Journal of Pure and Applied Mathematics: Advances and Applications
Volume 4, Number 2, 2010, Pages 187-219

LDA AND THE RELATION OF THE RELATIVE CONDITION NUMBER WITH SOME VARIABLES

H. SABERI NAJAFI and A. ZEINAL

Department of Mathematics
Faculty of Sciences
Guilan University
P. O. Box 41335-1914
Rasht
Iran
e-mail: hnajafi@guilan.ac.ir

Abstract

Linear discriminant analysis (LDA) is a well-known method for dimensionality reduction and classification. LDA finds a transformation matrix, which is readily computed by solving a generalized eigenvalue problem. The accuracy of LDA can be measured by the condition number of the generalized eigenvalue problem.

In this paper, we find the relation of the maximum of relative condition numbers related to the generalized eigenvalue problem with some variables.

1. Introduction

Discriminant analysis aims at the classification of an object into one of c given classes based on information from a set of p predictor features. Among the many available methods, the simplest and the most popular

2010 Mathematics Subject Classification: 47A70.

Keywords and phrases: linear discriminant analysis, transformation matrix, generalized eigenvalue problem.

Received September 4, 2010

approach is linear discriminant analysis (LDA) that projects high-dimensional data onto a low-dimensional space, where the data achieves maximum class separability [3, 6, 7].

The derived features in LDA are linear combinations of the original features, where the coefficients are from the transformation matrix. The optimal projection or transformation in classical LDA is obtained by minimizing the within-class distance and maximizing the between-class distance simultaneously, thus achieving maximum class discrimination. It has been applied successfully in many applications [1, 2, 4, 11] including face recognition and microarray gene expression data analysis.

The original LDA formulation, known as the Fisher linear discriminant analysis (FLDA) [5] deals with binary-class classifications. The key idea in FLDA is to look for a direction that separates the class means well (when projected onto that direction) while achieving a small variance around these means. Linear discriminant analysis is optimal if

(a) multivariate normal distribution for the observation in each of the groups;

(b) equal covariance matrix for all groups;

(c) equal prior probability for each group;

(d) equal costs for misclassification.

For evaluating the performance of LDA, apparent classification error rate (APER), expected cost of misclassification (ECM), total probability of misclassification (TPM), optimal error rate (OER), and actual error rate (AER) are used. We want to define a new measure which depends on eigenvectors of a generalized eigenvalue problem. As we know, the columns of the projection matrix in LDA are the eigenvectors of a generalized eigenvalue problem. So a criterion, which could measure the accuracy of the eigenvectors, will be appropriate for evaluating the performance of LDA. Relative condition number is an index, which has this property. Here, we consider some of the variables defined on the data or matrices W and B, and then investigate to see if there is any relations between the relative condition number and the mentioned variables.

2. LDA and Generalized Eigenvalue Problem

A data set of n data vectors in a p-dimensional space is denoted as

$$S = [O_1, O_2, \ldots, O_n] = [U_1, U_2, \ldots, U_c] \in R^{p \times n}, \tag{2.1}$$

where the data is clustered to c classes and each block has data vectors as follows:

$$U_i = [X_{i1}, \ldots, X_{in_i}].$$

Here, we assume that X_{ij} is an observation of a multivariate normal distribution and the class covariance matrices are identical. The between-class scatter matrix B and the within-class scatter matrix W are defined as

$$B = \sum_{i=1}^{c} n_i (\overline{X}_{i.} - \overline{X})(\overline{X}_{i.} - \overline{X})^T,$$

$$W = \sum_{i=1}^{c} \sum_{j=1}^{n_i} (X_{ij} - \overline{X}_{i.})(X_{ij} - \overline{X}_{i.})^T,$$

where

$$\overline{X_{i.}} = [\overline{X}_{i.1}, \ldots, \overline{X}_{i.p}]^T = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}, \tag{2.2}$$

and

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{c} \sum_{j=1}^{n_i} X_{ij}$$

are the centroid of the class i and the global centroid, respectively. The separability of classes in a data set can be measured by using the traces of these scatter matrices. The goal of linear discriminant analysis (LDA) is to find a transformation matrix $G^{p \times l}$ for some integer l with $l \leq p$ that defines a linear transformation

$$G^T : O_i \in R^{p\times l} \rightarrow y = G^T O_i \in R^{l\times 1},$$

and preserves the cluster structure by maximizing the between-class scatter and minimizing the within-class scatter. In the transformed space defined by G^T, the between-class scatter matrix $\widetilde{B}$ and the within-class scatter matrix $\widetilde{W}$ become

$$\widetilde{B} = G^T BG,$$

and

$$\widetilde{W} = G^T WG,$$

respectively. A commonly used criterion in LDA for finding an optimal clustered structure preserving transformation G^T is

$$\max_G trace((G^T WG)^{-1}(G^T BG)).$$

It is well known [6] that this criterion is satisfied when $l = c - 1$ and the columns of $G^{p\times(c-1)}$ are the eigenvectors corresponding to the $c-1$ largest eigenvalue for the following generalized eigenvalue problem,

$$BW = \lambda WV, \tag{2.3}$$

in which λ and V denote the eigenvalue and the eigenvector, respectively. When W is nonsingular, we face the following standard eigenvalue problem,

$$W^{-1}BV = \lambda V.$$

However, as in many applications such as information retrival [8] and face recognition [9], when the number of data items is smaller than the dimension of data space, W becomes singular.

3. Condition Number

Condition number plays an important role in modern numerical linear algebra. It explains the sensitivity of the solution of a problem to perturbation in the data. Relative condition number of a generalized eigenvalue problem is defined as follows.

Let x_i and y_i be, respectively, the right and left eigenvectors of a given eigenvalue λ_i :

$$Ax_i = \lambda_i Bx_i, \quad y_i^* A = \lambda_i y_i^* B,$$

and define the corresponding components,

$$\alpha_i = y_i^* Ax_i / (\|y_i\|_2 \|x_i\|_2), \quad \beta_i = y_i^* Bx_i / (\|y_i\|_2 \|x_i\|_2), \quad \lambda_i = \frac{\alpha_i}{\beta_i}.$$

In [10], it is shown that a perturbation in A and B of relative size ε :

$$\|\delta A\| \le \varepsilon \|A\|_2, \qquad \|\delta B\| \le \varepsilon \|B\|_2$$

may yield a perturbed eigenvalue $\widetilde{\lambda}_i$, but that the chordal distance:

$$\chi(\lambda_i, \widetilde{\lambda}_i) = \frac{|\alpha_i \widetilde{\beta}_i - \widetilde{\alpha}_i \beta_i|}{\sqrt{|\alpha_i|^2 + |\beta_i|^2}\sqrt{|\widetilde{\alpha}_i|^2 + |\widetilde{\beta}_i|^2}},$$

between the original and the perturbed eigenvalue is bounded by

$$\chi(\lambda_i, \widetilde{\lambda}_i) \le \varepsilon \frac{(\|A\|_2^2 + \|B\|_2^2)^{\frac{1}{2}}}{(|\alpha_i|^2 + |\beta_i|^2)^{\frac{1}{2}}} + o(\varepsilon^2),$$

and that there exist perturbations δA and δB, for which this bound is met. The quantity

$$\kappa(\lambda_i) = \frac{(\|A\|_2^2 + \|B\|_2^2)^{\frac{1}{2}}}{(|\alpha_i|^2 + |\beta_i|^2)^{\frac{1}{2}}},$$

is thus a valid relative condition number for λ_i in the sense that it measures how a perturbation of relative size ε in A and B affects λ_i in the chordal metric. We define

$$C = \max\{\kappa(\lambda_i)\}, \tag{3.1}$$

as the condition number of the generalized eigenvalue problem.

