Available online at www.sciencedirect.com
ScienceDirect
Journal of Approximation Theory 188 (2014) 57–68
Journal of Approximation Theory
www.elsevier.com/locate/jat

Full length article

Learning rate of support vector machine for ranking[☆]

Heng Chen[a,*], Di-Rong Chen[b]

[a] School of Statistics, Capital University of Economics and Business, Beijing 100070, PR China
[b] Department of Mathematics, Wuhan Textile University, Wuhan 430200, PR China

Received 5 January 2013; received in revised form 3 July 2014; accepted 26 August 2014
Available online 21 September 2014

Communicated by Ding-Xuan Zhou

Abstract

The ranking problem has received increasing attention in both the statistical and machine learning literature. This paper considers support vector machines for ranking. Under some mild conditions, a learning rate is established.

Keywords: Ranking rule; Support vector machine; Excess risk; U-statistic

1. Introduction and main result

In the ranking problem, one has to compare two different observations and decide the ordering between them. It has received increasing attention in both the statistical and machine learning literature.

The problem may be modeled in the framework of statistical learning (see [7,10]). Let (X, Y) be a pair of random variables taking values in $\mathcal{X} \times \mathcal{Y}$ with $\mathcal{X} \subset \mathbb{R}^d$, $\mathcal{Y} \subset \mathbb{R}$ being compact sets. The random observation X models some object and Y its real-valued label. Let (X', Y') denote a pair of random variables identically distributed with (X, Y) (with respect to the probability

☆ Research supported in part by NSF of China under grants 11171014 and 91130009, National Basic Research Program of China under grant 973-2010CB731900, and National Technology Support Program under grant 2012BAH05B01.
* Corresponding author.
E-mail address: chenheng_0913@163.com (H. Chen).

http://dx.doi.org/10.1016/j.jat.2014.08.004
0021-9045/

$\mathbb{P}$), and independent of it. In the ranking problem one observes X and X' but not their labels Y and Y'. We think about X being "better" than X' if $Y > Y'$. We are to construct a function $f : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, called a ranking rule, which predicts the ordering between objects in the following way: if $f(X, X') \geq 0$, we predict that X is better than X'. A ranking rule f has the property $f(x, x')f(x', x) \leq 0$. The performance of a ranking rule f is measured by the ranking risk

$$L(f) = \mathbb{P}(\text{sign}(Y - Y')f(X, X') < 0), \tag{1}$$

i.e., the probability that f ranks two randomly drawn instances incorrectly.

Let $\eta(X, X') = \mathbb{P}(Y > Y' \mid X, X')$. It is easily seen that the ranking rule

$$f^* := \text{sgn}(2\eta - 1)$$

minimizes the risk $L(f)$ in (1) over the class of all measurable functions. Obviously, $f^*(x_1, x_2) = -f^*(x_2, x_1)$.

In practice, the best rule f^* is unknown since the probability $\mathbb{P}$ is unknown. Instead, we are given n independent, identically distributed copies, $Z_1 = (X_1, Y_1), \ldots, Z_n = (X_n, Y_n)$. The empirical error of f

$$L_n(f) = \frac{1}{n(n-1)} \sum_{i \neq j} \mathbb{I}[\text{sign}(Y_i - Y_j)f(X_i, X_j) < 0] \tag{2}$$

is an estimator of $L(f)$, where $\mathbb{I}(\cdot)$ is the indicator function. Based on the empirical estimate, one may define the empirical risk minimizer

$$r_n = \arg\min_{f \in \mathcal{F}} L_n(f). \tag{3}$$

In [7, Corollary 6], under a noise assumption as (16) below, the authors established a fast convergence for VC classes $\mathcal{F}$:

$$\mathbb{P}\Big(L(r_n) - L^* \leq 2\Big(\inf_{f \in \mathcal{F}} L(f) - L^*\Big) + C\Big(\frac{\ln n + t}{n}\Big)^{\frac{1}{2-\alpha}}\Big) \geq 1 - e^{-t}, \quad t > 0,$$

where $L^* = L(f^*)$, and C and $\alpha \in (0, 1]$ are constants.

It is known that finding a minimizer of $L_n(f)$ over $\mathcal{F}$ is computationally difficult and not efficient. A commonly used approach is to replace the indicator function $\mathbb{I}$ by a convex and non-negative function φ defined on $\mathbb{R}$ and $L_n(f)$ by the empirical convex risk [7,10]

$$Q_n(f) = \frac{1}{n(n-1)} \sum_{i \neq j} \varphi_f(Z_i, Z_j), \tag{4}$$

where $\varphi_f(z, z') = \varphi(\text{sign}(y - y')f(x, x'))$ with $z = (x, y)$, $z' = (x', y')$.

Denote f_n the minimizer of empirical convex risk over a class $\mathcal{F}$ of rules:

$$f_n = \arg\min_{f \in \mathcal{F}} Q_n(f).$$

In [10], a fast convergence rate

$$\mathbb{P}\Big(Q(f_n) - \inf_{f \in \mathcal{F}} Q(f) \leq C_1 \max\Big(\frac{\ln n}{n}, \frac{1}{n^\gamma}\Big) + C_2 \frac{C_3 + t}{n}\Big) \geq 1 - e^{-t}, \quad t > 0, \tag{5}$$