TEACHING LINEAR MODELS BASED ON OPERATIONS RESEARCH IN STATISTICAL EDUCATION

Dong Q. Wang
Victoria University of Wellington, New Zealand
Bingshu (Nancy) Wu
Ballarat Grammar School, Australia
znw@bgs.vic.edu.au

Teaching the topic of linear models is a complex process. A new teaching process will be investigated which we will consider in two different ways: on one hand linear models and quadratic programming problems are formulated and solved by statistical methods; on the other hand the solution of the linear regression model with constraints makes use of the simplex methods of linear and quadratic programming.

INTRODUCTION

Teaching the topic of linear models is a complex process, a new teaching process will be investigated in statistics education. We will consider in two different ways: on one hand linear models and quadratic programming problems are formulated and solved by statistical methods, and on the other hand the solution of the linear regression model with constraints makes use of the simplex methods of linear or quadratic programming. The interaction between teaching Linear Models and Operations Research (OR) are explored by both statistical and operations research theories in statistics education. The aim of this paper is that students taking linear model courses as well as a course in a linear or non-linear programming (NLP) of Operations Research will realize that there is a definite connection between these problems.

These interactions are considered both from a statistical point of view and from an optimization point of view. A linear regression model is solved by a two-phase version of the simplex method and a statistical algorithm for solving quadratic programming problem is presented. In comparison with the nonlinear programming methods for solving quadratic programming problem, the latter has the following advantages:

(a) Statistics courses often form a core portion in most information science degree programs at Bachelor level. The algorithm based on basic statistical concepts is easy to understand, learn and apply.

(b) Some of the steps of the algorithm are included as built-in functions or procedures in most of the commonly used software packages like *SAS*, *R*, *MAPLE*, *MATHEMATICA* in statistic education.

(c) The algorithm avoids the usage of slack and artificial variables. Some examples of teaching linear models are given to illustrate the ideas.

There are a few algorithms to solve optimization problems for non-linear programming, for example, geometric programming is approximated by a generalization of the arithmetic-geometric mean inequality to solve algebraic non-linear models; Algorithms for solving special forms of non-linear models have been developed, especially for quadratic programming (see Wang, Chukova and Lai, 2005 and Wang, Chukova, Lai, 2004). In this paper we will focus on quadratic programming (or non-linear) problems.

Quadratic programming is concerned with the problem of minimizing (or maximizing) a quadratic objective function subject to linear inequality (or equality) constraints with non-negative values for unknown variables. There are a few approaches to solve QP problems, for example, simplex combinatorial approaches (see Best and Ritter, 1988); ellipsoidal approaches (see Chung and Murty, 1981); interior points methods (see Mehrotra and Sun, 1990) and Ben-Daya and AI-Sultan (1997) developed an exterior point algorithm for quadratic programming based on a penalty function approach using a single Newton method (see Ferrez, Fukuda and Liebling, 2005). This paper will discuss some approaches for solving QP problems based on a statistical point of view in statistical education.

For optimization problems of quadratic programming, we will consider the optimum value (maximum or minimum) of a function Z ($\beta 0$, $\beta 1$, …, βp) of p+1 parameters $\beta 0$, $\beta 1$, …, βp.

Due to the fact that max Z= −min Z, we need only consider minimization problems. A typical mathematical programming problem (MPP) consists of a single objective function, representing either a profit to be maximized or a cost to be minimized, and a set of constraints that circumscribe the decision variables. In the following sections, the relationships are given concerning the multivariate linear models; then a few approaches are presented to teach linear models; and finally some examples are given to show how to use these approaches in statistics education.

RELATIONSHIP BETWEEN LINEAR MODELS AND QP MODELS

Consider a quadratic programming problem QP (1)

Minimise $Z(\beta) = b'\beta + \frac{1}{2}\beta' D\beta$ .... QP(1)

subject to $A\beta \geq C$; and $\beta \geq 0$

where, for $k \leq p+1$, $A\{k\times(p+1)\}$, $D\{(p+1)\times(p+1)\}$ are matrices, $C\{k\times 1\}$, $\beta\{(p+1)\times 1\}$, $b\{(p+1)\times 1\}$ are column vectors, rank(A) = k and D is a symmetric, positive definite matrix, which can be decomposed as D=L'L (Note, matrices D and A have a well defined structure, with D often being a covariance matrix of random variables in statistical analysis), where L is a real upper triangular matrix with positive diagonal elements and can be obtained by the Choleski decomposition of D. The optimal value $Z(\beta)$ and solutions of QP(1) can be found based on Wolfe's method (see Wolfe 1959) which used the Kuhn-Tucker Necessary and Sufficient conditions (NSC), i.e., any point satisfying the NSC is solved for quadratic programming QP(1) through steps (see, e.g., Brown and Goodall, 1959).