4. The Relation of the Condition Number with Other Variables

We want to find the relation of the condition number of (2.3) with some variables defined on the data and on B and W matrices. For this purpose, we want to change the mean, variance, and correlation coefficients values of the features, used in LDA, so that their means and variances will be various and the random mechanism will also be used.

For this purpose, we use the polynomials and some random numbers subject to different distributions. For each of the mean, variance, and correlation coefficient, 10 choices are considered, and therefore $10 \times 10 \times 10 = 1000$ different conditions are made in which, we simulate the multivariate normal distributions. We do this when $n_i = 20$, $c = 4$, $p = 40$. We consider $n_i = 20$, $p = 40$ to follow the work in an undersample condition that makes W singular.

Considering

$$X_{ij} = [x_{ij1}, \ldots, x_{ijp}]^T,$$

as a simulated sample, we define

$$E(x_{ijk}) = m_{ik} \qquad i = 1, \ldots, c; \quad j = 1, \ldots, n_i;$$

$$\operatorname{corr}(x_{ijk}, x_{ijl}) = \rho_{kl} \qquad k = 1, \ldots, p;$$

$$\operatorname{Var}(x_{ijk}) = v_k \qquad l = 1, \ldots, p.$$

Then, we consider the values of v_k, the variance values for k-th features, as follows:

(1) A random number of the uniform distribution in interval (10, 1000).

(2) A random number of the uniform distribution in interval (k, 10k).

(3) A random number of the normal distribution with the mean of 50 and the standard deviation of 10.

(4) A random number of the normal distribution with the mean of $10\sqrt{k}$ and the standard deviation of $\sqrt{k}$.

(5) A random number of the Poisson distribution with the parameter of 20.

(6) A random number of the Poisson distribution with the parameter of $5k$.

(7) $10k + 20$.

(8) $5k^2 + 3k + 10$.

(9) $0.2k^3 + 0.3k^2 + 0.5k + 5$.

(10)

$$f(\frac{k+1}{2}) \qquad k = 1, 3, 5, \ldots, 39,$$

$$f(4 - \frac{k}{2} + 1) \qquad k = 2, 4, 6, \ldots, 40,$$

where

$$f(k) = 3k^2 + 6k + 15.$$

In number 10, we change the polynomials of degree 2 so that correlation coefficient between k and v_k will not be large.

We consider the values m_{ik}, the mean values for k-th features, as follows:

(1) Random numbers of the uniform distributions in intervals (200, 300), (250, 350), (300, 400) and (350, 450).

(2) Random numbers of the uniform distributions in intervals ($50k$, $60k$), ($60k$, $70k$), ($70k$, $80k$), and ($80k$, $90k$).

(3) Random numbers of the normal distributions with means and standard deviations of (500, 10), (800, 15), (1100, 20), and (1400, 25).

(4) Random numbers of the normal distributions with means and standard deviations of $(100\sqrt{k}, 10\sqrt{k})$, $(120\sqrt{k}, 11\sqrt{k})$, $(140\sqrt{k}, 12\sqrt{k})$, and $(160\sqrt{k}, 13\sqrt{k})$.

(5) Random numbers of the Poisson distributions with parameters of 80, 90, 100, and 110.

(6) Random numbers of the Poisson distributions with parameters of $50k$, $60k$, $70k$, and $80k$.

(7) $100k + 50$, $150k + 60$, $200k + 70$, $250k + 80$.

(8)

$$25k^2 + 10k + 100, \quad 35k^2 + 15k + 150,$$

$$45k^2 + 20k + 200, \quad 55k^2 + 25k + 250.$$

(9)

$$10k^3 + 5k^2 + 5k + 50, \quad 11k^3 + 7k^2 + 8k + 70,$$

$$12k^3 + 9k^2 + 11k + 90, \quad 13k^3 + 11k^2 + 14k + 110.$$

(10)

$$f_i\left(\frac{k+1}{2}\right) \qquad k = 1, 3, 5, \ldots, 39;$$

$$f_i\left(4 - \frac{k}{2} + 1\right) \qquad k = 2, 4, 6, \ldots, 40;$$

$$f_1(k) = 90k + 70; \qquad f_2(k) = 120k + 90;$$

$$f_3(k) = 150k + 110; \qquad f_4(k) = 180k + 130.$$

In number 10, we change the polynomial of degree 1 so that the correlation coefficient between k and m_{ik} will not be large.

We consider the values ρ_{kl}, the correlation coefficient between the k-th and l-th features, as follows:

(1) A random number of the uniform distribution in interval (0.05, 0.2).

(2) A random number of the uniform distribution in interval $(0, 0.03\sqrt{k+l})$.

(3) A random number of the uniform distribution in interval $(-(0.004k + 0.0015l), (0.004k + 0.0015l))$.

(4) A random number of the uniform distribution in interval $\left(\frac{-1}{k+l}, \frac{1}{k+l}\right)$.

(5) A random number of the normal distribution with the mean of –0.02 and the standard deviation of 0.005.

(6) A random number of the normal distribution with the mean of 0.2 and the standard deviation of 0.01.

(7) A random number of the normal distribution with the mean of 0.3 and the standard deviation of $0.001(k + l)$.

(8) A random number of the exponential distribution with the parameter of 0.01 multiplied by –1.

(9) A random number of the exponential distribution with the parameter of $\frac{1}{5\sqrt{k+l}}$.

(10) A random number of the exponential distribution with the parameter of $\frac{0.004k + 0.0015l}{2.5}$.

Now, we notice many variables defined on the data and matrices W and B. Corresponding with each simulation of these variables and that of the matrices W and B, we compute the condition number defined in (3.1) by replacing A and B with B and W, respectively. Using the multivariate linear regression, we find the following variables, which are related to ln C.

$$Y_1 = \frac{2}{p(p-1)} \sum_{i=1}^{p} \sum_{j=1}^{p} r_{ij},$$

where

$$r_{ij} = \frac{W_{ij}}{\sqrt{W_{ii}}\sqrt{W_{jj}}},$$

r_{ij} denotes the sample correlation coefficient between the i-th and j-th features. So Y_1 denotes mean of the sample correlation coefficients.

$$Y_2 = \frac{\overline{W_d}}{S_{W_d}},$$

where

$$\overline{W_d} = \frac{1}{p} \sum_{i=1}^{p} W_{ii},$$

and

$$S_{W_d} = \frac{1}{p} \sum_{i=1}^{p} (W_{ii} - \overline{W_d})^2,$$

W_{ii} denotes the variance estimation of the i-th feature. So Y_2 denotes the coefficient of variation for the sample variance of the features.

