Int. J. Engng Ed. Vol. 10, No. 4, pp. 325–330, 1994
Printed in Great Britain.

0742-0269/91 $3.00+0.00

Ridge Regression: An Application to College Admission*

S. HASANUDDIN AHMAD
MOHAMMAD S. ALJIFFRY
MUSTAFA M. ALIDRISI
Department of Industrial Engineering, King Abdulaziz University, PO Box 9027, *Jeddah*-21413, *Saudi Arabia*

In causal forecasting models multiple regression is one of the popular ones. However, due to high multicollinearity between the regressor variables multiple regression model parameters become unstable. One of the alternate techniques to overcome such problems is ridge regression. Ridge regression is not a very old find. Very few applications have so far been shown. While trying to set-up criterion for admission to an engineering programme based on some science course grades in high school examinations, we found that physics, chemistry and mathematics course grades were highly correlated. We then successfully used a ridge regression model to estimate the required GPA at college level. We also derived a relation which should be satisfied in order to be sure with a 95% degree of confidence that a student will maintain a GPA of at least 2.0. Using our ridge regression model and relation as mentioned above we considered some cases of students scores in high school science courses to study the possibility of achieving a higher GPA in college, i.e. at least 2.0.

INTRODUCTION

A PROBLEM often encountered in the use of multiple regression is the effect of multicollinearity on the regression model.

Often in multiple regression problems, the independent or regressor variables x_j are intercorrelated. In situations where this intercorrelation is very large, we say that multicollinearity exists. Multicollinearity can have serious effects on the estimates of the regression coefficients and on the general applicability of the estimated model.

The estimates of the regression coefficients are very imprecise when multicollinearity is present. If the prediction of new observations requires extrapolation, we generally would expect to obtain poor results. Extrapolation usually requires good estimates of the individual model parameters.

Multicollinearity arises for several reasons. It will occur when the analyst collects the data such that a constraint of the form $\Sigma_{j=1}^{k} a_j x_j = 0$ holds among the columns of the x matrix (the a_j are constants, not all zero). For example, if regressor variables are the components of a mixture, then such a constraint will always exist because the sum of the components is always constant. Usually, these constraints do not hold exactly and the analyst does not know that they exist.

There are several ways to detect the presence of multicollinearity. The reader interested in these may refer to Hocking *et al.* [1].

Several remedial measures have been proposed for resolving the problem of multicollinearity. Augmenting the data with new observations specifically designed to break up the approximate linear dependencies that currently exist is often suggested. However, sometimes this is impossible for economic reasons or because of physical constraints that relate the X_j. Another possibility is to delete certain variables from the model. This suffers from the disadvantage of discarding the information contained in the deleted variables.

Since multicollinearity primarily affects the stability of the regression coefficients, it would seem that estimating these parameters by some method that is less sensitive to multicollinearity than ordinary least squares would be helpful. Several methods have been suggested for this. Hoerl and Kennard [2] have proposed ridge regression as an alternative to ordinary least squares.

Good discussions of the practical use of ridge regression are in Johnson [3], Marquardt and Snee [4] and Vinod [5]. We have [6] in a study of setting up admission standards for the College of Engineering, King Abdulaziz University, encountered very high multicollinearity between mathematics, physics and chemistry. This problem was overcome by the use of ridge regression. Thus, in this paper an application of ridge regression to the problem of setting up a standard of admission based on the performance of students in high school mathematics, physics and chemistry is shown. Using ridge regression coefficients a standard of admission has been proposed in order to

* Paper accepted 12 December 1993.

help students achieve a minimum GPA of 2.0, as required by the college regulations.

RIDGE REGRESSION

In ridge regression, the parameter estimates are obtained by solving

$$b^*(k) = (X'X + kI)^{-1}X'Y \qquad (1)$$

where $I \geqslant 0$ is an identity matrix. To obtain ridge coefficients b^*s one must specify the constant k where $k \geqslant 0$. The ridge estimator $b^*(k)$ is not an unbiased estimate of b, as is the ordinary least squares estimator $\hat{b}$, but the mean square error of $b^*(k)$ will be smaller than the mean square error of $\hat{b}$. Thus the ridge regression seeks to find a set of regression coefficients that is more 'stable', in the sense of having a small mean square error. Since multicollinearity usually results in ordinary least-squares estimators that may have extremely large variances, ridge regression is suitable for situations where the multicollinearity problems exist.

