Ann Inst Stat Math (2008) 60:699–746
DOI 10.1007/s10463-008-0197-x

# Direct importance estimation for covariate shift adaptation

**Masashi Sugiyama · Taiji Suzuki · Shinichi Nakajima · Hisashi Kashima · Paul von Bünau · Motoaki Kawanabe**

Received: 4 January 2008 / Revised: 19 April 2008 / Published online: 30 August 2008


**Abstract** A situation where training and test samples follow different input distributions is called *covariate shift*. Under covariate shift, standard learning methods such as maximum likelihood estimation are no longer consistent—weighted variants according to the ratio of test and training input densities are consistent. Therefore, accurately estimating the density ratio, called the *importance*, is one of the key issues

in covariate shift adaptation. A naive approach to this task is to first estimate training and test input densities separately and then estimate the importance by taking the ratio of the estimated densities. However, this naive approach tends to perform poorly since density estimation is a hard task particularly in high dimensional cases. In this paper, we propose a direct importance estimation method that does not involve density estimation. Our method is equipped with a natural cross validation procedure and hence tuning parameters such as the kernel width can be objectively optimized. Furthermore, we give rigorous mathematical proofs for the convergence of the proposed algorithm. Simulations illustrate the usefulness of our approach.

**Keywords** Covariate shift · Importance sampling · Model misspecification · Kullback–Leibler divergence · Likelihood cross validation