Iteratively Reweighted PLS Estimation for Tensor Generalized Linear Regression

Haesong Choi

Florida State University
Department of Statistics

April 20, 2020

Overview

Partial Least Squares

Partial Least Squares

- Dateset : independent realizations of $\boldsymbol{X} \in \mathbb{R}^{n \times p}$ and $\boldsymbol{Y} \in \mathbb{R}^{n \times r}$
- The model assumes that the observations are iteratively projected into the low dimensional spaces of latent variables. (Wold, 1966)

$$\boldsymbol{X} = \boldsymbol{T}\boldsymbol{P}^\top + \boldsymbol{E}$$

$$\boldsymbol{Y} = \boldsymbol{T}\boldsymbol{Q}^\top + \boldsymbol{F}$$

- $\boldsymbol{T} = [\boldsymbol{t}_1, \ldots, \boldsymbol{t}_d] \in \mathbb{R}^{n \times d}$: the matrix of the latent variables,
  $\boldsymbol{P} \in \mathbb{R}^{p \times d}$ and $\boldsymbol{Q} \in \mathbb{R}^{r \times d}$: the matrices of the loading vectors
  $\boldsymbol{E}, \boldsymbol{F}$: residual matrices and d : the number of iterations
- PLS reduces the predictor $\boldsymbol{X}$ to a sufficiently lower dimensional latent matrix $\boldsymbol{T}$.
- Well suited for high dimensional regression because of its efficient solution for dimension reduction.

IRPLS Estimation for GLM

IRPLS Estimation for GLM

- Marx (1996) extends iteratively reweighted PLS into the framework of generalized linear models (GLM) by an iterative method in which each step involves solving a weighted least squares problem of the form.
- For GLM (McCullagh and Nelder, 1983),
  $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)^\top \in \mathbb{R}^{n \times p}$: Predictor matrix (not contain $\mathbf{1}_p$)
  $\boldsymbol{Y} \in \mathbb{R}^{n \times 1}$: independent random variables y_i which belongs to an exponential family with probability density (mass) function

$$y_i \sim f(y_i; \theta_i, \phi) = \exp\left[\left\{\frac{y_i\theta_i - b(\theta_i)}{a(\phi)}\right\} + c(y_i, \phi)\right] \tag{1}$$

where $\phi > 0$:a dispersion parameter, θ :the natural parameter of the distribution, and $a(\cdot), b(\cdot), c(\cdot)$: known functions

IRPLS Estimation for GLM

- In the GLM setting, the mean $\mu = E(\boldsymbol{Y}|\boldsymbol{X})$ is connected to a linear predictor $\boldsymbol{X}\boldsymbol{\beta}$ via

$$g(\mu_i) = \eta_i = \alpha_i + \boldsymbol{\beta}^\top \boldsymbol{x}_i \tag{2}$$

  where g is a strictly increasing or decreasing and twice differentiable function. η_i : the linear systematic part with intercept α and the coefficient vector $\boldsymbol{\beta} \in \mathbb{R}^p$.
- The log-likelihood for $\boldsymbol{Y}$ is

$$\ell(\alpha, \boldsymbol{\beta}; \boldsymbol{X}) = \sum_{i=1}^{n} \left\{ \frac{y_i\theta_i - b(\theta_i)}{\phi} \right\} + c(y_i, \phi), \tag{3}$$

  where $a(\phi) \equiv \phi$ is constant over all i.
- Solution for the likelihood estimator of $\boldsymbol{\beta}$: Iterative Fishers' Method of Scoring algorithm

IRPLS Estimation for GLM

- By introducing the diagonal matrix $\boldsymbol{W}^{(r)} = \mathrm{diag}(w_i^{(r)}) \in \mathbb{R}^{n \times n}$

$$w_i^{(r)} = \frac{1}{\phi V(\mu_i^{(r)})(g'(\mu_i^{(r)}))^2},$$

it turns out that the updates can be written as

$$\beta^{(r+1)} = (\boldsymbol{X}^\top \boldsymbol{W}^{(r)} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{W}^{(r)} \boldsymbol{z}^{(r)} \tag{4}$$

i.e. the estimated parameters for a weighted least squares regression of $\boldsymbol{z}^{(r)}$ on $\boldsymbol{X}$ with weights $\boldsymbol{W}^{(r)}$, where

$$\boldsymbol{z}^{(r)} = \boldsymbol{\eta}^{(r)} + (\boldsymbol{y} - \boldsymbol{\mu}^{(r)}) g'(\boldsymbol{\mu}^{(r)}). \tag{5}$$