Considering Corr as the correlation coefficient, we define

$$Y_3 = \mathrm{Corr}(I, W_d),$$

where

$$I = [1, 2, \ldots, p]^T,$$

and

$$W_d = [W_{11}, \ldots, W_{pp}]^T,$$

Y_3 denotes the correlation coefficient between the row number of S, defined by (2.1), and the feature variance of the related row.

$$Y_4 = \frac{S_E}{\overline{\overline{E}}},$$

where

$$\overline{\overline{E}} = \frac{1}{p^2}\sum_{i=1}^{p}\sum_{j=1}^{p} E_{ij},$$

and

$$S_E^2 = \frac{1}{p^2}\sum_{i=1}^{p}\sum_{j=1}^{p}(E_{ij} - \overline{\overline{E}})^2,$$

with

$$E_{ij} = W_{ij} - \min_{i,j} W_{ij},$$

Y_4 denotes the coefficient of variation for the transformed entries of matrix W.

$$Y_5 = \frac{S_D}{\overline{\overline{D}}},$$

where

$$\overline{\overline{D}} = \frac{1}{p^2}\sum_{i=1}^{p}\sum_{j=1}^{p} D_{ij},$$

and

$$S_D^2 = \frac{1}{p^2}\sum_{i=1}^{p}\sum_{j=1}^{p}(D_{ij} - \overline{\overline{D}})^2,$$

with

$$D_{ij} = B_{ij} - \min_{i,j} B_{ij},$$

Y_5 denotes the coefficient of variation for the transformed entries of matrix B.

$$Y_6 = \frac{S_{\overline{E}}}{\overline{\overline{E}}},$$

where

$$S^2_{\overline{E}} = \frac{1}{p}\sum_{i=1}^{p}(\overline{E_{i.}} - \overline{\overline{E}})^2,$$

with

$$\overline{E_{i.}} = \frac{1}{p}\sum_{j=1}^{p}E_{ij},$$

Y_6 denotes the coefficient of variation for the mean of the rows for the transformed W.

$$Y_7 = \frac{S_{S^2}}{\overline{\overline{S^2}}},$$

where

$$\overline{S^2} = \frac{1}{p}\sum_{i=1}^{p}S^2_{W_{i.}},$$

and

$$S^2_{S^2} = \frac{1}{p}\sum_{i=1}^{p}(S^2_{W_{i.}} - \overline{S^2})^2,$$

with

$$S^2_{W_{i.}} = \frac{1}{p}\sum_{j=1}^{p}(W_{ij} - \overline{W_{i.}})^2,$$

where

$$\overline{W_{i.}} = \frac{1}{p}\sum_{j=1}^{p} W_{ij},$$

Y_7 denotes the coefficient of variation subject to coefficients of variation for the entries of each row of W.

$$Y_8 = \frac{1}{p}\sum_{i=1}^{p} V_{E_{i.}},$$

where

$$V_{E_{i.}} = \frac{S_{E_{i.}}}{\overline{E_{i.}}},$$

and

$$S^2_{E_{i.}} = \frac{1}{p}\sum_{j=1}^{p}(E_{ij} - \overline{E_{i.}})^2,$$

Y_8 denotes the mean of coefficients of variation of the entries for each transformed row of W.

$$Y_9 = \frac{S_{\overline{D}}}{\overline{\overline{D}}},$$

where

$$S^2_{\overline{D}} = \frac{1}{p}\sum_{i=1}^{p}(\overline{D_{i.}} - \overline{\overline{D}})^2,$$

and

$$\overline{D_{i.}} = \frac{1}{p}\sum_{j=1}^{p} D_{ij},$$

Y_9 denotes the coefficient of variation for the row mean of the transformed B.

$$Y_{10} = \frac{1}{p}\sum_{j=1}^{p} V_{..j},$$

where

$$V_{..j} = \frac{S_{X_{..j}}}{\overline{X_{..j}}},$$

with

$$\overline{X_{..j}} = \frac{1}{c}\sum_{i=1}^{c} \overline{X_{i.j}},$$

and

$$S^2_{X_{..j}} = \frac{1}{c}\sum_{i=1}^{c} (\overline{X_{i.j}} - \overline{X_{..j}})^2,$$

$\overline{X_{i.j}}$ is defined in (2.2), and Y_{10} denotes the mean of coefficient of variation corresponding with the features so that the coefficient of variation corresponding with each feature is computed through the c value, corresponding with the c class.

Considering

$$V_{ij} = \frac{\sqrt{W_{jj}}}{\overline{X_{i.j}}},$$

we define

$$Y_{11} = \frac{S_{\overline{V}}}{\overline{\overline{V}}},$$

where

$$\overline{\overline{V}} = \frac{1}{c}\sum_{i=1}^{c} \overline{V_{i.}},$$

and

$$S^2_{\overline{V}} = \frac{1}{c}\sum_{i=1}^{c}(\overline{V_{i.}} - \overline{\overline{V}})^2,$$

with

$$\overline{V_{i.}} = \frac{1}{p}\sum_{j=1}^{p} V_{ij}.$$

For each class, the mean of coefficient of variation for the features is computed. Y_{11} denotes the coefficient of variation of the computed values corresponding with the classes.

$$Y_{12} = \frac{1}{c}\sum_{i=1}^{c} S^2_{V_i},$$

where

$$S^2_{V_i} = \frac{1}{p}\sum_{j=1}^{p}(V_{ij} - \overline{V_{i.}})^2.$$

For each class, the variance for the coefficient of variation of the features is computed. Y_{12} denotes the mean of the computed values subject to classes.

For the mentioned simulated data, the range of variation for C and the other variables have been computed as follows:

$$0.01 \le Y_1 \le 0.32, \quad 0.18 \le Y_2 \le 1.12, \quad -0.22 \le Y_3 \le 1.00,$$

$$0.28 \le Y_4 \le 1.18, \quad 0.30 \le Y_5 \le 1.98, \quad 0.01 \le Y_6 \le 0.54,$$

$$0.40 \le Y_7 \le 1.63, \quad 0.47 \le Y_8 \le 1.10, \quad 0.11 \le Y_9 \le 1.12,$$

$$0.05 \le Y_{10} \le 0.35, \quad 0.04 \le Y_{11} \le 0.47, \quad 0.00 \le Y_{12} \le 5.95,$$

$$171.63 \le C \le 2.53 \times 10^3.$$

We first, used the ordinary least square regression, but we faced heteroskedasticity problem. To remove this problem, we used weighted least square regression, which resulted the 18 models.

The models 1 through 17 use both the variables defined on matrices B and W, and the variables defined on the data. To minimize the colinearity, a few variables were entered each time. Some models include constant value and some lack it. Having constant value will bring about a higher tolerance and a lower vif, and lacking it will bring about a bigger eigenvalue and a smaller condition index.

These models are presented in Tables 1 through 17, on pages 205, 206, 206, 207, 207, 208, 208, 209, 209, 210, 210, 211, 211, 212, 212, 213, and 213 respectively.

Now, we want to convey a model, which can only be defined on the entries of B or W matrices. We use the variable Y_5 defined on the entries of B matrix, to present a quadratic model so that the colinearity will be minimized. Regarding [9], $\phi_1(v)$ and $\phi_2(v)$, which are defined as follows, are orthogonal and therefore their colinearity will be minimized.

$$v = \frac{\sqrt{Y_5}}{\max \sqrt{Y_5}} = \frac{\sqrt{Y_5}}{1.41} \qquad 0 < v < 1,$$

$$u = \frac{2v - \min v - 1}{1 - \min v} = \frac{2v - 1.39}{0.61},$$

$$\phi_1(v) = v - \frac{\min v + 1}{2} = v - 0.695,$$

$$\phi_2(v) = \frac{(1 - \min v)^2}{12}(-1 + 3u^2) = \frac{(0.61)^2}{12}(-1 + 3u^2),$$

and therefore the Model 18 on page 214 will be obtained.

