Lecture 6: Machine Learning

Zhentao Shi

March 20, 2017

Machine Learning

From the view of an econometrician, machine learning is a set of data fitting procedures that focus on out-of-sample prediction. The simplest illustration is in the regression context. We repeat a scientific experiment for n times, which generates a dataset $(y_i, x_i)_{i=1}^n$. What would be the best way to predict y_{n+1} from the same experiment if we observe x_{n+1}? Mean squared errors (MSE) is a popular criterion for comparison when the response variable is continuous.

In modern scientific analysis, the number of covariates x_i can be enormous. In DNA microarray analysis, we are looking for the association of a sympton to genes. Variable selection is crucial here because we have scientific evidence that only a small handful of genes are involved.

Variable Selection

Back to economic analysis, the problem of variable selection is not well appreciated, though applied economists are sellecting varibles implicitly. They rely on their prior knowledge to choose some relevant variables from a large number of potential candidates. For example, the UK Living Costs and Food Survey has been widely used for analysis of demand theory and family consumption. A questionaire in this survey consists of thousand of questions. It is not clear what are the most relevant variables in general.

The most well-known variable selection method in a regression is the least-absolute-shrinkage-and-selection-operator (Lasso) (Tibshirani 1996). Upon on the usual OLS criterion function, Lasso penalizes the L_1 norm of the coefficients.

$$(2n)^{-1}(Y - X\beta)'(Y - X\beta) + \lambda\|\beta\|_1$$

where $\lambda \geq 0$ is a tuning parameter. In a wide range of values of λ, Lasso can shrink some coefficients exact to 0, which suggests that these variables are likely to be irrelevant in the regression. However, later research finds that Lasso cannot consistently select the relevant variables from the irrelevant ones.

Anothe successful variable selection is smoothly-clipped-absolute-deviation (SCAD) (Fan and Li 2001). Their crition function is

$$(2n)^{-1}(Y - X\beta)'(Y - X\beta) + \sum_{j=1}^{d} \rho_\lambda(|\beta_j|)$$

where

$$\rho_\lambda(\theta) = \lambda\left\{1\{\theta \leq \lambda\} + \frac{(a\lambda - \theta)_+}{(a-1)\lambda} \cdot 1\{\theta > \lambda\}\right\}$$

for some $a > 2$ and $\theta > 0$. This is a non-convex function, and Fan and Li (2001) establish the so-called *oracle property*, which means that an estimator can achieve variable selection consistency and (pointwise) asymptotic normality simultaneously.

The follow-up *adaptive Lasso* (Zou 2006) enjoys selection consistency. Adaptive Lasso is a two step scheme: 1. First run a Lasso and save the estimator $\hat{\beta}^{(1)}$. 2. Solve

$$(2n)^{-1}(Y - X\beta)'(Y - X\beta) + \lambda \sum_{j=1}^{d} w_j |\beta_j|$$

where $w_j = \frac{1}{\left|\hat{\beta}_j^{(1)}\right|^a}$ and $a \geq 1$ is a constant. (Common choice is $a = 2$).

In R, `glmnet` or `LARS` implements Lasso, and `ncvreg` does SCAD. We can set the adaptive Lasso weight via the argument `penalty.factor` in `glmnet`.

Prediction

If the purpose of the regression is not variable selection, but for accurate prediciton of the response variables, then there are other methods available.

boosting, forward selection, and reweighting (Bai and Ng)

regression tree, bagging (Killian and Inoue), average of subsampling

three steps in econometrics: consistency -> asymptotic normality -> efficiency

rebuke from ML data is big, don't worry accuracy inference is not interested, prediction matters no DGP is considered, nothing to converge to

In regression context, explore all sorts of nonlinear relationship.

References

Fan, Jianqing, and Runze Li. 2001. "Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties." *Journal of the American Statistical Association* 96 (456). Taylor & Francis: 1348–60.

Tibshirani, Robert. 1996. "Regression Shrinkage and Selection via the Lasso." *Journal of the Royal Statistical Society. Series B (Methodological)*. JSTOR, 267–88.

Zou, Hui. 2006. "The Adaptive Lasso and Its Oracle Properties." *Journal of the American Statistical Association* 101 (476). Taylor & Francis: 1418–29.