To obtain the ridge regression estimator from Eq. 1, one must specify a value for the constant k. Generally, there is an 'optimum' k for any problem, but the most simple approach is to solve Eq. 1 for several values of k in the interval $0 \leqslant k \leqslant 1$. Then a plot of the values of $b^*(k)$ against k is constructed. This display is called the ridge trace. The appropriate value of k is chosen subjectively by inspection of the ridge trace. Typically, a value for k is chosen such that relatively stable parameter estimates are obtained. In general, the variance of $b^*(k)$ is a decreasing function of k, while the squared bias $[b - b^*(k)]^2$ is an increasing function of k. Choosing the value of k involves trading off these two properties of $b^*(k)$. However, we can find optimum k by using the following formula [2].

$$k < \frac{\text{MSE}}{\hat{b}'\hat{b}} \qquad (2)$$

where MSE is the mean squared error of the regression equation and $\hat{b}$ is the estimated vector of regression coefficients of the model.

DATA COLLECTION

Applied types of projects, such as the one under consideration, do need historical data. The sources of our data are within the university. It was collected, edited and prepared in the proper format for analysis.

At the first stage high school data was collected from the academic affairs department for the students whose computer numbers start with 80 up to the students whose computer number starts with 85. In all, approximately 2000 files of students were consulted and the appropriate data were recorded and entered into a data file, thus accumulating a raw data set of 1736 students records. Out of these students records a random sample of 270 were selected for our analysis for 99% accuracy. All the columns, except the last one, in Table 1 give a sample of such sample data.

Table 1. Sample of scaled high school and college core GPA

Student number	Science subject			GPA
	Mathematics	Physics	Chemistry	
1	88	60	71	1.80
2	72	62	63	2.30
3	87	80	80	2.50
4	99	90	88	3.35
5	94	76	90	3.20
6	86	71	59	3.67
7	71	66	74	2.23
8	86	77	79	3.31
9	78	82	82	2.07
10	72	62	63	2.30

After data checking and editing was completed, each student mean grade point in college core courses was calculated as follows:

$$\text{GPA} = \frac{\text{Weighted total of college core grades}}{\text{Total of college core credit hours}}$$

$$\text{GPA} = \frac{\Sigma(Y_i)(n_i)}{\Sigma n_i}$$

where GPA is the college core grade point average, Y_i is the grade in course i and n_i is the number of credit hours for the course i. The GPA appears in the last column of Table 1.

ANALYSIS

Correlation ratios

A quick look at Table 2 shows that grades in mathematics, physics and chemistry are highly correlated with college GPA. Also, for example, Fig. 1 suggests a linear fit through the origin. Our physical phenomenon also confirms this assumption. A student who scores a failing grade in x_i, i.e. ith course does not obtain admission to the college. Hence, the question of GPA ($= y$) does not arise. In other words $x_i \leqslant 49 \Rightarrow \text{GPA} = 0$, as the minimum pass grade is 50.

Thus, in each case a regression line without intercept was set up and found to be a very good fit. Table 2, for the sample, gives the necessary characteristics of regression lines $E[Y|x_j]$.

Multiple regression

We have already discussed in the previous section the reasons of fitting a regression line without an intercept in the case of one variable. The same logic holds in the case of multiple regression as well. A student who fails in mathematics, physics and chemistry (all three of them) will be denied admission to the college. Hence, the question of a non-zero GPA of college core courses for an X

Table 2. Regression characteristics of physics, chemistry and mathematics

Subject and symbol	Regression coefficient (intercept zero)	Correlation ratio	*F*-statistics (all highly significant)
Physics (x_1)	0.6931	0.954	2737.93
Chemistry (x_2)	0.6887	0.947	2361.07
Mathematics (x_3)	0.6238	0.956	2918.73