In all the above 18 models, we have

$$\text{Tolerance} > 0.25 \qquad \text{VIF} < 4$$

$$\text{Eigenvalue} > 0.01 \qquad \text{Condition Index} < 15.$$

Therefore, the probability of colinearity is low.

For analyzing the residuals, we used the normality test and the charts of residuals. The normality test of residuals were carried out using the Lillifors test and K-stat values, the corresponding statistic values mentioned in tables related to models. The charts of residuals for the 18 above-mentioned models are displayed in Figures 1 through 18 on pages 214, 214, 215, 215, 215, 215, 216, 216, 216, 217, 217, 217, 218, 218, 218, 219, 219 and 219, respectively.

5. Conclusion

We found out that there is a relationship between the condition number and some variables. These variables have been defined either on the data or on the matrices W and B. The positive and negative coefficients of variables indicate increasing and decreasing effect of the variable values on the condition number, respectively. So, these variables could be considered as the criteria for measuring the performance of LDA. Some of them are controllable, that is, by following some principles, we can control their values to decrease the condition number.

The positive coefficients for the variables $\frac{1}{Y_8}$ and $\frac{1}{Y_{12}}$ as well as those of the variables $Y_1, Y_2, Y_3, Y_4, Y_5, Y_6, Y_7, Y_9, Y_{10}$, and Y_{11} in models indicate the magnitude of their values will increase the index C. Then, the magnitude of the values of variables Y_8 and Y_{12} indicates the high accuracy of LDA and the other variables magnitude represents the low accuracy of LDA.

Noticing the decreasing effect of the variable Y_2, Y_3, Y_1, Y_{10}, and Y_{11} on the accuracy of LDA, we should respectively observe what follows. When choosing the features in LDA, it is necessary to select the features, whose variances are close to one another. When the data matrix is made, we should allocate the features to the rows so that the linear correlation coefficient between the feature variance and the row number is small. We should choose the features so that their correlation coefficient mean is small. We should choose the features in LDA so that their means in c

class are close to one another. We should choose the features so that the mean of coefficient of variation corresponding with classes are close to each other.

Noticing the increasing effect of the variable Y_{12} on the accuracy of LDA, we should select the features so that their coefficient of variation are far from each other.

None of the above-mentioned models is a general model, but according to them, we can simulate more to achieve general models.

The Model 18 can result in presenting a general model in which the performance of LDA is evaluated only based on the changes of matrix B, and there is no need to notice the existing changes in data for evaluating the performance of LDA. The other advantage of this model is that, it can lead us to a general model for each generalized eigenvalue problem $AV = \lambda BV$, in which the changes in A entries determine the condition number.

6. Tables and Figures

The tables and figures of the 18 models follow the references. The figures include the histograms of the residuals accompanied by the normal fit and the charts of residuals against the predicted values. The tables include an independent variable and dependent ones related to each model. They also include non-standard coefficient, standard coefficient, t, significance, tolerance, vif related to the independent variables, and eigenvalue and condition index related to the dimensions. The results indicate that the above-mentioned models are fairly appropriate.

References

[1] P. N. Belhumeour, J. P. Hespanha and D. J. Kriegman, Eigenfaces vs Fisherfaces: Recognition using class specific linear projection, IEEE Trans. Pattern Analysis and Machine Intelligence 19(7) (1997), 711-720.

[2] L. Chen, H. Liao, M. K. J. Lin and G. Yu, A new LDA-based face recognition system which can solve the small sample size problem, Pattern Recognition 33 (2000), 1713-1726.

[3] R. O. Duda, P. E. Hart and D. J. Stork, Pattern Classification, Wiley, New York, 2001.

[4] S. Dudoit, J. Fridly and T. P. Speed, Comparison of discrimination methods for the classification of tumors using gene expression data, Journal of the American Statistical Association 97 (2002), 77-87.

[5] R. A. Fisher, The use of multiple measurements in taxonomic problems, Annals of Eugenics 7 (1936), 179-188.

[6] K. Fukunaga, Introduction to Statistical Pattern Classification, Academic Press, 1990.

[7] T. Hasti, R. Tibshirani and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference and Prediction, Springer, 2001.

[8] G. Kowalski, Information Retrieval System: Theory and Implementation, Kluwer, Academic Publishers, 1997.

[9] M. Shacham and N. Brauner, Minimizing the effect of colinearity in polynomial regression, Ind. Eng. Chem. Res. 36 (1997), 4405-4412.

[10] G. W. Stewart, Perturbation Theory for the Generalized Eigenvalue Problem, Academic Press, New York, 1978.

[11] D. L. Swets and J. Weng, Using discriminant eigenfeatures for image retrieval, Pattern Analysis and Machine Intelligence 18 (1996), 831-836.

Table 1. Model 1

Dep Var:$(\ln C)^{1.05}$					
Model	Non-standard Coeff.		Standard Coeff.		t
	B	Std. Err.	Beta	Std. Err.	
Constant	0.352	0.165	− − −	− − −	2.140
Y_7	1.382	0.108	0.093	0.007	12.748
Y_9	17.646	0.148	0.873	0.007	119.560
Y_{11}	17.169	0.356	0.352	0.007	48.254
Sig	Tol	VIF	Dim	Eig	Cond
0.033	− − −	− − −	1	3.665	1.000
0.000	1.000	1.000	2	0.148	4.974
0.000	0.993	1.008	3	0.145	5.035
0.000	0.992	1.008	4	0.042	9.314
K-stat	Sig	R-square	Adj R-square	F	Sig
0.0265	0.0944	0.947	0.947	5.973×10^3	0.000

Table 2. Model 2

Dep Var:$(\ln C)^{1.32}$					
Model	Non-standard Coeff.		Standard Coeff.		t
	B	Std. Err	Beta	Std. Err	
Constant	– 6.738	0.477	– – –	– – –	–14.132
Y_7	3.247	0.294	0.082	0.007	11.040
Y_5	28.575	0.221	0.995	0.008	129.290
Y_{10}	56.294	1.280	0.338	0.008	43.964
Sig	Tol	VIF	Dim	Eig	Cond
0.000	– – –	– – –	1	3.602	1.000
0.000	1.000	1.000	2	0.230	3.961
0.000	0.938	1.066	3	0.132	5.224
0.000	0.938	1.066	4	0.032	9.870
K-Stat	Sig	R-Square	Adj R-Square	F	Sig
0.0278	0.0628	0.945	0.945	5.669×10^3	0.000

Table 3. Model 3

Dep Var:$(\ln C)^{1.119}$					
Model	Non-standard Coeff.		Standard Coeff.		t
	B	Std. Err	Beta	Std. Err	
Constant	0.357	0.177	– – –	– – –	2.015
Y_3	0.988	0.099	0.074	0.007	9.951
Y_9	22.728	0.193	0.875	0.007	117.683
Y_{11}	21.836	0.466	0.349	0.007	46.894
Sig	Tol	VIF	Dim	Eig	Cond
0.044	– – –	– – –	1	3.333	1.000
0.000	1.000	1.000	2	0.462	2.685
0.000	0.993	1.007	3	0.145	4.790
0.000	0.992	1.008	4	0.059	7.522
K-Stat	Sig	R-Square	Adj R-Square	F	Sig
0.0284	0.0525	0.945	0.945	5.739×10^3	0.000

