Mach Learn (2012) 86:233–272
DOI 10.1007/s10994-011-5263-6

Learning by extrapolation from marginal to full-multivariate probability distributions: decreasingly naive Bayesian classification

Geoffrey I. Webb · Janice R. Boughton · Fei Zheng · Kai Ming Ting · Houssam Salem

Received: 8 December 2009 / Accepted: 15 September 2011 / Published online: 13 October 2011

Abstract Averaged n-Dependence Estimators (AnDE) is an approach to probabilistic classification learning that learns by extrapolation from marginal to full-multivariate probability distributions. It utilizes a single parameter that transforms the approach between a low-variance high-bias learner (Naive Bayes) and a high-variance low-bias learner with Bayes optimal asymptotic error. It extends the underlying strategy of Averaged One-Dependence Estimators (AODE), which relaxes the Naive Bayes independence assumption while retaining many of Naive Bayes' desirable computational and theoretical properties. AnDE further relaxes the independence assumption by generalizing AODE to higher-levels of dependence. Extensive experimental evaluation shows that the bias-variance trade-off for Averaged 2-Dependence Estimators results in strong predictive accuracy over a wide range of data sets. It has training time linear with respect to the number of examples, learns in a single pass through the training data, supports incremental learning, handles directly missing values, and is robust in the face of noise. Beyond the practical utility of its lower-dimensional variants, AnDE is of interest in that it demonstrates that it is possible to create low-bias high-variance generative learners and suggests strategies for developing even more powerful classifiers.

Keywords Bayesian learning · Classification learning · Probabilistic learning · Averaged one-dependence estimators · Naive Bayes · Semi-naive Bayesian learning · Learning without model selection · Ensemble learning · Feating

Editor: Peter Flach.

G.I. Webb (✉) · J.R. Boughton · F. Zheng · K.M. Ting · H. Salem
Faculty of Information Technology, Monash University, Clayton, VIC, 3800, Australia
e-mail: Geoff.Webb@monash.edu

J.R. Boughton
e-mail: Janice.Boughton@monash.edu

K.M. Ting
e-mail: Kaiming.Ting@monash.edu