

Consistent Cross-Validation for Tuning Parameter Selection in High-Dimensional Variable Selection

Yang Feng*

Columbia University, New York, NY, USA yangfeng@stat.columbia.edu

Yi Yu

Fudan University, Shanghai, China yu.yi@gmx.com

For variable selection in high-dimensional settings, we systematically investigate the asymptotic behavior of several cross-validation methods (e.g., K-fold cross-validation) for selecting the penalty parameter in the popular penalized maximum likelihood method. Especially, we unveil a counter-intuitive theoretical result that the shrinkage problem with LASSO penalty is not the true reason for the over-selection phenomenon in K-fold cross-validation under certain technical conditions. After delineating the properties of the traditional cross-validation method, we propose a new procedure for selecting tuning parameters named Consistent Cross-Validation (CCV), which is shown to enjoy the model selection consistency property. Extensive simulations and real data analysis are conducted, supporting the theoretical results.

Key Words: generalized linear models; high-dimensional variable selection; penalty; restricted MLE; consistent cross-validation