Table 4. Model 4

Dep Var:$(\ln C)^{1.15}$					
Model	Non-standard Coeff.		Standard Coeff.		t
	B	Std. Err	Beta	Std. Err	
Y_3	1.075	0.106	0.033	0.003	10.145
Y_9	25.316	0.172	0.716	0.005	147.475
Y_{11}	24.002	0.400	0.290	0.005	59.940
Sig	Tol	VIF	Dim	Eig	Cond
0.000	0.574	1.741	1	2.413	1.000
0.000	0.255	3.926	2	0.441	2.338
0.000	0.256	3.902	3	0.145	4.077
K-Stat	Sig	R-Square	Adj R-Square	F	Sig
0.0283	0.0551	0.994	0.994	5.521×10^4	0.000

Table 5. Model 5

Dep Var:$(\ln C)^{1.14}$					
Model	Non-standard Coeff.		Standard Coeff.		t
	B	Std. Err	Beta	Std. Err	
Y_1	5.658	0.552	0.034	0.003	10.259
Y_9	24.511	0.167	0.713	0.005	147.129
Y_{11}	23.382	0.390	0.291	0.005	59.979
Sig	Tol	VIF	Dim	Eig	Cond
0.000	0.528	1.893	1	2.456	1.000
0.000	0.253	3.955	2	0.399	2.481
0.000	0.253	3.955	3	0.145	4.113
K-Stat	Sig	R-Square	Adj R-Square	F	Sig
0.0284	0.0521	0.994	0.994	5.579×10^4	0.000

Table 6. Model 6

Dep Var:$(\ln C)^{1.24}$					
Model	Non-standard Coeff.		Standard Coeff.		t
	B	Std. Err	Beta	Std. Err	
Y_1	8.236	0.794	0.039	0.004	10.409
Y_5	19.178	0.140	0.682	0.005	137.118
Y_{11}	34.059	0.528	0.327	0.005	64.514
Sig	Tol	VIF	Dim	Eig	Cond
0.000	0.536	1.865	1	2.421	1.000
0.000	0.301	3.320	2	0.406	2.443
0.000	0.289	3.456	3	0.174	3.735
K-Stat	Sig	R-Square	Adj R-Square	F	Sig
0.0228	0.2445	0.993	0.993	4.444×10^4	0.000

Table 7. Model 7

Dep Var:$(\ln C)^{1.14}$					
Model	Non-standard Coeff.		Standard Coeff.		t
	B	Std. Err	Beta	Std. Err	
Constant	1.644	0.190	− − −	− − −	8.635
Y_1	5.787	0.612	0.074	0.008	9.458
Y_5	13.374	0.121	0.869	0.008	110.544
Y_{11}	24.193	0.516	0.369	0.008	46.907
Sig	Tol	VIF	Dim	Eig	Cond
0.000	− − −	− − −	1	3.329	1.000
0.000	1.000	1.000	2	0.420	2.820
0.000	0.995	1.005	3	0.176	4.356
0.000	0.995	1.005	4	0.065	7.152
K-Stat	Sig	R-Square	Adj R-Square	F	Sig
0.0263	0.0979	0.939	0.939	5.093×10^3	0.000

Table 8. Model 8

Dep Var:$(\ln C)^{1.14}$					
Model	Non-standard Coeff.		Standard Coeff.		t
	B	Std. Err	Beta	Std. Err	
Constant	1.209	0.216	− − −	− − −	5.589
Y_2	2.041	0.201	0.077	0.008	10.129
Y_5	13.280	0.116	0.873	0.008	114.605
Y_{11}	23.774	0.522	0.347	0.008	45.537
Sig	Tol	VIF	Dim	Eig	Cond
0.000	− − −	− − −	1	3.592	1.000
0.000	1.000	1.000	2	0.195	4.290
0.000	0.992	1.008	3	0.160	4.732
0.000	0.992	1.008	4	0.052	8.300
K-Stat	Sig	R-Square	Adj R-Square	F	Sig
0.0276	0.0671	0.943	0.943	5.461×10^3	0.000

Table 9. Model 9

Dep Var:$(\ln C)^{1.1}$					
Model	Non-standard Coeff.		Standard Coeff.		t
	B	Std. Err	Beta	Std. Err	
Constant	2.254	0.164	− − −	− − −	13.729
Y_6	4.916	0.481	0.078	0.008	10.217
Y_5	11.490	0.102	0.872	0.008	113.203
Y_{11}	20.774	0.457	0.350	0.008	45.414
Sig	Tol	VIF	Dim	Eig	Cond
0.000	− − −	− − −	1	3.274	1.000
0.000	0.999	1.001	2	0.493	2.576
0.000	0.992	1.008	3	0.167	4.422
0.000	0.991	1.009	4	0.065	7.107
K-Stat	Sig	R-Square	Adj R-Square	F	Sig
0.0243	0.1722	0.941	0.941	5.329×10^3	0.000

Table 10. Model 10

Dep Var:$(\ln C)^{1.08}$					
Model	Non-standard Coeff.		Standard Coeff.		t
	B	Std. Err	Beta	Std. Err	
Constant	0.883	0.153	− − −	− − −	5.770
Y_6	4.583	0.426	0.080	0.007	10.761
Y_9	19.726	0.167	0.875	0.007	117.901
Y_{11}	19.107	0.403	0.351	0.007	47.358
Sig	Tol	VIF	Dim	Eig	Cond
0.000	− − −	− − −	1	3.305	1.000
0.000	0.999	1.001	2	0.490	2.596
0.000	0.993	1.007	3	0.145	4.770
0.000	0.992	1.008	4	0.059	7.488
K-Stat	Sig	R-Square	Adj R-Square	F	Sig
0.0256	0.1211	0.946	0.945	5.771×10^3	0.000

Table 11. Model 11

Dep Var:$(\ln C)^{1.14}$					
Model	Non-standard Coeff.		Standard Coeff.		t
	B	Std. Err	Beta	Std. Err	
Y_6	5.648	0.503	0.035	0.003	11.222
Y_9	24.529	0.164	0.714	0.005	149.616
Y_{11}	23.474	0.503	0.292	0.005	61.345
Sig	Tol	VIF	Dim	Eig	Cond
0.000	0.602	1.660	1	2.387	1.000
0.000	0.256	3.901	2	0.468	2.258
0.000	0.258	3.882	3	0.145	4.055
K-Stat	Sig	R-Square	Adj R-Square	F	Sig
0.0265	0.0929	0.994	0.994	5.685×10^4	0.000

Table 12. Model 12

Dep Var:$(\ln C)^{1.24}$					
Model	Non-standard Coeff.		Standard Coeff.		t
	B	Std. Err	Beta	Std. Err	
Y_6	8.051	0.727	0.038	0.003	11.077
Y_5	19.199	0.138	0.682	0.005	138.949
Y_{11}	34.236	0.518	0.329	0.005	66.127
Sig	Tol	VIF	Dim	Eig	Cond
0.000	0.609	1.642	1	2.353	1.000
0.000	0.305	3.283	2	0.473	2.229
0.000	0.297	3.367	3	0.174	3.680
K-Stat	Sig	R-Square	Adj R-Square	F	Sig
0.0230	0.2339	0.993	0.993	4.502×10^4	0.000

