A general index for admission decisions

The University of Tokyo   Sei, Tomonari

Problem setting

Consider a multivariate data, where every variable has the meaning that a larger value indicates higher score. Our aim is to separate the observations into two groups: one is at overall higher score and the other is not. The problem is unsupervised.

We suppose that a separation procedure is based on a weighted sum of the variables. A naive weighting method is to take the sum of standardized variables. However, this method has the following paradoxical phenomena: the average of the higher-score group for a particular variable may be less than that of the lower-score group. We provide an algorithm without such shortcomings, as long as random decisions at the borderline are allowed.

Method

Let $\boldsymbol{x}_{(1)}, \ldots, \boldsymbol{x}_{(n)} \in \mathbb{R}^d$ be scores of n examinees on d subjects. Suppose that the ratio α of admission is given. Let $(\boldsymbol{w}, c)$ be a solution to the following convex optimization problem:

$$\begin{aligned} &\text{Minimize} \quad \sum_{i=1}^{d}(-\log w_i) + \frac{1}{n}\sum_{t=1}^{n}\ell_\alpha(\boldsymbol{w}^\top \boldsymbol{x}_{(t)} - c) \\ &\text{subject to} \quad (\boldsymbol{w}, c) \in \mathbb{R}_+^d \times \mathbb{R}, \end{aligned}$$

where ℓ_α is the check-loss function [1]: $\ell_\alpha(u) = \max(u/\alpha, -u/(1-\alpha))$. Then a general index is defined by $g_t = \boldsymbol{w}^\top \boldsymbol{x}_{(t)} - c$. The t-th student is admitted if $g_t > 0$. We have the following properties as desired:

$$\frac{\sum_t \mathbb{I}_{\{g_t>0\}}}{n} \approx \alpha, \quad \frac{\sum_t x_{ti}\mathbb{I}_{\{g_t>0\}}}{n\alpha} - \frac{\sum_t x_{ti}\mathbb{I}_{\{g_t<0\}}}{n(1-\alpha)} \approx \frac{1}{w_i} > 0.$$

References

[1] Koenker, R. (2005). *Quantile Regression*, Cambridge University Press.