

คุณสมบัติการลู่เข้าและความทนทานต่อข้อมูลรบกวนของซัพพอร์ตเวกเตอร์แมชชีน
CONVERGENCE AND NOISE-TOLERANCE PROPERTIES OF SUPPORT VECTOR
MACHINES

นิตยา เกิดประสพ, กิตติศักดิ์ เกิดประสพ, ปาริฉัตร รัตนศฤงค์, ลักษมี โขมโนทัย และ
ธรรมศักดิ์ เขียวนิเวศน์

Nittaya Kerdprasop, Kittisak Kerdprasop, Parichat Rattanasaring, Laksamee Khomnotai and
Thammasak Thianniwet

Data Engineering and Knowledge Discovery (DEKD) Research Unit, School of Computer
Engineering, Suranaree University of Technology, Nakhon Ratchasima, Thailand,

E-mail address: nittaya@ccs.sut.ac.th, kerdpras@ccs.sut.ac.th

บทคัดย่อ

ข้อมูลในความเป็นจริงมักจะมีปริมาณมากและมีความผิดพลาดแฝงอยู่ ข้อมูลที่ผิดพลาดจะ
รบกวนความเที่ยงตรงของอัลกอริทึมเรียนรู้ นอกจากนี้ปริมาณที่มากเกินไปของข้อมูลยังส่งผลให้
อัลกอริทึมเรียนรู้ทำงานได้ช้าลง ซึ่งจากผลการศึกษาในอดีตพบว่าการเรียนรู้จากข้อมูลสุ่มให้
ผลลัพธ์เป็นโมเดลที่เที่ยงตรงได้เช่นเดียวกับการเรียนรู้จากข้อมูลทั้งหมด งานวิจัยนี้มีจุดมุ่งหมายที่
จะศึกษาคุณสมบัติการลู่เข้าและความทนทานต่อข้อมูลรบกวนของอัลกอริทึมซัพพอร์ตเวกเตอร์
แมชชีน ผลการทดสอบพบว่าซัพพอร์ตเวกเตอร์แมชชีน มีการลู่เข้าสู่ผลการสังเคราะห์โมเดลที่ดี
ที่สุดในเวลารวดเร็ว ด้วยปริมาณข้อมูลสุ่มประมาณ 300-500 รายการข้อมูล และมีความทนทานต่อ
ข้อมูลรบกวนที่ปรากฏในหลากหลายตำแหน่งได้ในระดับที่ค่อนข้างสูง

Abstract

Real-world data tend to be huge and noisy. Noise, or random error, in data can reduce
learning performance in terms of learning accuracy. Learning concept model from a very large
data also causes a performance problem even with the most efficient algorithm. It has been
shown in the previous studies that many learning algorithms can learn from a sample of data and
produce as accurate model as does learning from the whole population. We are interested in
studying the convergence and noise-tolerance properties of the support vector machines, a
statistical learning algorithm introduced recently and gained popularity rapidly. The experimental
results reveal that support vector machines converge very fast with a sufficient sample of 300-500
instances. The noise-tolerant level is quite high when noise occurs at various positions.

Published in: Proceedings of 30th Congress on Science and Technology of Thailand, Impact
Exhibition and Convention Center, Bangkok, Thailand, October 19-21, 2004.