Table 13. Model 13

Dep Var:$(\ln C)^{1.1}$					
Model	Non-standard Coeff.		Standard Coeff.		t
	B	Std. Err	Beta	Std. Err	
Constant	1.160	0.249	− − −	− − −	4.658
Y_4	2.851	0.383	0.060	0.008	7.438
Y_5	11.564	0.108	0.867	0.008	107.182
Y_{11}	21.097	0.460	0.371	0.008	45.865
Sig	Tol	VIF	Dim	Eig	Cond
0.000	− − −	− − −	1	3.665	1.000
0.000	0.999	1.001	2	0.181	4.501
0.000	0.995	1.005	3	0.125	5.424
0.000	0.995	1.005	4	0.030	11.132
K-Stat	Sig	R-Square	Adj R-Square	F	Sig
0.0221	0.2920	0.935	0.935	4.791×10^3	0.000

Table 14. Model 14

Dep Var:$(\ln C)^{1.15}$					
Model	Non-standard Coeff.		Standard Coeff.		t
	B	Std. Err	Beta	Std. Err	
Constant	– 1.284	0.289	– – –	– – –	– 4.449
Y_4	3.354	0.432	0.058	0.008	7.773
Y_9	25.452	0.219	0.877	0.008	116.225
Y_{11}	24.296	0.528	0.347	0.008	46.009
Sig	Tol	VIF	Dim	Eig	Cond
0.000	– – –	– – –	1	3.709	1.000
0.000	0.999	1.001	2	0.145	5.050
0.000	0.993	1.008	3	0.117	5.628
0.000	0.992	1.008	4	0.028	11.429
K-Stat	Sig	R-Square	Adj R-Square	F	Sig
0.0282	0.0562	0.944	0.944	5.566×10^3	0.000

Table 15. Model 15

Dep Var:$(\ln C)^{1.22}$					
Model	Non-standard Coeff.		Standard Coeff.		t
	B	Std. Err	Beta	Std. Err	
Constant	– 1.428	0.245	– – –	– – –	– 5.833
$\frac{1}{Y_{12}}$	0.001	0.000	0.076	0.007	10.598
Y_9	33.069	0.268	0.887	0.007	123.286
Y_{11}	29.354	0.671	0.316	0.007	43.751
Sig	Tol	VIF	Dim	Eig	Cond
0.000	– – –	– – –	1	3.199	1.000
0.000	0.986	1.014	2	0.607	2.295
0.000	0.987	1.013	3	0.135	4.873
0.000	0.978	1.023	4	0.059	7.354
K-Stat	Sig	R-Square	Adj R-Square	F	Sig
0.0253	0.1290	0.949	0.949	6.189×10^3	0.000

Table 16. Model 16

Dep Var:$(\ln C)^{1.09}$					
Model	Non-standard Coeff.		Standard Coeff.		t
	B	Std. Err	Beta	Std. Err	
Constant	2.601	0.148	− − −	− − −	17.540
$\frac{1}{Y_{12}}$	0.001	0.000	0.104	0.007	14.412
Y_5	11.116	0.091	0.885	0.007	122.205
Y_{11}	18.885	0.431	0.317	0.007	43.644
Sig	Tol	VIF	Dim	Eig	Cond
0.000	− − −	− − −	1	3.162	1.000
0.000	0.981	1.019	2	0.619	2.261
0.000	0.982	1.018	3	0.154	4.538
0.000	0.976	1.025	4	0.065	6.951
K-Stat	Sig	R-Square	Adj R-Square	F	Sig
0.0268	0.0860	0.949	0.948	6.132×10^3	0.000

Table 17. Model 17

Dep Var:$(\ln C)^{1.1}$					
Model	Non-standard Coeff.		Standard Coeff.		t
	V	Std. Err	Beta	Std. Err	
Constant	1.059	0.279	− − −	− − −	3.792
$\frac{1}{Y_8}$	1.329	0.200	0.054	0.008	6.660
Y_5	11.562	0.108	0.867	0.008	106.609
Y_{11}	21.049	0.462	0.370	0.008	45.533
Sig	Tol	VIF	Dim	Eig	Cond
0.000	− − −	− − −	1	3.681	1.000
0.000	1.000	1.000	2	0.180	4.517
0.000	0.995	1.005	3	0.116	5.639
0.000	0.995	1.005	4	0.022	12.798
K-Stat	Sig	R-Square	Adj R-Square	F	Sig
0.0215	0.3292	0.935	0.935	4.738×10^3	0.000

Table 18. Model 18

Dep Var:$(\ln C)^{0.00001}$					
Model	Non-standard Coeff.		Standard Coeff.		t
	B	Std. Err	Beta	Std. Err	
Constant	1.000	0.000	– –	– –	5×10^{7}
ϕ_1	6×10^{-6}	0.000	1.016	0.020	49.628
ϕ_2	-6×10^{-6}	0.000	– 0.199	0.020	– 9.741
Sig	Tol	VIF	Dim	Eig	Cond
0.000	------	------	1	1.908	1.000
0.000	0.508	1.969	2	0.840	1.507
0.000	0.508	1.969	3	0.252	2.749
K-Stat	Sig	R-Square	Adj R-Square	F	Sig
0.0242	0.1741	0.788	0.788	1.85×10^{3}	0.000

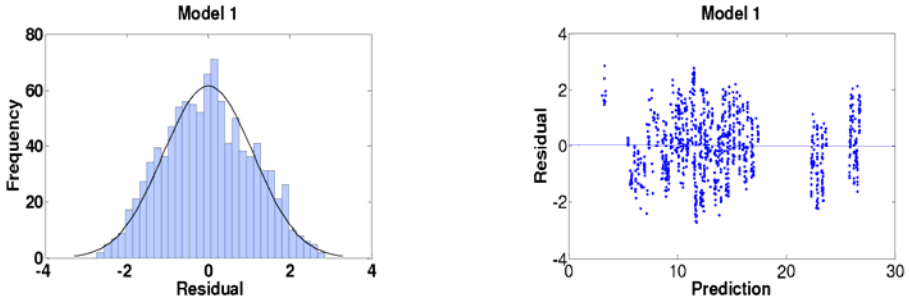

Figure 1. Model 1.

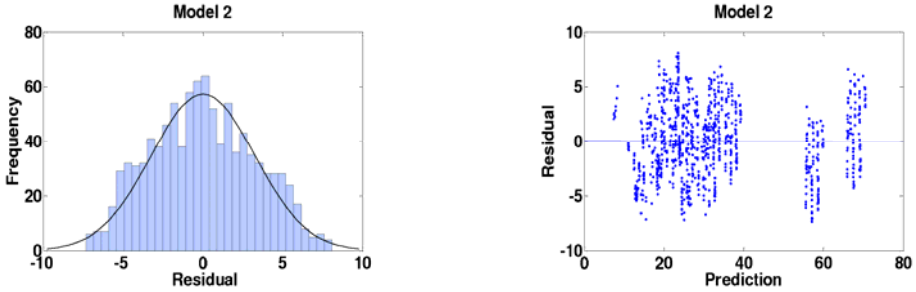

Figure 2. Model 2.

