$L_2: y = \phi_0 + \phi x.$$ (15)

Now, we consider the geometric regression method under minimal local projection. Let $k=n$ data divisions of variable X under T be ready under minimal local projection. Then every three neighbors x_{i-1}, x_i, x_{i+1} are fitted to a local quadratic polynomial curve $x = g_i(u) = a_i u^2 + b_i u + c_i$ in the local coordinate system xOu, where correlation coordinates $u \in \{u_0, u_0+1, u_0+2\}$. Then, k local curves are joined to form manifold M_X according to correlation T. Similarly, local quadratic polynomial curve $y = h_i(u)$ and curve manifold M_Y are ready for variable Y.

Being different from the above traditional method of using the statistic of n pairs of data D_i, the geometric regression uses curve manifolds and the integral of curves. Let the same integral region $[u_0, u_0+2]$ be adopted to cover three neighbor data D_{i-1}, D_i and D_{i+1}. As the central cutting of a local curve for overlapping data divisions, the $x_i = g_i(u_0+1)$ may be used as the approximation of any x on $x = g_i(u)$ in the short region $[u_0+0.67, u_0+1.33]$. Similarly, the $y_i = h_i(u_0+1)$ is the approximation of any y on $y = h_i(u)$.

Let every variable be described by $k=n$ local curves. Then the coefficients of regression equation L_1 may be found as per theorem 2.1:

$$\hat{\phi} = Lyx/Lxx,$$ (16)

where

$$Lyx = \sum_{i=1}^{k} \int_{u_0}^{u_0+2} h_i(u) g_i(u) du - k(u_0 + 2 - u_0)\overline{y}\,\overline{x}$$

$$\approx \sum_{i=1}^{k} 2 h_i(u_0+1) g_i(u_0+1) - 2k\overline{y}\,\overline{x}$$

$$= \sum_{i=1}^{k} 2 y_i x_i - 2k\overline{y}\,\overline{x}.$$ (17)

Similarly,

$$Lxx \approx \sum_{i=1}^{k} 2 x_i x_i - 2k\overline{x}\,\overline{x}.$$ (18)

Hence,

$$\hat{\phi} \approx \phi,$$ (19)

$$\hat{\phi}_0 = \overline{y} - \hat{\phi}\overline{x} \approx \phi_0,$$ (20)

$$L_3: Y = \hat{\phi} X + \hat{\phi}_0.$$ (21)

Thus, L_3 is approximately the same as L_2.

7.2 Proof of Theorem 3.2

Proof. The relation between two variables is first discussed, and then more variables added.
(1) Predictor variable X and response variable Z have a linear dependent relation.
Let n pairs of data $D_i=\{x_i, z_i\}$ ($i=1\sim n+1$, $i\neq k$) be given and satisfy the linear regression equation L_1:

$$L_1: z=a_1x+a_0. \tag{22}$$

It is known that good inner correlation T holds for D_i and the corresponding geometrical correlation is represented by M_X and M_Z.

Without losing generality, let test datum x_k rank in the middle of data $\{x_i\}$ ($i=1\sim n+1$) under correlation T. As x_k is far away from M_X, its neighbors, and z_k is obtained by L_1:

$$z_k=a_1x_k+a_0, \tag{23}$$

the z_k is far away from M_Z, its neighbors.

Now, we consider the linear geometric regression. According to T, the $n+1$ data $\{x_i\}$ are projected to curve manifold M_X, where a quadratic polynomial curve C_X is for the neighbors of x_k (where u is the correlation coordinate):

$$C_X: \ x'(u)=\beta_2u^2+\beta_1u+\beta_0 . \tag{24}$$

Then, the corresponding point of x_k on C_X is (x'_k, u_k), or x_k is projected to x'_k:

$$x'_k=x'(u_k)= \beta_2u_k^2+\beta_1u_k+\beta_0 . \tag{25}$$

Thus, the (x_k, u_k) far away from M_X is projected to (x'_k, u_k) on M_X (see Figure 1).

Let the geometric regression project M_X to M_Z or (x, u) to (z, u) by linear regression equation L_2:

$$L_2: z=b_1x+b_0. \tag{26}$$

With (x'_k, u_k) on M_X for L_2, we have

$$z'_k=b_1x'_k+b_0. \tag{27}$$
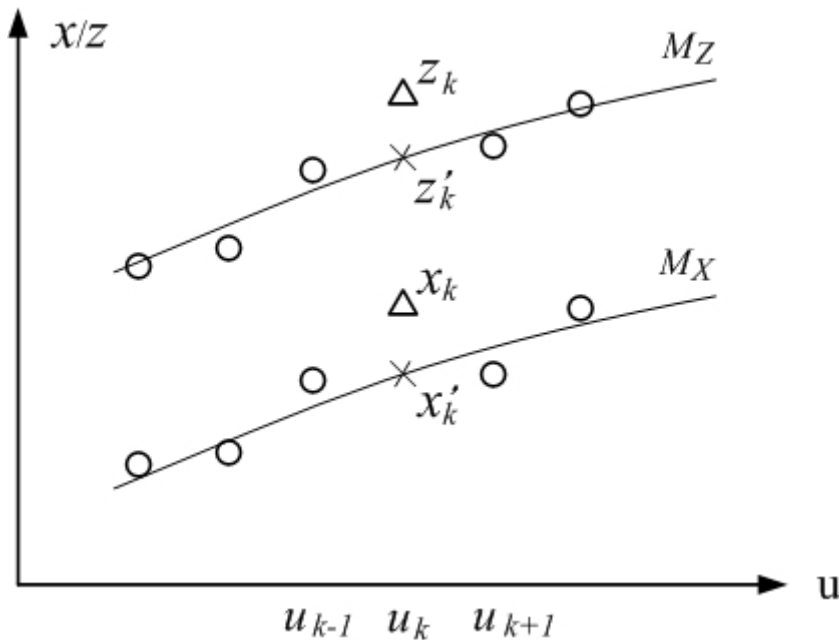
Hence, the (z'_k, u_k) is on M_Z.

Figure 1. x_k far away from M_X is projected to x'_k on M_X, and then the x'_k is projected to z'_k on M_Z.

(2) $q>2$ variables with a linear dependent relation.
Without loss of generality, let predictor variables X and Y and response variable Z have a linear dependent relation, their n groups of training data $D_i=\{x_i, y_i, z_i\}$ ($I=1\sim n+1$, $I\neq k$) satisfy linear regression equation L_3:

$$L_3: z=a_1x+a_2y+a_0. \tag{28}$$

It is known that the prediction z_k is far away from M_Z by L_3 on account of test datum x_k far away from M_X.

i) When only datum x_k is far away from M_X or its neighbor data, the z_k is affected mainly by the x_k. This case is similar to (1) so that we have the same conclusion as (1).

ii) When both a_1 and b_1 are positive (or negative), and both positive x_k and y_k are far away from M_X and M_Y This case is similar to (1) so that we have the same conclusion as (1).

iii) Other cases except i) and ii) for the x_k far away from M_X:
Let the linear geometric regression project M_X and M_Y to M_Z or (x, y, u) to (z, u) by linear regression equation L_4:
$$L_4: z=b_1x+b_2y+b_0. \qquad (29)$$

And let the x_k and y_k be projected to the (x'_k, u_k) and (y'_k, u_k) on curve manifold M_X and M_Y respectively. With x'_k and y'_k for L_4, we then have
$$z'_k= b_1x'_k+ b_2y'_k+b_0. \qquad (30)$$

Hence, the (z'_k, u_k) is on M_Z, and we say that the z'_k is on M_Z for simplification.