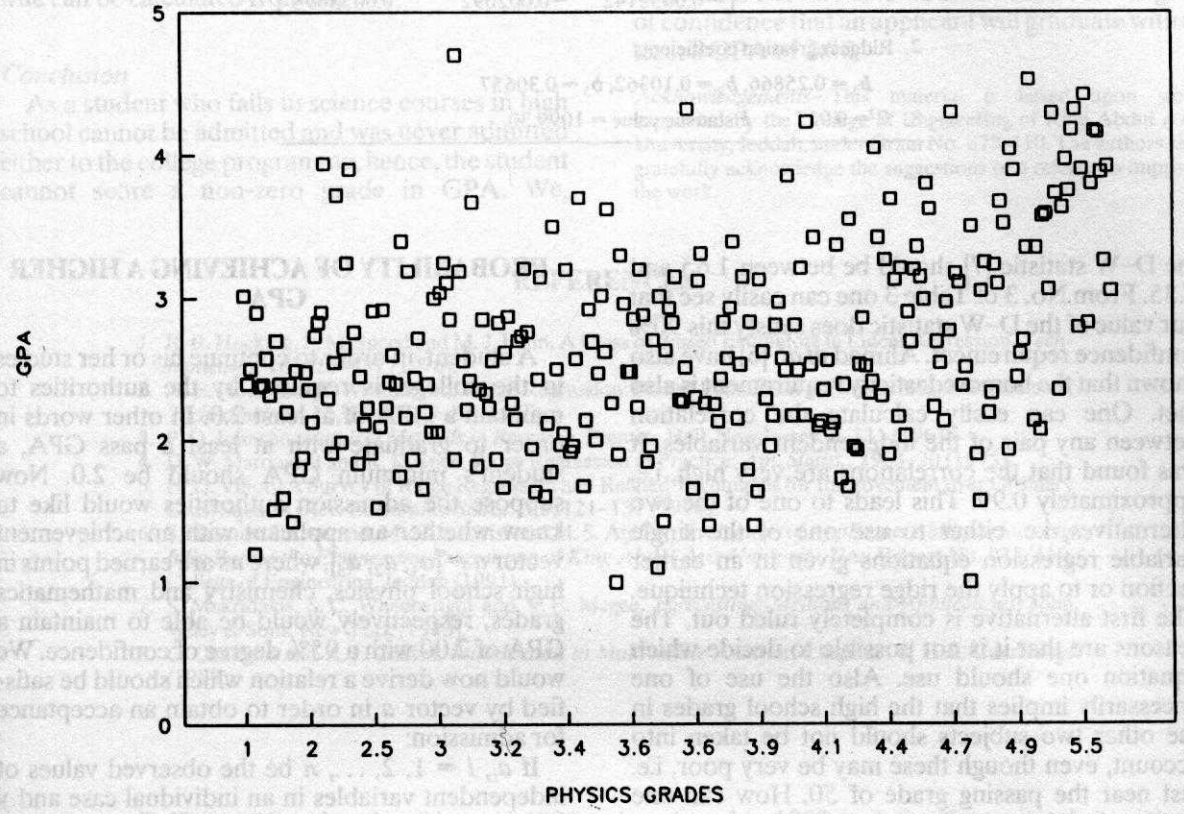

Fig. 1. Scattergram of physics grade *v* GPA.

which contains at least one component $x_i < 50$ does not arise. In other words

$$X' = [x_1, x_2, x_3] \rightarrow \text{GPA} = 0$$

where at least one x_i is $\leqslant 49$.

Because of this actual life physical phenomenon we omit the intercept parameter b_0 from our model. Thus, the model under study with y being the core courses GPA was

$$y = b_1x_1 + b_2x_2 + b_3x_3 + e \qquad (3)$$

where the x_i are already explained in Table 2 and e is random error.

Before analysing the data of Table 1 we converted these to an earned points system. A score of 50 means 1 earned point. A score of 54 means 1.4 earned point. Thus, a vector of scores [60, 70, 85] will become [2, 3, 4.5].

Table 3 gives some characteristics of our multiple regression model based on the earned points sytem. One can easily see the estimates of regression coefficients given in No. 1 of Table 3. Thus, our model based on earned points of data as shown in Table 1 and Eq. 3 is derived as

$$y = 0.26023\,x_1 + 0.10019\,x_2 + 0.30840\,x_3 \quad (4)$$

Discussion on problems

A multiple regression model may face one or more of the following problems;

1. linearity in coefficients;
2. independence of residuals;
3. homosedasticity;
4. multicollinearity.

We did not face problems of linearity in coefficients. It is easily seen that our linear model in Eq. 4 gives a very high value of R^2, i.e. 0.925. To test the independence of residual one usually computes a Durban–Watson statistic. For a 90% confidence