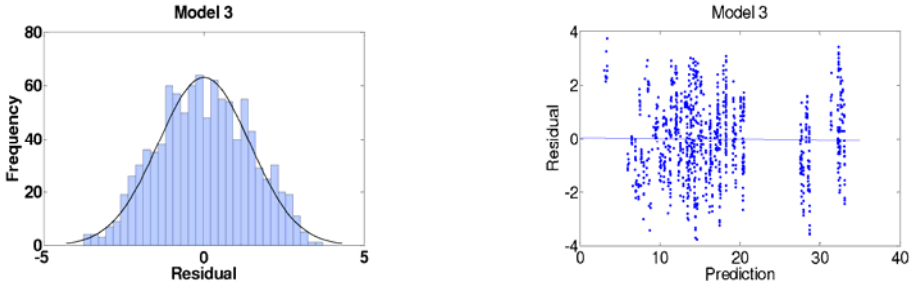

Figure 3. Model 3.

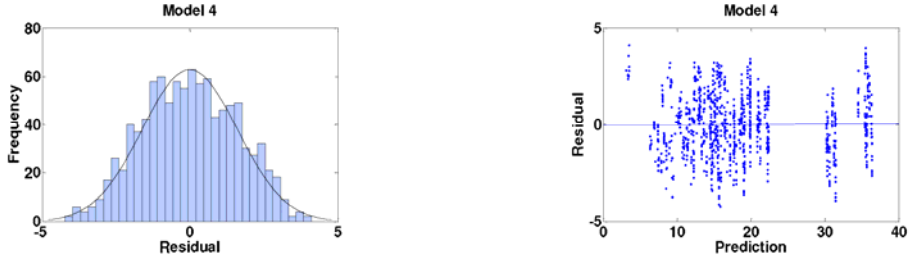

Figure 4. Model 4.

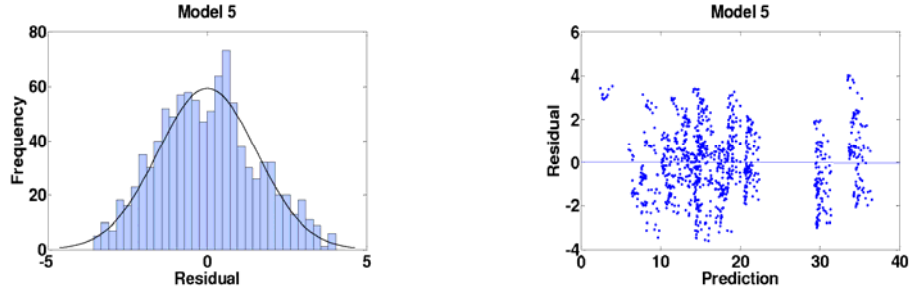

Figure 5. Model 5.

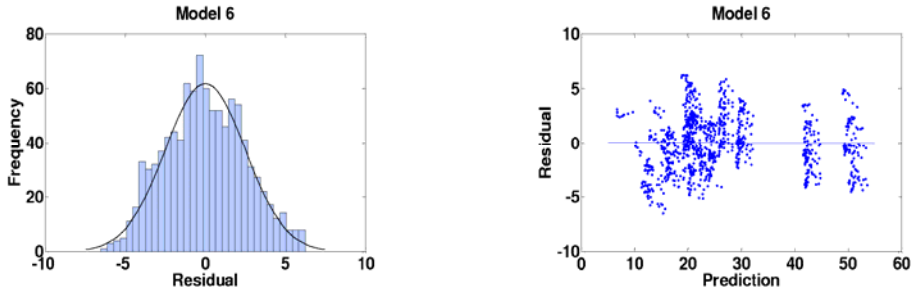

Figure 6. Model 6.

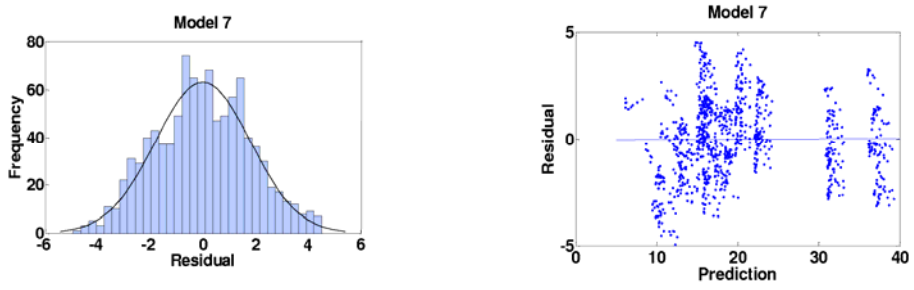

Figure 7. Model 7.

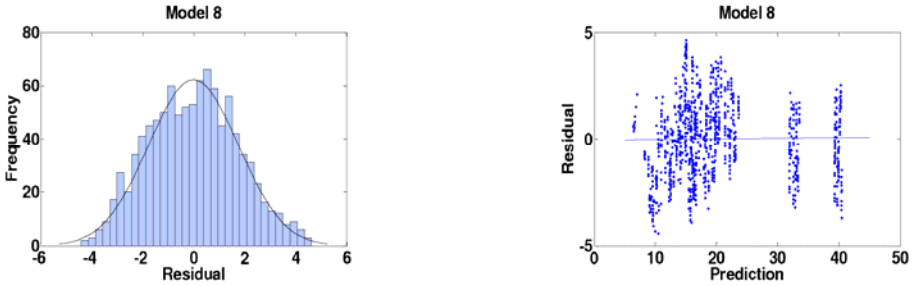

Figure 8. Model 8.

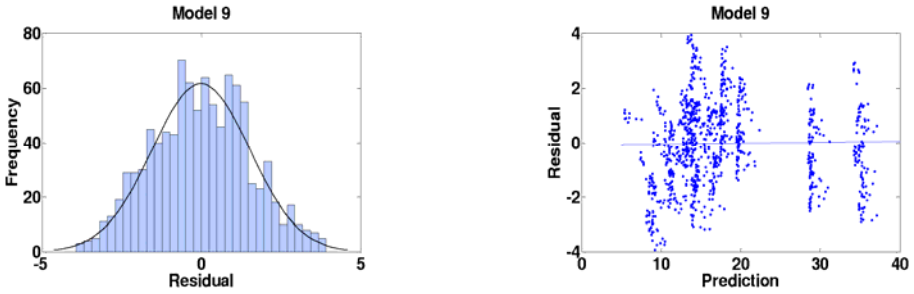

Figure 9. Model 9.

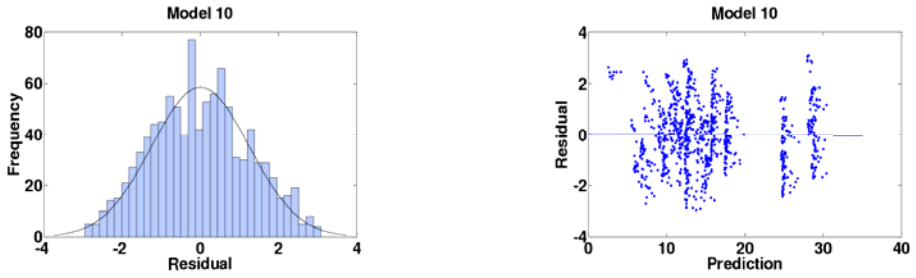

Figure 10. Model 10.

