Middle-East Journal of Scientific Research 8 (1): 125-128, 2011
ISSN 1990-9233

Logical Comments on Goal Programming Approach Based on Median

Ramin Seyedi

Islamic Azad University, Maragheh Branch, Maragheh, Iran

Abstract: In this paper we see some logical comments for Models in paper "An experimental comparison of the new goal programming and the linear programming approaches in the two-group discriminant problems", also discuss some logical notes about the modelling and other primal subjects.

Key words: Goal programming · Linear programming · Classification · Model selection

INTRODUCTION

A previous by Bal *et al*. [1], consider a new linear programming and two goal programming models for two-group classification problems. When these approaches are applied to the data of real life or ofsimulation, proposed new models perform well both in separating the groups and the group membership predictions of new objects. In discriminant analysis somelinear programming models determine the attribute weights and the cut-off valuein two steps, but suggested models determine simultaneously all of these valuesin one step. Moreover, the results of simulation experiments show that suggested proposed models outperform significantly than existing linear programming and statistical approaches in attaining higher average hit ratios. Deductive logic explicates the notion of a valid argument and develops a formalism how to discern valid inferences that preserve the truth in passing from the premises to the conclusions. Then the premises logically entail the conclusion. Hence, deductive logic studies principles and criteria of truth-preserving inference. It is a formal science in the sense that the meaning of the symbols does not affect soundness or validity of the conclusions. Inductive logic tries to generalize the idea of logical entailment to inferences where the truth of the premises does not guarantee the truth of the conclusions. Still, the truth of the premises might indicate the truth of the conclusion and it is the point of inductive logic to make the vague and informal notion of truth-indication more precise. The central concepts become confirmation and evidential support: it is not asked whether the premises logically entail the conclusion but whether they give good reasons to assert the conclusion and to which degree they support it. In particular, inductive logic is supposed to quantify the effects of observation and measurement on the epistemic status of general hypotheses and theories. Most empirical science sinfer from data to general hypotheses and as deductive relations between the oryand evidence seldom hold, the degree of support is of particular interest. There inductive logic comes into play, figuring out which hypotheses are best confirmed by the data. Usually, inductive reasoning in science proceeds along the lines of the mathematical theory of probability. A probabilistic entailment has the general form.

$$\phi_1^{x_1},\ldots,\phi_n^{x_n} \models \psi^{\mathrm{y}}$$

Where ψ and ϕ_i denote sentences of a given language and y and x_i denote the corresponding probabilities. In particular, y denotes the posterior probability which the premises sentences with a given probability impose on the conclusion. Many frequentist techniques are highly sensitive to underlying assumptions so that human expertise and scientific understanding are required for a sensible implementation. Consequently, I conclude that statistics mainly addresses practical worries about using data in making decisions, predicting events or describing the mechanisms of a system. More precisely, statistics contains a patchwork of different approaches. Choosing one of them is highly sensitive to modeling assumptions, specification of goals, error tolerance etc. and there are no conclusive arguments for a particular method. Hence, comparison of different methods is only possible relative to far-reaching assumptions, blurring the prospects for conceptual unification of statistics.

Corresponding Author: Ramin Seyedi, Islamic Azad University, Maragheh Branch, Maragheh, Iran.

The way how genuinely scientific insights enter the statistical model analysis suggests that statistics resembles an empirical science more than a sophisticated inductive logic. This claim can be substantiated by the numerical turn in statistics: computer-based design of statistical methods and their simulation-based evaluation become more and more important [1-3].

LCM Model and LPMED Models: From a previous papers [4-5], we can see that Linear programming models and many of the others determine the attribute weights and cut-off value. And divide the process of their model into two steps: the first constitutes the determination of attribute weights and the second determines the cut-off value for the classification. In its first step their model makes use of an objective function minimizing the sum of deviations from the group mean classification scores. The LCM model can be formulated as follows:

$$(LCM1)\ Min \sum (d_i^- + d_i^+)$$

s.t.

$$\sum_{j=1}^{k} w_j (x_{i,j} + \mu_{1,j}) + (d_i^- - d_i^+) = 0, i \in G_1$$

$$\sum_{j=1}^{k} w_j (x_{i,j} + \mu_{2,j}) + (d_i^- - d_i^+) = 0, i \in G_2$$

$$\sum_{j=1}^{k} w_j (\mu_{1,j} - \mu_{2,j}) \geq 0$$

Where as $(d_i^-, d_i^+) \geq 0, (i = 1,2,\ldots,n), w_j) j = 1,2,\ldots,k),$, are unrestricted variables.

