SQL

This is a resource page for the Statistical Quantile Learning (SQL) method described in Statistical Quantile Learning for Large, Nonlinear, and Additive Latent Variable Models. It estimates additive (deep) generative models or nonlinear factor models. Compared to machine learning models (such as VAE and GAN), SQL is:

- simple,
- scalable,
- fast,
- consistent,
- perform well for large and high-dimensional data (p large).

The algorithm is easy to use and does not necessitate the usage of libraries such as tensorflow or torch. SQL is currently available as R package. New features and improvements will be available soon.

Background

SQL estimates the additive model,

$$X_j = \sum_{l=1}^{q} g_{j,l}(Z_l) + \epsilon_j$$

where Z_l are normally distributed latent factors. More details are available in Statistical Quantile Learning for Large, Nonlinear, and Additive Latent Variable Models.

Installation

You can install the development version of SQL from GitHub with:

```
devtools::install_github("jbodelet/SQL/sql")
```

Usage

Fit additive factor models

```
library(sql)

# q= 1 factor:
sim <- simulate_afm(n = 150, p = 200)
sql <- SQL(sim$data)
hist(sql$factor, breaks = 30)
abs( cor(sim$factor, sql$factor) )
plot(sql)

# q= 2 factor:
q <- 2
sim <- simulate_afm(n = 150, p = 200, q = 2)

sql <- SQL(sim$data, q = 2, d= 6)
```

```
sql
abs( cor(sim$factor, sql$factor) )
```

Dependencies

The sql package depends on R libraries Matrix, matrixcalc, and Rfast.

Repository author

Julien Bodelet – post-doctoral researcher in Statistical learning, CHUV, Lausanne, Switzerland

Reference

Statistical Quantile Learning for Large, Nonlinear, and Additive Latent Variable Models

Cite as

Bodelet, Julien, Guillaume Blanc, Jiajun Shan, Graciela Muniz Terrera, and Oliver Y. Chen. 2023. *Statistical Quantile Learning for Large, Nonlinear, and Additive Latent Variable Models*. https://arxiv.org/abs/2003.13119.