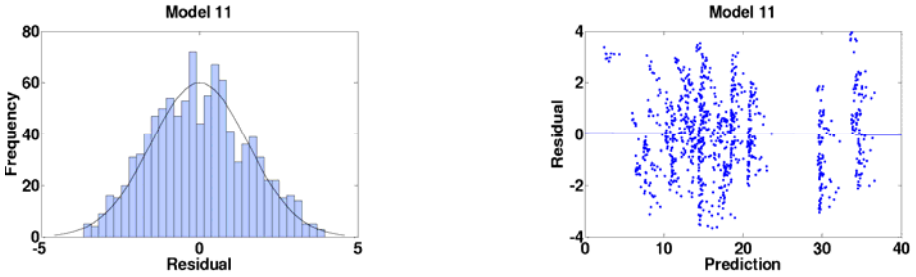

Figure 11. Model 11.

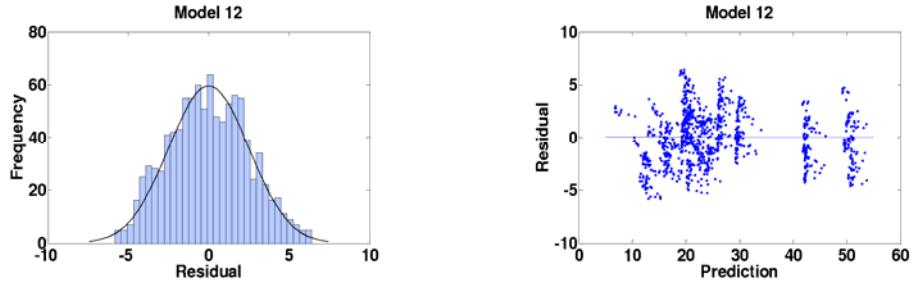

Figure 12. Model 12.

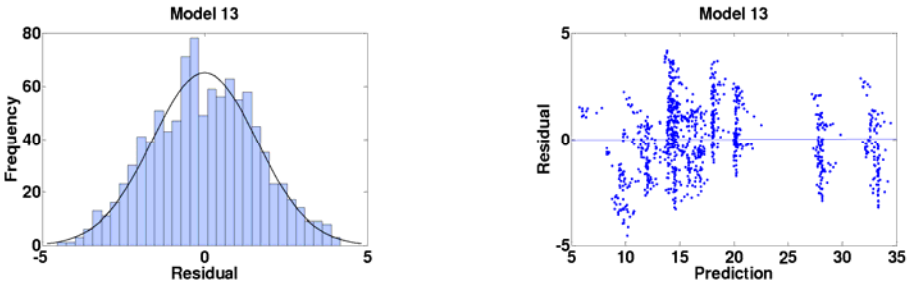

Figure 13. Model 13.

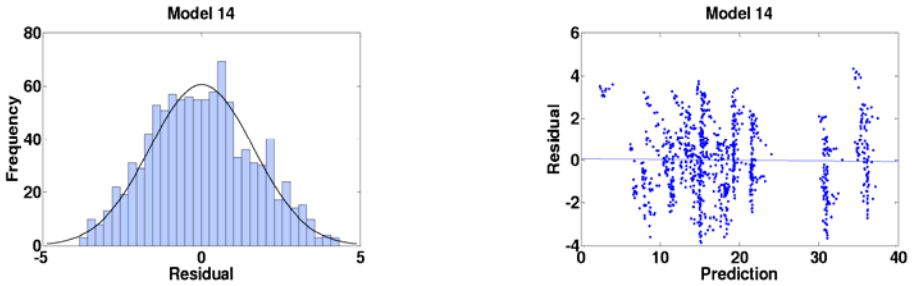

Figure 14. Model 14.

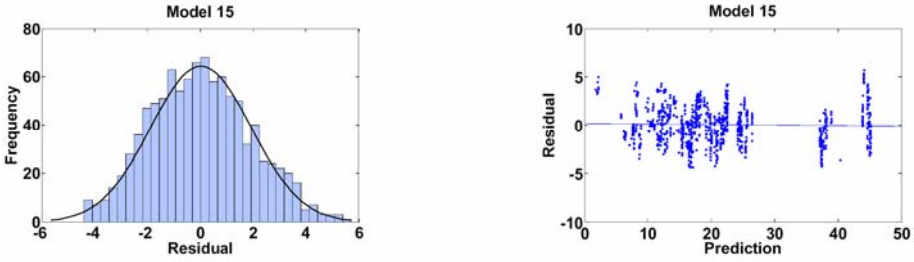

Figure 15. Model 15.

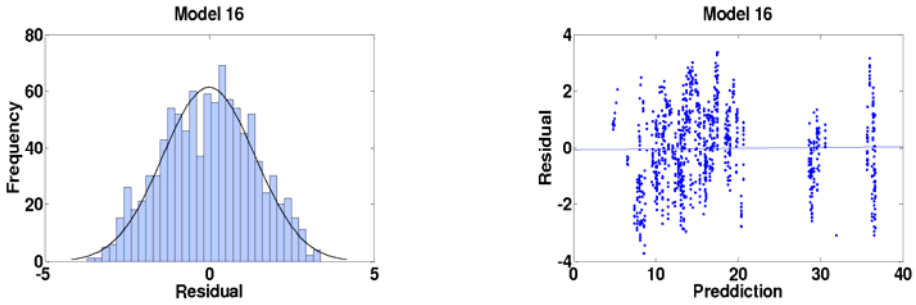

Figure 16. Model 16.

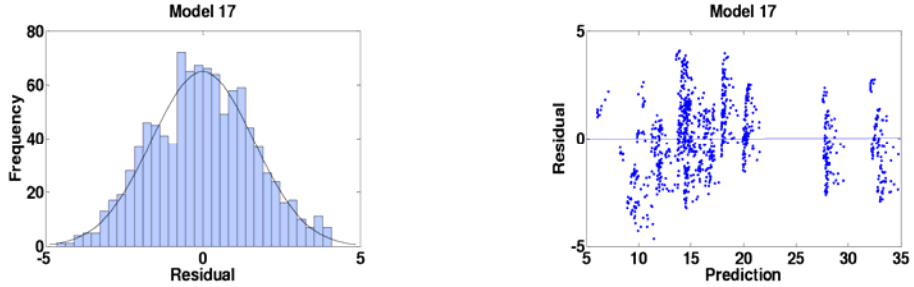

Figure 17. Model 17.

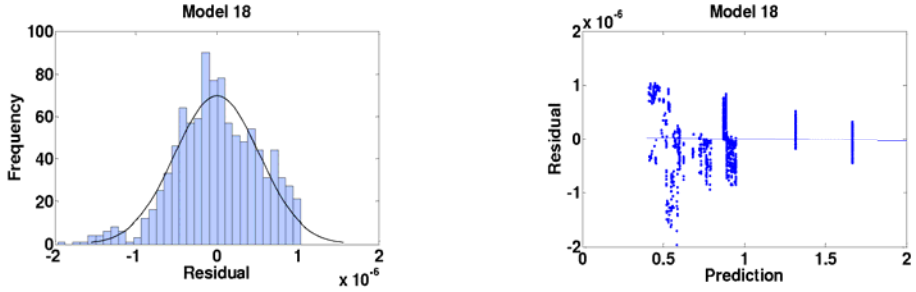

Figure 18. Model 18.