Forecasting Discrete Outcomes Using Many Highly Correlated Predictors

Anh Tran*

October, 2022

Abstract

In this paper, we develop boosting algorithms Markov Chain Factor Adjusted Naive Bayes (MCFANB) AdaBoost and MCFANB ε-Boost for forecasting binary outcomes when time series and cross-sectional dependence both exist in high-dimensional covariates as in economic and financial data. In particular, the base learner of MCFANB AdaBoost algorithm is the modified Naïve Bayes classifier in which Markov Chain properties of the outcome variable are assumed to capture dependence among time series binary classes, and high dimensional covariates are transformed from highly correlated covariates to weak correlated ones by Class-dependent Factor Adjusted procedure. We also propose the regularized version MCFANB ε-Boost in which the 1-norm of the weights of the base classifiers is restricted to a fixed value. The algorithms' consistency properties are examined in the case that the samples are independent and identically distributed (i.i.d) and in the case that samples are non-i.i.d but from stationary weakly dependent sequences (β-mixing). In the simulation study, the proposed methods are shown to outperform other competing methods in achieving low test errors. Finally, in the empirical study, we apply our proposed methods to forecast economic recessions with economic and financial data.

Keywords: Adaboost, classification, β-mixing, economic recession forecasting

*University of California, Riverside, Department of Economics, E-mail: atran166@ucr.edu