By this model, $w_j = 1,2,\ldots,k)$, the attribute weights are found and then the object scores are obtained. In this model, it is reached to the weights by making object scores close to their group mean scores. And then the object scores are used in the following model and the classification is made:

$$(LCM2) Min \sum_{i=1}^{n} h_i$$

s.t.

$$S_i + h_i \geq c, i \in G_1$$

$$S_i - h_i \leq c, i \in G_2$$

Whereas $h_i \geq 0$, (I = 1,2,...,n) and c is an unrestricted variable. LCM model, minimizes the sum of individual deviations of the classification scores from their group mean classification scores. Instead of the mean, we can use the median in their model [4, 6-7], because the median is the point that minimizes the total l1-norm distance from all points to it. And this model based on median is called as the LPMED. This LPMED model minimizes the deviations of individual classification scores from their group median classification scores in a two-group classification problem. Similar to the LCM model, the LPMED model formulated as follows:

$$(\text{LPMED1})\ Min \sum (d_i^- + d_i^+)$$

s.t.

$$\sum_{j-1}^{k} w_j (x_{i,j} + med_{1,j}) + (d_i^- - d_i^+) = 0, i \in G_1$$

$$\sum_{j-1}^{k} w_j (x_{i,j} + med_{2,j}) + (d_i^- - d_i^+) = 0, i \in G_2$$

$$\sum_{j=1}^{k} wj (med_{1,j} - med_{2,j}) \geq 0$$

Whereas $(d_i^-, d_i^+) \geq 0, (i = 1,2,\ldots,n), w_j) j = 1,2,\ldots,k),$ are unrestricted variables and $med_{1,j}$ is the median of the jth variable in G1 and med_{2j} is the median of the jth variable in G2. In this model, in the first step the weights w_j are found after the solution to the LPMED1. Here the weights are found by making the object classification scores close to their group median scores. Using these weights the classification scores for each object are evaluated and then the assignment of objects to groups are made by the following the

LPMED2 Model:

$$(\text{LPMED2}) Min \sum_{i=1}^{n} h_i$$

s.t.

$$S_i + h_i \geq c, i \in G_1$$

$$S_i - h_i \leq c, i \in G_2$$

Whereas $h_i \geq 0$, (I = 1,2,...,n) and c is an unrestricted variable. Like in LCM models, The classification is made in two independent steps. (see [1, 6, 7]).

Logical Comments on Model Selection: A lot of the modern debate in statistics and applied sciences focuses on the issue of model selection how to filter a set of candidate models as to obtain a predicatively successful and explanatorily helpful model. To select a model which can be used in further study of the phenomena is such an important decision that we ought to treat it as an integral part of statistical inference. Model selection thus involved the fitting of models to empirical data as well as decisions on the complexity of the model and finding the causally relevant factors. A suitable selection strategy has to evaluate the model selection uncertainty, i.e. to account for the problem that the same data which are used for selecting a model family are also used for fitting the model and estimating the parameters which leads to undue optimism towards the selected model. This problem, sometimes also called selection bias, is a serious problem for statistical inference "Statisticians admit this privately, but they(we) continue to ignore the difficulties because it is not clear what else could or should be done " To base statistical inference on several sensible candidate models is a natural attempt to mitigate the problem. This is the rationale of model averaging: Instead of using a single fitted model as the basis of statistical inference, the inference is based on an average of all candidate models. The subsequent discussion of model selection methods show that satisfactory inference methods are highly sensitive to prior assumptions, goals of inference and substantial scientific insights into the underlying process. Statistical methods are optimal only relative to a variety of external, pragmatic factors: Which types of error do we want to address? What are the practical consequences of a fallacious inference? What is the structure of the random error? Do we have nested or non-nested, linear or non-linear models? And so forth. It turns out to be impossible to make a neat separation between the logical and the decision-theoretic part in statistical inference. Statistics must not be described as a branch of mathematics that miraculously transforms messy data and vague assumptions into a trustworthy posterior distribution. This would neglect the many uncertainties in the process. Instead, statistics seems to be much closer to empirical work and scientific modeling: The most interesting and fruitful questions about models in science deal with the interplay of scientific inquiry and mathematical modeling. Being able to address such questions with the help of statistical tools has yielded an incredible progress, making statistics an indispensable part of empirical science. These questions are beyond the realms of formal theories of inference as inductive logic [4]. The theoretical properties which we can deduce about a model like LCM and LPMED, do not decide alone over its adequacy. There are a lot of different error types and none of the available model selection criteria takes care of all of them. Instead, simulations that resemble typical applications are used in order to study the properties of the proposed criterion. They help us to see whether the criterion is sufficiently robust and applicable in a variety of circumstances. In particular, it is important to check whether the constraints given by the intended application are adequately transformed to the parameters of the simulation (e.g. number of candidate models, linearity, etc). This evaluation of the model selection criteria has a quasi-empirical character and due to the increasing computer power, such approaches become more and more popular. In particular, the results of the simulation analysis can lead to the introduction of adhoc information criteria that are adapted to model selection under specific circumstances.