Table 3. Multiple regression model characteristics

1. Regression parameters estimates
   $b_1 = 0.26023, b_2 = 0.10019, b_3 = 0.30840$
2. MSE $= \hat{s}^2 = 0.57008 = s^2$
3. Durban–Watson (D–W) statistic value = 1.759
4. n = number of students considered = 270
5. $k = 3.33$
6. $(X'X + kI)^{-1} = \begin{bmatrix} 0.0068392 & -0.003177 & -0.0032142 \\ -0.003177 & 0.0062727 & -0.002692 \\ -0.0032142 & -0.002692 & 0.0053698 \end{bmatrix}$
7. Ridge regression coefficients
   $b_1 = 0.25866, b_2 = 0.10362, b_3 = 0.30657$
8. $R^2 = 0.925$ F statistic value = 1099.30

the D–W statistic [7] should be between 1.65 and 2.35. From No. 3 of Table 3 one can easily see that our value of the D–W statistic does satisfy this 90% confidence requirement. Ahmad *et al.* [6] have also shown that the homosedasticity requirement is also met. One can easily calculate the correlation between any pair of the independent variables. It was found that the correlations are very high, i.e. approximately 0.99. This leads to one of the two alternatives, i.e. either to use one of the single variable regression equations given in an earlier section or to apply the ridge regression technique. The first alternative is completely ruled out. The reasons are that it is not possible to decide which equation one should use. Also the use of one necessarily implies that the high school grades in the other two subjects should not be taken into account, even though these may be very poor, i.e. just near the passing grade of 50. How can one justify admitting a student whose grade vector, say is [50, 50, 80] or [1, 1, 4] conversion to earned points after scaling? Moreover, in spite of a good correlation between two variables one can still use both of them in the model [8]. This then may result in high values of variances, i.e. a problem of multicollinearity.

Ridge regression model

In order to overcome the problem of multicollinearity one applies the ridge regression technique. The technique is discussed in detail earlier. The estimates of parameters are found using Eq. 1. A value of k which stabilizes the estimates of parameters b_i is to be found. This is obtained as 3.33, given in No. 5 of Table 3. Thus, we can estimate stable b_is at $k = 3.33$. Using values given in No. 7 of Table 3, we get

$$y = 0.25866x_1 + 0.10362x_2 + 0.30657x_3 \quad (5)$$

where x_i are earned points.

This is the equation which would be used for prediction of GPA.

PROBABILITY OF ACHIEVING A HIGHER GPA

A student, in order to continue his or her studies in the college, is required by the authorities to maintain a GPA of at least 2.0. In other words in order to graduate with at least a pass GPA, a student's minimum GPA should be 2.0. Now suppose the admission authorities would like to know whether an applicant with an achievement vector $a = [a_1, a_2, a_3]$, where as are earned points in high school physics, chemistry and mathematics grades, respectively would be able to maintain a GPA of 2.00 with a 95% degree of confidence. We would now derive a relation which should be satisfied by vector a in order to obtain an acceptance for admission:

If a_i, $i = 1, 2, \ldots, n$ be the observed values of independent variables in an individual case and y be the value to be predicted then a t-distribution [7] can be used to calculate the probability of obtaining y. The formula is given by

$$P[y \geqslant \hat{y}] = P[t_{n-p-1} \geqslant (y - \hat{y})/[Var(y - \hat{y})]^{1/2}] \quad (6)$$

where $\hat{y}$ is the estimated value from Eq. 5 and $\text{Var}(y - \hat{y}) = s^2[1 + a'(x'x + kI)^{-1}a]$.

Observe that for Eq. 6 to be $\geqslant 0.95$ with our large sample size n ($\geqslant 200$), the value of t is -1.645.

Now substituting $y = 2$ in the regression relation Eq. 5 for $\hat{y}$, and values of s^2 and $(X'X + kI)^{-1}$ from Table 3 into Eq. 6 and squaring the relation within the bracket we obtain

$$1.034a_1 + 0.413a_2 + 1.226a_3 - 0.056a_1^2 - 0.001a_2^2 - 0.086a_3^2 - 0.063a_1a_2 - 0.169a_1a_3 - 0.072a_2a_3 \leqslant 2.457 \quad (7)$$

Let us consider a couple of cases to see if these can be accepted for admission with a degree of confidence of 95% that the applicants will graduate with a minimum GPA of 2.0.

Case 1: $a = [4.5, 5.0, 5.0]$, and the left side of Eq. 7 is 2.52, hence one cannot be 95% confident about this case.

Case II: $a = [5, 4.5, 5]$ and the left side of Eq. 7 is 2.35, hence one can be 95% confident that an applicant having an achievement vector a in high school physics, chemistry and mathematics will graduate with at least a GPA of 2.0.

Also note that if one claims that he or she would achieve a GPA of 2 with their hard work in spite of the fact that $\hat{y} < 2$, the probability of this claim to be true can be calculated from Eq. 6.

Conclusion

As a student who fails in science courses in high school cannot be admitted and was never admitted either to the college programme, hence, the student cannot score a non-zero grade in GPA. We, therefore, deleted the intercept parameter from our analysis.

