Iteratively reweighted least squares

The method of **iteratively reweighted least squares** (**IRLS**) is used to solve certain optimization problems with objective functions of the form of a *p*-norm:

$$\underset{\boldsymbol\beta}{\operatorname{arg\,min}} \sum_{i=1}^{n} \big| y_i - f_i(\boldsymbol\beta) \big|^p,$$

by an iterative method in which each step involves solving a weighted least squares problem of the form:[1]

$$\boldsymbol\beta^{(t+1)} = \underset{\boldsymbol\beta}{\operatorname{arg\,min}} \sum_{i=1}^{n} w_i(\boldsymbol\beta^{(t)}) \big| y_i - f_i(\boldsymbol\beta) \big|^2.$$

IRLS is used to find the maximum likelihood estimates of a generalized linear model, and in robust regression to find an M-estimator, as a way of mitigating the influence of outliers in an otherwise normally-distributed data set. For example, by minimizing the least absolute errors rather than the least square errors.

One of the advantages of IRLS over linear programming and convex programming is that it can be used with Gauss–Newton and Levenberg–Marquardt numerical algorithms.

Contents

Examples

L_1 minimization for sparse recovery

IRLS can be used for ℓ_1 minimization and smoothed ℓ_p minimization, $p < 1$, in compressed sensing problems. It has been proved that the algorithm has a linear rate of convergence for ℓ_1 norm and superlinear for ℓ_t with $t < 1$, under the restricted isometry property, which is generally a sufficient condition for sparse solutions.[2][3] However, in most practical situations, the restricted isometry property is not satisfied.

L^p norm linear regression

To find the parameters $\boldsymbol\beta = (\beta_1, \ldots,\beta_k)^T$ which minimize the L^p norm for the linear regression problem,

$$\underset{\boldsymbol\beta}{\arg\min}\big\|\mathbf{y}-X\boldsymbol\beta\big\|_p = \underset{\boldsymbol\beta}{\arg\min}\sum_{i=1}^n |y_i - X_i\boldsymbol\beta|^p,$$

the IRLS algorithm at step $t + 1$ involves solving the weighted linear least squares problem:[4]

$$\boldsymbol\beta^{(t+1)} = \underset{\boldsymbol\beta}{\arg\min}\sum_{i=1}^n w_i^{(t)} |y_i - X_i\boldsymbol\beta|^2 = (X^{\rm T}W^{(t)}X)^{-1}X^{\rm T}W^{(t)}\mathbf{y},$$

where $W^{(t)}$ is the diagonal matrix of weights, usually with all elements set initially to:

$$w_i^{(0)} = 1$$

and updated after each iteration to:

$$w_i^{(t)} = \big|y_i - X_i\boldsymbol\beta^{(t)}\big|^{p-2}.$$

In the case $p = 1$, this corresponds to least absolute deviation regression (in this case, the problem would be better approached by use of linear programming methods,[5] so the result would be exact) and the formula is:

$$w_i^{(t)} = \frac{1}{\big|y_i - X_i\boldsymbol\beta^{(t)}\big|}.$$

To avoid dividing by zero, regularization must be done, so in practice the formula is:

$$w_i^{(t)} = \frac{1}{\max\left\{\delta, \big|y_i - X_i\boldsymbol\beta^{(t)}\big|\right\}}.$$

where δ is some small value, like 0.0001.[5] Note the use of δ in the weighting function is equivalent to the Huber loss function in robust estimation. [6]

See also

- Feasible generalized least squares
- Weiszfeld's algorithm (for approximating the geometric median), which can be viewed as a special case of IRLS

Notes

1. C. Sidney Burrus, *Iterative Reweighted Least Squares (https://cnx.org/exports/92b90377-2b34-49e4-b26f-7fe572db78a1@12.pdf/iterative-reweighted-least-squares-12.pdf)*
2. Chartrand, R.; Yin, W. (March 31 – April 4, 2008). "Iteratively reweighted algorithms for compressive sensing". *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2008*. pp. 3869–3872. doi:10.1109/ICASSP.2008.4518498 (https://doi.org/10.1109%2FICASSP.2008.4518498).
3. Daubechies, I.; Devore, R.; Fornasier, M.; Güntürk, C. S. N. (2010). "Iteratively reweighted least squares minimization for sparse recovery". *Communications on Pure and Applied Mathematics*. **63**: 1–38. arXiv:0807.0575 (https://arxiv.org/abs/0807.0575). doi:10.1002/cpa.20303 (https://doi.org/10.1002%2Fcpa.20303).

4. Gentle, James (2007). "6.8.1 Solutions that Minimize Other Norms of the Residuals". *Matrix algebra*. Springer Texts in Statistics. New York: Springer. doi:10.1007/978-0-387-70873-7 (https://doi.org/10.1007%2F978-0-387-70873-7). ISBN 978-0-387-70872-0.
5. William A. Pfeil, *Statistical Teaching Aids (http://www.wpi.edu/Pubs/E-project/Available/E-project-050506-091720/unrestricted/IQP_Final_Report.pdf)*, Bachelor of Science thesis, Worcester Polytechnic Institute, 2006
6. Fox, J.; Weisberg, S. (2013),*Robust Regression (http://users.stat.umn.edu/~sandy/courses/8053/handouts/robust.pdf)*, Course Notes, University of Minnesota

References

- Numerical Methods for Least Squares Problems by Åke Björck (https://web.archive.org/web/20070810222123/http://www.mai.liu.se/~akbjo/LSPbook.html) (Chapter 4: Generalized Least Squares Problems.)
- Practical Least-Squares for Computer Graphics. SIGGRAPH Course 11 (http://graphics.stanford.edu/~jplewis/lscourse/SLIDES.pdf)

External links

- Solve under-determined linear systems iteratively (https://stemblab.github.io/irls/)