It should be clear by now that a solution of the model selection problem is more than the solution of an intricate mathematical problem. Human expertise is required to decide which form of modeling is most appropriate. It is clear that these priorities must be set by scientists, not by mathematicians. Only they understand the objects of mathematical modeling sufficiently well to assess the adequacy of a particular discrepancy function or the importance of model parsimony in the relevant context.

The results thus suggest a close collaboration between mathematically minded statisticians and working scientists in order to find the most adequate model selection method in a particular problem. Indeed, this is the route statistics has taken in the last decade, with a lot of statistical literature stemming from researchers that are not located in a mathematics or statistics department. The increased interest in statistical methods among researchers whose primary interests are outside mathematics and statistics shows that a crucial point has been realized: In order to design efficient and helpful statistical methods, scientific understanding and mathematical sophistication have to go hand in hand. It turns out to be impossible to make a neat separation between the logical and the decision-theoretic part in statistical inference. Statistics must not be described as a branch of mathematics that miraculously transforms messy data and vague assumptions into a trustworthy posterior distribution. This would neglect the many uncertainties in the process. Instead, statistics seems to be much closer to empirical work and scientific modeling: The most interesting and fruitful questions about models in science deal with the interplay of scientific inquiry and mathematical modeling. Being able to address such questions with the help of statistical tools has yielded an incredible progress, making statistics an indispensable part of empirical science. These questions are beyond the realms of formal theories of inference as inductive logic [4, 8-10].

REFERENCES

1. Bal, H., H. Orkcu and S. Calabioglu, 2006. An experimental comparison of the new goal programming and the linear programming approaches in the two-group discriminant problems, Computers, Industrial Engineering. Follow this Model in All References.
2. Gilani, A. and M. Padberg, 2002. Alternative methods of linear regression. Mathematical and Computer Modelling.
3. Fitelson, B., 2005. Inductive logic, Philosophy of Science, An Encyclopedia, Routledge, London.
4. Sprenger, J., 2009. Statistics between inductive logic and empirical science, J. Applied Logic, pp: 7.
5. Akaike, H., 1995. Information theory as an extension of the maximum likelihood principle, B.N. Petrov, F. Csaki, (Eds.), Second International Symposium on Information Theory, Akademiai Kiado, Budapest.
6. Bajgier, S.M. and A.V. Hill, 1982. An experimental comparison of statistical and linear programming approaches to the discriminant problem, Decision Sciences.
7. Sueyoshi, T., 1995. Discriminant Analysis view of goal programming. European J. Operational Research.
8. Urnham, P., 1998. Year D.R. Anderson, K Model Selection and Inference: A Practical Information-Theoretic Approach. Springer, New York.
9. Bozdogan, H., 1999. ICOMP. A new model selection criterion, In: H.H. Bock, (Ed.), Classification and Related Methods of Data Analysis, Elsevier Science, Amsterdam.
10. Zucchini, W., 2000. An introduction to model selection, J. Mathematical Psychology.