Applications of boosting to economics

Rodney Beard

Outline

What is boosting?

Machine learning vs Econometrics

Two example papers

Software

Discussion

What is boosting?

- Ensemble approach
- Weak learners
- Boosting is a recursive updating method (Family of Algorithms - many boosting methods)
- Many different algorithms in the boosting family.
- Method of steepest descent (Deep mathematical foundations).

What is econometrics?

Economics theory in its relation to statistics and mathematics - Ragnar Frisch , 1930.

Another definition

'Econometrics' as a special type of economic analysis in which the general theoretical approach-often formulated in explicitly mathematical terms-is combined frequently through the medium of intricate statistical procedures-with empirical measurement of economic phenomena - Gerhard Tintner 1953.

Steps towards artificial intelligence

The problems of heuristic programming-of making computers solve really difficult problems-are divided into five main areas: Search, Pattern-Recognition, Learning, Planning, and Induction. - Marvin Minsky (1961, Steps towards artificial intelligence.

Machine learning, 1959

The programming of a digital computer to behave in a way which. if done by human beings or animals, would be described as involving the process of learning. - Arthur Samuel 1959.

- Luo, Ye, and Martin Spindler. 2017. "L2-Boosting for Economic Applications." American Economic Review, 107 (5): 270–73. DOI: 10.1257/aer.p20171040
- Salvadé, N., Hillel, T. (2024). RUMBoost: Gradient Boosted Random Utility Models. arXiv preprint arXiv:2401.11954.

L_2 boosting for economic applications

- What is L_2 boosting?
- Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, 29(5), 1189–1232. `http://www.jstor.org/stable/2699986`
- Bühlmann, P., Yu, B. (2003). Boosting with the L_2 loss: regression and classification. *Journal of the American Statistical Association*, 98(462), 324-339.
- L_2 just means quadratic loss (L_2 - norm minimization).

Problem

- High-dimensional problems (more features than data)
- alternative to LASSO (regularization)
- Designed for Linear models

Experimental school - Economics

- Causal ionference
- Gold standard: Controlled experiment (Randomized control trial)
- Not always possible to do.

The reflection problem

Here is a problem from everyday life. Suppose you observe the almost simultaneous movements of a person and his reflection in a mirror. does the mirror image induce the person's movements, does the image reflect the person's movements, or do the person and image move together in response to a common external stimulus? data alone cannot answer this question. - Mansky *Identification problems in the social sciences* 1995.

Similar problem arises with separating individual and group behaviour - implications for public policy.

Identification problem

- Instrumental variables (Identification problem - system identification)
- This is related to inverse modelling originally.
- Example: supply and demand form a system of equations.
- Data generated by the system. So we aren't doing function approximation here but system identification (major difference between machine learning and econometrics)

Model framework

We are in the world of linear models, so linear regression. Linear identification is well understood, non-linear identification less so, and we care about identification because of the reflection problem.

$$y_i = x_i'\beta + \epsilon_i, i = 1, \ldots, n$$

where x_i and β are both column vectors of length p_n and ϵ_i is a scalar error term with $E(\epsilon|x) = 0$. $p_n >> n$ is possible, in which case degrees of freedom would prevent fitting using ordinary least squares, but NOT using boosting.

L_2 boosting algorithm

1. Initializaion $\beta^0 = 0$, $f^0 = 0$, set m_{stop} the max number of iterations, and set m =0 (iteration counter).
2. (m+1)-th step $U_i^m = y_i - x_i'\beta^m$.
3. For each predictor variable $x_j, j = 1, \ldots, p$ calculate the correlation with the residuals and select the predictor that is most correlated with the residuals.
4. $\beta^{m+1} = \beta^m + \gamma_{j^m} e_{j^m}^m$ where $e_{j^m}^m$ is the index vector of the j-th predictor. So these are both vectors. $f^{m+1} = f^m + \gamma_{j^m}^m x_{j^m}^m$
5. increment m and repeat from step 2 as long a $m < m_{stop}$

Applications

- Instrumental variables (How do federal court decisions impact GDP, use characteristics of appelate court judges as instruments (z)
- Treatment selection (Does the GDP of different countries converge even if the initial wealth differs?). Problem is to select institutional and cultural variables as instruments.

Instrumental variables

$$y_i = \beta d_i + \epsilon_i,$$

$$d_i = z_i \Pi + \nu_i$$

$$(\epsilon, \nu)\ N(0, \Sigma), iid$$

$\Pi = C\tilde{\Pi}$ where $\tilde{\Pi}$ is a sparse design matrix. C is a regression coefficient that will need to be updated via boosting but is initally chosen according to a specific rule.

Treatment selection

$$y_i = d_i \alpha_0 + x_i' \theta_g + \xi_i$$

$$d_i = x_i' \theta_m + \nu_i$$

d_i here is the example corresponds to initial wealth levels and x_i to other institutional and cultural factors. These are the instruments that need to be selected for using boosting.

Software packages

- R package l2boost`https://cran.r-project.org/web/packages/l2boost/`
- Python KTBoost `https://github.com/fabsig/KTBoost`

- What are peoples impressions of the paper?
- Why linear models?
- What is being predicted here?
- No prediction(?), classification or clustering?
- What's going on?
- How does it compare to LASSO?