

Principal components-driven lasso

Grzegorz Mika

Abstract

In this paper the properties of the principal components least absolute shrinkage and selection operator (pcLASSO) are studied. The pcLASSO was introduced by J.K Tay, J.Friedmann and R. Tibshirani in 2018 in paper 'Principal component-guided sparse regression' as a new method for supervised learning, especially suited to wide data where the number of features is much greater than the number of observations. The method combines the lasso sparsity penalty with the quadratic penalty that shrinks the coefficients toward the leading principal components of the featured matrix. This method is especially powerful when the features are grouped. In this case the pcLASSO shrinks each group-wise coefficient of the solution toward the leading principal component of the solution exploiting the assignment of the features to the groups much better in comparison to classic lasso method. The theoretical properties of pcLASSO are investigated for finite samples where oracle inequalities are proved. Thereafter the asymptotic behavior of the selected estimator is investigated using the L_2 - norm and the Kullback-Leibler divergence. Showed oracle inequalities allow to select the optimal hyperparameter in a much more computationally efficient way in comparison to traditionally used methods like cross-validation. The asymptotic properties proved in this paper show that under some mild conditions the pcLASSO estimator attains the optimal rate of convergence over the wide class of problems. In the second part, the results of wide simulation studies are presented. The properties of the pcLASSO estimator are compared with those of the traditional LASSO, PCA method, ridge regression and elastic net. This comparison points out to classes of problems for which the pcLASSO estimator exhibits significantly better performance in comparison to other methods. The theory is illustrated on both simulated and real data for small and big samples.