In a multivariate linear model (MLM) we consider the squares of differences between the predicted and observed values and add up these squared differences across all the predictions, we get a number called the sum of squared errors (SSE). From a statistical point of view we want the SSE to be as small as possible, i.e. to minimize SSE with the constraints of non-negative variables β as QP(2).

Minimise $SSE(\beta) = \sum \varepsilon^2 = (Y - X\beta)'(Y - X\beta)$ .... QP(2)

where the ε are the residuals. A general least squares linear regression problem (i.e. Regression Programming (RP)) is obtained as QP(3) as follows:

Minimise $Q(\beta) = (Y - X\beta)'(Y - X\beta)$ .... QP(3)

subject to $A\beta \geq C$; and $\beta \geq 0$

where $\beta \in R^{p+1}$ is the unknown and nonnegative vector, and for $n \geq p+1$ and $k \leq p+1$, $X\{n\times(p+1)\}$, $A\{k\times(p+1)\}$, are constant matrices, and $Y\{n\times 1\}$, $C\{k\times 1\}$, $\varepsilon\{n\times 1\}$ are column vectors. Moreover $\varepsilon \sim N(0, \varepsilon^2 I)$, $X'X \geq 0$ and rank(A) = k. The solution of the linear regression with constraints (LRWC) is a subject of the Karush-Kuhn-Tucker theorem.

Let the column vector X = L be the explanatory variable and the column vector $Y = -\frac{1}{2}(L')^{-1}b$ be taken as the response variable in multivariate regression analysis QP(3). We know that quadratic programming is concerned with minimizing a convex quadratic function subject to linear inequality constraints. The unique solution of a quadratic programming problem QP(1) exists provided that the feasible region is non-empty (the QP has a feasible space), and the relative minimum optimal value is also a global optimal value. Since $\{\beta : \beta \in R_+^{p+1}, A\beta \geq C\}$ is a closed convex set, then $\hat{\beta}$ is a unique optimal solution for model QP(2). The relationships are given between multivariate linear model and operations research by Wang (2005); Wang also gave the following theorem to show the relationship among the three models QP(1), QP(2) and QP(3).

Let $\hat{\beta}^* = (\hat{\beta}_0^*, \hat{\beta}_1^*, \hat{\beta}_2^*, ..., \hat{\beta}_p^*)^T$ be the least squares estimators without constraints in a linear regression model QP(2) and $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_p)^T$ be the least squares estimators in the model QP(3) (or QP(1)), both $\hat{\beta}^T$ and $\hat{\beta}^{*T} \in R_+^{p+1}$, then we have

(i) The relationship of optimal values between QP(1) and QP(3) is given by

$$Z(\hat{\beta}) = Q(\hat{\beta}) - \frac{1}{4} b' D^{-1} b$$

(ii) The relationship of the optimal solutions of QP(2) and QP(3) is

$$\hat{\beta} = \left[I - (X'X)^{-1}A'H^{-1}A\right]\hat{\beta}^{*} + (X'X)^{-1}A'H^{-1}C$$

where $H = A(X'X)^{-1}A'$. The details of the full proof are given by Wang (2005).

TEACHING LINEAR MODELS BASED ON QP MODELS

We will apply the relationship as above to teach the QP(1), QP(2) and QP(3).

Teaching Linear Models QP(2) Using the Phase Method

Considering QP(2) we differentiate Q(β) with respect to each of the β and set equal to zero, giving the normal equations in multivariate linear regression. For given observation X and Y, the simplex method is applied to the initial table, the solutions of a simple linear model are obtained based on the operations research methods.