Consequently we can obtain $\boldsymbol{\eta}^{(r+1)}$ by plugging in (2) via

$$\boldsymbol{\eta}^{(r+1)} = \boldsymbol{\alpha}^{(r)} \mathbf{1} + \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{W}^{(r)} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{W}^{(r)} \boldsymbol{z}^{(r)}. \tag{6}$$

IRPLS Estimation for GLM

- Latent factor models of $\boldsymbol{X}$ and $\boldsymbol{z}$ in (5) : With $d \ll p$,

$$\boldsymbol{X} = \sum_{j=1}^{d} \boldsymbol{t}_j \boldsymbol{p}_j^\top + \boldsymbol{E}_d \tag{7}$$

$$\boldsymbol{z} = \sum_{j=1}^{d} \boldsymbol{q}_j \boldsymbol{t}_j + \boldsymbol{f}_d \tag{8}$$

where $\boldsymbol{t}_j \in \mathbb{R}^n$ denotes latent variables, $\boldsymbol{p}_j \in \mathbb{R}^p$ is a loading vector and $\boldsymbol{E}_k$ is a residual matrix. The $\boldsymbol{q}_j$ are scalar coefficients, and $\boldsymbol{f}_d \in \mathbb{R}^n$ is an residual vector.

IRPLS Estimation for tensor GLM

IRPLS Estimation for tensor GLM

- For a response variable Y, m-dimensional tensor predictor $\boldsymbol{X} \in \mathbb{R}^{p_1 \times \cdots \times p_M}$ which follows a CANDECOMP/PARAFAC (CP) decomposition (a rank-R decomposition), the response is assumed to belong to an exponential family where the linear systematic part is of the form,

$$\begin{aligned} g(\mu) = g(E(Y|\boldsymbol{X})) &= \alpha + \langle \boldsymbol{B}, \boldsymbol{X} \rangle \\ &= \alpha + \langle \sum_{r=1}^{R} \beta_1^{(r)} \circ \cdots \circ \beta_M^{(r)}, \boldsymbol{X} \rangle && (9) \\ &= \alpha + \langle (\boldsymbol{B}_M \odot \cdots \odot \boldsymbol{B}_1)\mathbf{1}_R, \mathrm{vec}\boldsymbol{X} \rangle && (10) \end{aligned}$$

  where $g(\cdot)$ is a strictly increasing link function, $\boldsymbol{B}_m = [\beta_m^{(1)}, \ldots, \beta_m^{(R)}] \in \mathbb{R}^{p_m \times R}, m = 1, \ldots, M$, and $\boldsymbol{B}_M \odot \cdots \odot \boldsymbol{B}_1 \in \mathbb{R}^{\prod_m p_m \times R}$
- $\langle \boldsymbol{B}, \boldsymbol{X} \rangle = \langle \mathrm{vec}\boldsymbol{B}, \mathrm{vec}\boldsymbol{X} \rangle = \sum \beta_{i_1,\ldots,i_M} x_{i_1,\ldots,i_M}$, and $\boldsymbol{B} \approx [\![\boldsymbol{B}_1, \ldots, \boldsymbol{B}_M]\!]$.

IRPLS Estimation for tensor GLM

- The log-likelihood function given i.i.d. data $\{(y_i, \boldsymbol{x}_i), i = 1, \ldots, n\}$:

$$\ell(\alpha, \boldsymbol{B}_1, \ldots, \boldsymbol{B}_M; \boldsymbol{X}) = \sum_{i=1}^{n} \left\{ \frac{y_i \theta_i - b(\theta_i)}{\phi} \right\} + \sum_{i=1}^{n} c(y_i, \phi).$$

- Zhou (2013) introduced block relaxation algorithm for efficiently maximizing $\ell(\alpha, \boldsymbol{B}_1, \ldots, \boldsymbol{B}_M)$.
- A key point is that $g(\mu)$ is not linear in $(\boldsymbol{B}_1, \ldots, \boldsymbol{B}_M)$ jointly, but linear in $\boldsymbol{B}_m$ individually in (10). So, this algorithm alternatively updates $\boldsymbol{B}_m, m = 1, \ldots, M$, while keeping other parameters fixed.

