Journal of Machine Learning Research 14 (2013) 3493-3494 Submitted 12/12; Revised 8/13; Published 11/13

Comment on "Robustness and Regularization of Support Vector Machines" by H. Xu et al. (Journal of Machine Learning Research, vol. 10, pp. 1485-1510, 2009)

Yahya Forghani[*] YAHYAFOR2000@YAHOO.COM
Hadi Sadoghi H-SADOGHI@UM.AC.IR
Department of Computer
Ferdowsi University of Mashhad
Azadi Sq., Mashhad, IRAN

Editor: Ingo Steinwart

Abstract

This paper comments on the published work dealing with robustness and regularization of support vector machines (Journal of Machine Learning Research, Vol. 10, pp. 1485-1510, 2009) by H. Xu et al. They proposed a theorem to show that it is possible to relate robustness in the feature space and robustness in the sample space directly. In this paper, we propose a counter example that rejects their theorem.

Keywords: kernel, robustness, support vector machine

1. Comment

Firstly, it must be stated that Xu et al. (2009) made a good study of robustness and regularization of support vector machines. They proposed the following theorem to show that it is possible to relate robustness in the feature space and robustness in the sample space directly:

Theorem (Xu et al., 2009) *Suppose that the kernel function has the form* $k(x,x') = f(\|x-x'\|)$*, with* $f: R^+ \to R$ *a decreasing function. Denote by* H *the RKHS space of* $k(.,.)$ *and* $\phi(.)$ *the corresponding feature mapping. Then we have any* $x \in R^n, w \in H$ *and* $c > 0$*,*

$$\sup_{\|\delta\|\le c} \langle w, \phi(x-\delta)\rangle = \sup_{\|\delta_\phi\|_H \le \sqrt{2f(0)-2f(c)}} \left\langle w, \phi(x) - \delta_\phi \right\rangle.$$

The following counter example rejects the mentioned theorem. However, this theorem is a standalone result in the appendix of the paper of Xu et al. (2009), which is not used anywhere else in the paper of Xu et al. (2009). Thus, the main result and all other results of Xu et al. (2009) are not affected in any way.

Counter example. Let $\phi(.)$ be the feature mapping of Gaussian kernel function. We have $\|\phi(x)\|_H = 1$. Let $w = \phi(x)$. Therefore, $\langle w, \phi(x)\rangle = \|w\|_H$, and

$$\sup_{\|\delta\|\le c} \langle w, \phi(x-\delta)\rangle = \|w\|_H. \tag{1}$$

[*]. Also at Islamic Azad University, Mashhad Branch, Mashhad, IRAN.

Moreover,

$$\sup_{\|\delta_\phi\|_H \le \sqrt{2f(0)-2f(c)}} \left\langle w, \phi(x) - \delta_\phi \right\rangle =$$

$$\sup_{\|\delta_\phi\|_H \le \sqrt{2f(0)-2f(c)}} \langle w, \phi(x) \rangle + \sup_{\|\delta_\phi\|_H \le \sqrt{2f(0)-2f(c)}} \left\langle w, \delta_\phi \right\rangle =$$

$$\|w\|_H + \sup_{\|\delta_\phi\|_H \le \sqrt{2f(0)-2f(c)}} \left\langle w, \delta_\phi \right\rangle =$$

$$\|w\|_H + \|w\|_H \sqrt{2f(0) - 2f(c)}. \tag{2}$$

According to Equation (1) and (2), and since f is a decreasing function, for any $c > 0$, we have

$$\sup_{\|\delta\| \le c} \langle w, \phi(x - \delta) \rangle \le \sup_{\|\delta_\phi\|_H \le \sqrt{2f(0)-2f(c)}} \left\langle w, \phi(x) - \delta_\phi \right\rangle .$$

End of counter example.

The exact spot that the error has been occurred in the mentioned theorem is Equation (19) of the paper of Xu et al. (2009). There it has been claimed that the image of the RKHS feature mapping is dense, which unfortunately is not true. Indeed, because $\langle \phi(x), \phi(x) \rangle = K(0)$ where $K(.)$ is the kernel function, the image of the feature mapping is in a ball of radius $\sqrt{K(0)}$.

References

Huan Xu, Shie Mannor, and Constantine Caramanis. Robustness and regularization of support vector machines. *Journal of Machine Learning Research*, 10:1485–1510, 2009.