Mach Learn (2013) 90:127–160
DOI 10.1007/s10994-012-5303-x

Mass estimation

Kai Ming Ting · Guang-Tong Zhou · Fei Tony Liu · Swee Chuan Tan

Received: 17 June 2011 / Revised: 20 May 2012 / Accepted: 1 June 2012 / Published online: 26 June 2012

Abstract This paper introduces mass estimation—a base modelling mechanism that can be employed to solve various tasks in machine learning. We present the theoretical basis of mass and efficient methods to estimate mass. We show that mass estimation solves problems effectively in tasks such as information retrieval, regression and anomaly detection. The models, which use mass in these three tasks, perform at least as well as and often better than eight state-of-the-art methods in terms of task-specific performance measures. In addition, mass estimation has constant time and space complexities.

Keywords Mass estimation · Density estimation · Information retrieval · Regression · Anomaly detection

1 Introduction

> 'Estimation of densities is a universal problem of statistics (knowing the densities one can solve various problems).' —Vapnik (2000).

Density estimation has been the base modelling mechanism used in many techniques designed for tasks such as classification, clustering, anomaly detection and information

Editor: Tong Zhang.

K.M. Ting (✉) · F.T. Liu · S.C. Tan
Gippsland School of Information Technology, Monash University, Melbourne, Vic 3842, Australia
e-mail: kaiming.ting@monash.edu

F.T. Liu
e-mail: tony.liu@monash.edu

S.C. Tan
e-mail: james.tan@monash.edu

G.-T. Zhou
School of Computing Science, Simon Fraser University, Burnaby, BC, V5A 1S6, Canada
e-mail: zhouguangtong@gmail.com