Problems of Discriminant Functions by Mathematical Programming

- Do I discovery new world of discrimination as same as Christopher Columbus ? -

Shuichi Shinmura
Seikei University
3-3-1 Kichijoji Kitamachi, Musashino, Tokyo 180-8633, JAPAN.

E-mail:shinmura@econ.seikei.ac.jp
http://sun.econ.seikei.ac.jp/~shinmura/
http://www.lindo.jp/
http://www.lindo.com/

1. Introduction

- Discrimination is very important in the industry, and it is very essential in the science.
- It is mainly approached by statistics, mathematical programming (MP) and SVM based on pattern recognition.
  - In 1930's, linear discriminant function (LDF) is introduced by Fisher.
  - In 1950's, pattern recognition starts to recognize the characters.
  - In 1970's (?), regression and discriminant models are defined by MP.
    - Stam (1997) seriously asks us "Why have statisticians rarely used Lp-norm methods?".
  - In 1995, Vapnik gives us new idea named SVM that is influenced by pattern recognition (concept of margin) and statistical approach.
    - evaluation by real data and generalization ability
  - I develop several MP discriminant functions such as IP-OLDF, LP-OLDF, IPLP-OLDF, Revised IP-OLDF, Revised LP-OLDF and successive Revised LP-OLDF.
    - Especially, IP-OLDF based MMN (Minimum Misclassification Number) by integer programming finds new know ledges about discriminant analysis.

2. What are problems?

- After 1970's, there are many researches by MP based discrimnant models.
  - Researchers in this area misunderstand that
    - most of MP models in other area are uniquely approached by MP.
    - But, discriminant analysis are already surveyed by statistical approach.
  - These models don't add new know ledges about discriminant analysis and have no information about 95% significant intervals.
- SVM is welcome by statistician by several reasons. Why?
  - It is evaluated by training and evaluation data as same as statistics.
  - New idea about maximum margin gives us generalization ability.
- But, I doubt following points.
  - There is no doubt about generalization ability for Hard margin SVM.
  - Successive Revised LP-OLDF is superior to Soft margin SVM. The latter only choose better discriminant hyper-plane among Hard margin SV and Revised LP-OLDF.
  - I doubt Kernel SVM is only curve fitting?

3. My ten year's works about discriminant models

- First stage (1998- 2004)
  - IP-OLDF gives us many new knowledge about discriminant analysis such as MMN.

    MIN Σe_i

    $y_i *(\mathbf{x_i}' \mathbf{b}+1)>=-c*e_i$

    $\mathbf{x_i} =(x_{i1}, x_{i2}, \cdots, x_{ip})$:p-dimensional feature vectors

    $y_i = 1$ for $\mathbf{x_i} \in$class1, $y_i =-1$ for $\mathbf{x_i} \in$class2

    $\mathbf{b}$: p-dimensional discriminant coefficient

    e_i: 0/1 decision variable corresponding to each case c: 99999 (BigM constant)
    - Discriminant coefficients become optimal convex polyhedron, the misclassification number of interior points of those equal to MMN.
    - But it sometimes seek wrong MMN if data isn't in general position.
  - LP-OLDF is one of Lp norm models. There are many researches from 1970's.
  - IPLP-OLDF is faster algorithm to seek an estimation of MMN.
- Second stage (2005-2007)
  - Revised IP-OLDF always seek right MMN.

    MIN Σe_i

    $y_i *(\mathbf{x_i}' \mathbf{b}+b_0)>= 1-c*e_i$
  - Every MP based discriminant models are derived by Successive Revised LP-OLDF. This model is faster than Revised IP-OLDF and Soft Margin SVM.
- Third stage(2008-)
  - We can construct 95% confidence interval of every statistics such as discriminant coefficients and MMN , etc. by bootstrap method and LINGO developed by LINDO Systems Inc.
    - Generate re-sampling sample by Speakeasy.
    - Apply Revised IP-OLDF for these bootstrap samples and 95% confidence interval.

4. Conclusion

- Revised IP-OLDF is defined by MMN criterion.
  - MMN decreases monotonously on forward stepwise models.
    - So, it can finds the minimum dimension of linear separable spaces if data is linear separable.
  - Stepwise methods, AIC and Cp statistics select wrong models if data is linear separable such as Swiss Bank Note (Flury & Rieduyl (1988)).
  - We can understand discrimination by MMN or Optimal Convex Polyhedron.
- Restricted Revised LP-OLDF is superior to S-SVM and sometimes finds MMN.
  - This model is faster than Revised IP-OLDF by IP and S-SVM by quadratic programming.
- Bootstrap statistics and LINGO (the second solver) supported by LINDO Systems Inc. can estimate 95% significant interval by re-sampling data.
  - Ordinary statistics offers 95% significant interval only for sample mean, standard deviation, skewness, kurtosis, ratio, chi-square test, F-test of regression and t-test of regression coefficient.
    - Bootstrap offers 95% significant interval for every sample statistics.

References

Flury,B. & Rieduyl,H. (1988). Multivariate Statistics: A Practical Approach. Cambridge University Press.

Stam, A. (1997). Nontraditional approaches to statistical classification: Some perspectives on Lp-norm methods. Annals of Operations Research, 74, 1-36.

Schrage, L. (2006). Optimization Modeling with LINGO (6 editions.). LINDO Systems Inc.

Shinmura S. (1998). Optimal Linear Discriminant Functions using Mathematical Programming, Bulletin of The Computational Statistics of Japan 11(2), 93-105.

Shinmura,S.(2000). A new algorithm of the linear discriminant functions using integer programming. New Trends in Probability and Statistics 5, 133-142.

Shinmura S.(2002). Lecture Seriece; "The fusion of OR and Statistics", Journal of the Operations Research Society of Japan 47(1), 38-45. 47(2), 109-113. 47(3), 172-185. 47(4), 244-250. 47(5), 315-321. (Japanese).

Shinmura S.(2004). New Algorithm of Discriminant Analysis using Integer Programming, IPSI 2004 Pescara VIP Forum, 1-18.

Shinmura S.(2007). Overviews of Discriminant Function by mathematical Programming, Bulletin of The Computational Statistics of Japan 20(1-2), 59-94.

Vapnik,V.(1995). The Nature of Statistical Learning Theory. Springer-Verlag,1995.