Since we found that mathematics, physics and chemistry were highly correlated, our multiple regression equation could not be used for prediction due to instability of the parameters. We then estimated ridge parameter k and re-estimated regression parameters which give stable parameters. This new equation is used for prediction.

We also derived a relation which should be satisfied if one wants to be sure with a 95% degree of confidence that an applicant will graduate with at least a GPA of 2.00.

Acknowledgements—This material is based upon work supported by the College of Engineering of King Abdul Aziz University, Jeddah, under Grant No. 078-410. The authors also gratefully acknowledge the suggestions of a referee to improve the work.

REFERENCES

1. R. R. Hocking, F. M. Speed and M. J. Lynn, A Class of Biased Estimators in Linear Regression, *Technometrics*, **18**, 425–437 (1976).
2. A. E. Hoerl and R. W. Kennard, Ridge Regression: Biased Estimation of Non-orthogonal Problems, *Technometrics*, **12**, 55–67, 69–82 (1970).
3. P. O. Johnson, *Statitical Methods in Research*, Prentice Hall, New Jersey, p. 331 (1961).
4. D. W. Marquardt and R. D. Snee, Ridge Regression in Practice. *Am. Statist.*, **29**, 3–20 (1975).
5. H. D. Vinod, A Survey of Ridge Regression and Related Techniques for Improvement over Ordinary Least Squares, *Rev. Economics Statist.*, **60**, 121–131 (1978).
6. S. Hasanuddin Ahmad, M. M. Alidrisi and M. S. Aljiffry, *Evaluation of Applicants Ability to Successfully Pursue the Engineering Programme of King Abdul Aziz University*, Res. Report No. 078/410, College of Engineering, Jeddah (1991).
7. S. Makridakis, S. C. Wheelwright and V. E. Mcgee, *Forecasting: Methods and Applications*, John Wiley & Sons, New York (1983).
8. C. Chatfield and A. J. Collins, *Introduction to Multivariate Analysis*, Chapman & Hall, Cambridge (1989).

S. Hasanuddin Ahmad was born in 1938 and graduated in 1958 from the University of Karachi, Pakistan. He received his first master's degree in 1960 from the same university and his second master's degree in 1963 from Arizona State University. He joined the Department of Mathematics, at the University of Karachi as an assistant professor. He received his PhD in 1970 in industrial engineering from Arizona State University and rejoined the University of Karachi. He became associate professor in 1973. Later he joined the University of Nigeria. He has been a member of several academic bodies of the Universities in Pakistan and Nigeria. He is now an associate professor in the College of Engineering, King Abdul Aziz University. He is the author of two books on mathematics and has published a number of papers in the field of reliability, network and systems in various journals such as *IEEE Transactions on Reliability*, *Microelectronics and Reliability* and the *International Journal of System Science*, etc.

Mohammed Sadiq Al Jiffry was born in Mekkah, Saudi Arabia in 1952. He received a BSc degree in industrial engineering from Oklahoma State University at Stillwater, Oklahoma, USA. He worked as a graduate and research assistant in 1977, 1978 and 1984 at the school of Industrial Engineering and Management during his postgraduate study. He served as a chairman of the Industrial Engineering Department at the College of Engineering at King Abdulaziz University, in Jeddah, Saudi Arabia from 1985 to 1989. Dr Jiffry obtained many research grants at the university and national levels. He is also the author of many publications on the topics of operational research, forecasting and other industrial engineering topics in local as well as international journals and conferences. He is also curently co-authoring four books on industrial engineering related topics. Currently Dr Jiffry is an associate professor in industrial engineering and the general supervisor of the King Abdulaziz University Services Department.

Mustafa M. Alidrisi was born in Saudi Arabia in 1952. He received a BSc degree in systems engineering from KFUPM in Dharan, Saudi Arabia in 1974 and an MSc and PhD in industrial and operations engineering from the University of Michigan in Ann Arbor in 1976 and 1981,

respectively. He served as the chairman of the Industrial Engineering Department at the College of Engineering at the King Abdul Aziz University in Jeddah, Saudi Arabia from 1983 to 1985 and the vice dean of the College of Engineering for Research and Academic Affairs from 1985 to 1989. Dr Alidrisi has worked as a consultant for many leading organizations in the Kingdom of Saudi Arabia. He has also obtained many research grants at the national as well as the university level. He is also the author of many publications (in international journals) on reliability, optimization, decision analysis and other industrial engineering topics. Currently he is associate professor of industrial engineering and editor-in-chief of King Abdul Aziz University *Journal of Engineering Sciences*.