Using Linear Models to Teach the QP(1) Z (β)

Solving QP(1) based on the linear model algorithms, from QP(1) we used slack variable vector S{k×1} for constraints to get standard form, and QP(1) is written as a linear program. The algorithm of the theorem may be used to solve the QP(1) Z(β): firstly solve the goal programming problem (with unknown variables S and d) to find the constants d; secondly use the relationship $\beta = D^{-1}A'd - D^{-1}b$ between β and d to get the optimal values β, then finally the optimal solution is found for QP(1).

Using Least Squares to Teach the QP(3)

A stepwise algorithm for searching for the solution to a QP(1) is explored on the basis of statistical theory (see Wang, 2005). It is shown that quadratic programming can be reduced to an appropriately formulated QP(3) with equality constraints and nonnegative variables. This approach allows us to obtain a simple algorithm to solve a QP model. The applicability of the suggested algorithm is illustrated with some numerical examples in research papers (see Wang, Chukova and Lai, 2005). Quadratic programming with zero-one variables is provided by Wang (2005); the problem can be reduced to search the extreme points of a zonotope.

EXAMPLES

Example 1: Simple linear regression:

Consider the example with 2 dimensions and sample size n=10 observations {(x_i, y_i), i=1, 2, ..., 10} which was given by Wang (2005). Using the SSE to minimize SSE(β) with respect to β, then QP(2) is solved using simplex methods (a special method of operations research) in teaching linear models.

Example 2: Multivariate linear models:

This example was given by Fang, Wang and Wu (1982) for n = 9, and p+1=6. In order to manufacture concrete, asphalt, big and small stones, crushed stones, grit, sands and rock powder are required. Denote these variables as *X*0, *X*1, …, *X*5 respectively. Different types of sieves are used to filter these elements and the mixture, *Y*, of them. The percentage of each of these variables that pass through the sieves are set to be *x*0, *x*1, …, *x*5 and y respectively. Using β0, β1, …, β5 to represent the percentage of these six elements in the mixture, they should satisfy β ≥ 0. According to the science of architecture, the total passing rate should fall into the given range and the closer to the middle point *Y*, the better. Therefore the problem becomes QP(3) which is equivalent to the least squares method for a linear regression problem (see Wang, 2005). Firstly we can find the optimal solutions β without the constraints for QP(2), and then we obtain the solution of QP(1).

REFERENCES

Adler, I. and Monteriro, M. D. C. (1989). Interior path following primal-dual algorithms, part II: Convex quadratic programming. *Mathematical Programming*, 44(1), 43-66.

Ben-Daya, M. and AI-Sultan, K. S. (1997). A new penalty function algorithm for convex quadratic programming. *European Journal of Operational Research*, 101(1), 155-163.

Best, M. J. and Ritter, K. (1988). A quadratic programming algorithm. *Zeitschrift fur Operations Research*, 32, 271-297.

Brown, B. M. and Goodall, C. R. (1995). Application of quadratic programming in statistics. In F. P. Kelly (Ed.), *Probability, Statistics and Optimization*, (pp. 341-350). New York: John Wiley Sons Ltd.

Chukova, S., Lai, C. D. and Wang, D. Q. (2005). Reducing quadratic programming problem to regression problem: Stepwise algorithm. *European Journal of Operational Research*, 164(1), 79-88.

Chukova, S., Lai, C. D. and Wang, D. Q. (2004). On the relationship between regression analysis and mathematical programming, *Journal of Applied Mathematics and Decision Sciences*, 8(2), 131-141.

Chung, S. J. and Murty, K. G. (1981). Polynomially bounded algorithms for convex quadratic programming. *Nonlinear Programming*, 4, 439-485.

Fang, K. T., Wang, D. Q. and Wu, G. F. (1982). A class of constrained regression: Programming regression. *Computing Mathematics*, 1(1), 57-69.

Ferrez, J. A., Fukuda, K. and Liebling, T. M. (2005). Solving the fixed rank convex quadratic maximization in binary variables by a parallel zonotope. *European Journal of Operational Research,* 166(1), 35-50.

Mehrotra, S. and Sun, J. (1990). An algorithm for convex quadratic programming that requires O(n{3.5} L) arithmetic operations. *Mathematics of Operations Research,* 15, 342-363.

Wang, D. Q. (2005). Solving quadratic and regression programming. *Proceedings of 40th Annual Conference of ORSNZ,* (pp. 155-162).

Wolfe, P. (1959). The simplex method for quadratic programming. *Econometrica*, 27, 382-398.