Weighted orthogonal greedy algorithm for prediction under covariate shift

Shinpei Imori[1]

[1]*Hiroshima University, Japan*

Abstract

The orthogonal greedy algorithm (OGA) is known as one of the good methods for prediction in high-dimensional regression models. There have been theoretical studies of OGA in previous studies (e.g., Ing and Lai, 2011, Statistica Sinica). When aiming at the prediction, we need to consider the distribution of the test data, which is usually (or implicitly) assumed to follow the same distribution as the training one.

In this talk, we attempt to mitigate this restriction by allowing the test and training distributions of the explanatory variables not to be the same, although the regression models (i.e., the conditional distribution of the response variable given the explanatory variables) in the training and test data are assumed to be the same. This situation is called the covariate shift introduced in Shimodaira (2000, Journal of Statistical Planning and Inference). We consider the high-dimensional linear regression models, where the number of explanatory variables can be larger than the sample size. Note that we do not assume that the true regression model is correctly specified.

The original OGA may not work well under the covariate shift because the original OGA is not defined by taking into account of the difference between the training and test distributions. So, we propose to use a weighted version of OGA. The weight function is based on the density ratio between the training and test explanatory variables, and the weight function is estimated from the observed data. Then, under some mild conditions, we can derive a convergence rate of the weighted OGA with respect to a conditional prediction error for incorrectly specified models. This result indicates that our proposed method has a promising prediction accuracy to predict the test response variable under the covariate shift.

This is a joint work with Professor Ching-Kang Ing at National Tsing Hua University, Taiwan.

Keywords: Orthogonal greedy algorithm; Covariate shift; High-dimensional analysis.