



Quantile Regression for Large-scale Applications

Jiyan Yang, Xiangrui Meng, Michael W. Mahoney

(Submitted on 1 May 2013)

Quantile regression is a method to estimate the quantiles of the conditional distribution of a response variable, and as such it permits a much more accurate portrayal of the relationship between the response variable and observed covariates than methods such as Least-squares or Least Absolute Deviations regression. It can be expressed as a linear program, and, with appropriate preprocessing, interior-point methods can be used to find a solution for moderately large problems. Dealing with very large problems, \emph{e.g.}, involving data up to and beyond the terabyte regime, remains a challenge. Here, we present a randomized algorithm that runs in nearly linear time in the size of the input and that, with constant probability, computes a $(1+\epsilon)$ approximate solution to an arbitrary quantile regression problem. As a key step, our algorithm computes a low-distortion subspace-preserving embedding with respect to the loss function of quantile regression. Our empirical evaluation illustrates that our algorithm is competitive with the best previous work on small to medium-sized problems, and that in addition it can be implemented in MapReduce-like environments and applied to terabyte-sized problems.

Comments: 32 pages; long version of a paper appearing in the 2013 ICML

Subjects: **Data Structures and Algorithms (cs.DS)**; Distributed, Parallel, and Cluster Computing (cs.DC); Numerical Analysis (cs.NA); Machine Learning (stat.ML)

Cite as: [arXiv:1305.0087 \[cs.DS\]](https://arxiv.org/abs/1305.0087)
(or [arXiv:1305.0087v1 \[cs.DS\]](https://arxiv.org/abs/1305.0087v1) for this version)

Submission history

From: Michael Mahoney [[view email](#)]

[v1] Wed, 1 May 2013 05:21:03 GMT (87kb)

Which authors of this paper are endorsers?

Link back to: [arXiv](#), [form interface](#), [contact](#).

Download:

- [PDF](#)
- [PostScript](#)
- [Other formats](#)

Current browse context:

cs.DS

[< prev](#) | [next >](#)

[new](#) | [recent](#) | [1305](#)

Change to browse by:

cs

[cs.DC](#)

[cs.NA](#)

stat

[stat.ML](#)

References & Citations

- [NASA ADS](#)

DBLP - CS Bibliography

[listing](#) | [bibtex](#)

[Jiyan Yang](#)

[Xiangrui Meng](#)

[Michael W. Mahoney](#)

Bookmark (what is this?)

