The Regularization Path for Nonlinear LASSO Regression

Gang Wang, Dit-Yan Yeung, Frederick H. Lochovsky
Department of Computer Science and Engineering
Hong Kong University of Science and Technology
Hong Kong, China
{wanggang, dyyeung, fred}@cse.ust.hk

Topic: learning algorithms.
Presentation: poster.

1 Abstract

LASSO, first proposed by Tibshirani (1996), regularizes ordinary least square regression with an L_1 regularizer. It can lead to sparse solutions by shrinking the coefficients of the irrelevant or redundant features to zero. Theoretical analysis (Ng, 2004) shows that LASSO is particularly effective when there exist many irrelevant features but only very few training examples. Efron et al. (2004) developed the Least Angle Regression (LARS) algorithm, which can fit the coefficient path for the linear least square regression problem regularized with the L_1 norm. An important finding is that the coefficient path is piecewise linear and hence it is efficient to explore the entire solution path by monitoring the breakpoints only.

Since we are more interested in the nonlinear regression function, we use $\mathcal{D} = \{K(\mathbf{x}, \mathbf{x}_1), ..., K(\mathbf{x}, \mathbf{x}_n)\}$ as a dictionary of basis functions and trace the whole regularization path of the nonlinear LASSO regression. Compared with the support vector regression (SVR) where there are two hyperparameters ϵ and λ, the user can only consider one regularization hyperprameter in this model. The regularization path is also piecewise linear hence we can efficiently explores every solution with all regularization hyperprameter values. The experimental results show that the solution of nonlinear LASSO often has more sparse representation than the SVR solution when they have similar generalization errors. Thus it benefits more representational and computational advantages. We have prepared some videos[1] to illustrate the regularization path for the nonlinear regression LASSO.

[1] www.cse.ust.hk/~wanggang/sol_path/nlLASSO.htm

References

Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *The Annals of Statistics*, *32*, 407499.

Ng, A. Y. (2004). Feature selection, l1 vs. l2 regularization, and rotational invariance. *Proceedings of the 21th International Conference on Machine Learning (ICML-04)*..

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B*, *58*, 267–288.