IRPLS Estimation for tensor GLM

- Given i.i.d. $\{(y_i, \boldsymbol{x}_i^*), i = 1, \ldots, n\}$, when updating $\boldsymbol{B}_m$

$$
\begin{aligned}
g(\mu_i) &= \alpha + \langle \boldsymbol{B}, \boldsymbol{X} \rangle \\
&= \alpha + \langle \boldsymbol{B}_m, \boldsymbol{X}_{(m)}(\boldsymbol{B}_M \odot \cdots \odot \boldsymbol{B}_{m+1} \odot \boldsymbol{B}_{m-1} \odot \cdots \odot \boldsymbol{B}_1) \rangle \\
&= \alpha + \mathrm{vec}^\top \boldsymbol{B}_m \mathrm{vec}\{\boldsymbol{X}_{(m)}(\boldsymbol{B}_M \odot \cdots \odot \boldsymbol{B}_{m+1} \odot \boldsymbol{B}_{m-1} \odot \cdots \odot \boldsymbol{B}_1)\} \\
&= \alpha + \boldsymbol{\beta}^* \boldsymbol{x}_i^*.
\end{aligned}
$$

  where $\boldsymbol{\beta}^*$ is the coefficients for linear relationship between $g(\mu_i)$ and $\boldsymbol{x}_i^*$. (parameters $\mathrm{vec}\boldsymbol{B}_m \in \mathbb{R}^{p_m \cdot R}$)
- Then our new log likelihood for $\boldsymbol{Y}$ is

$$
\ell^*(\alpha, \boldsymbol{\beta}^*; \boldsymbol{X}^*) = \sum_{i=1}^{n} \left\{ \frac{y_i \theta_i - b(\theta_i)}{\phi} \right\} + c(y_i, \phi),
$$

  where $\boldsymbol{X}^* = (\boldsymbol{x}_1^*, \ldots, \boldsymbol{x}_n^*)^\top$ and we can apply IRPLS (Marx (1996)) to new GLM.

Algorithm 1: IRPLS Estimation for tensor GLS

Result: $\alpha, \boldsymbol{B}_1, \ldots, \boldsymbol{B}_M$

[0] Initialize $\alpha^{(0)} = \arg\max_\alpha \ell(\alpha, \mathbf{0}, \ldots, \mathbf{0})$, $\boldsymbol{B}_m^{(0)} \in \mathbb{R}^{p_m \times R}$ a random matrix for $m = 1, \ldots, M$.

repeat

for $m = 1, \ldots, M$ **do**

[1] Define $\boldsymbol{x}_i^* = \mathrm{vec}\{\boldsymbol{X}_{(m)}(\boldsymbol{B}_M \odot \cdots \odot \boldsymbol{B}_{m+1} \odot \boldsymbol{B}_{m-1} \odot \cdots \odot \boldsymbol{B}_1)\}$, $i = 1, \ldots, n$

[2] Initialize $\boldsymbol{E}_0 = \boldsymbol{X}^* \in \mathbb{R}^{n \times p_m \cdot R}$; $\boldsymbol{f}_0 = \psi(\boldsymbol{y}) \in \mathbb{R}^n$; $\boldsymbol{W} = \{\phi \boldsymbol{V}(\boldsymbol{\mu})\}^{-1} h'[\psi(\boldsymbol{y})]^2 \in \mathbb{R}^{n \times n}$

for $k = 1, \ldots, K$ **do**

[3] Define $\boldsymbol{r}_k$ = the coefficient for the WLS regression of $\boldsymbol{f}_{k-1}$ on $\boldsymbol{E}_{k-1}$ with $\boldsymbol{W}$.

[4] Define a latent variable $\boldsymbol{t}_k = \boldsymbol{E}_{k-1}\boldsymbol{r}_k \in \mathbb{R}^n$.

[5] Scale $\boldsymbol{t}_k$ with center (=Weighted mean of $\boldsymbol{t}_k$ with $\boldsymbol{W}$) and scale (=**SS**).

[6] Define $\boldsymbol{q}_k$ = the coefficient for the WLS regression of $\boldsymbol{f}_{k-1}$ on $\boldsymbol{t}_k$ with $\boldsymbol{W}$.

[7] Deflate the response vector $\boldsymbol{f}_k = \boldsymbol{f}_{k-1} - \boldsymbol{t}_k \boldsymbol{q}_k$.

[8] Define $\boldsymbol{E}_k$ = Residuals for the WLS regression of $\boldsymbol{E}_{k-1}$ on $\boldsymbol{t}_k$ with $\boldsymbol{W}$.

end for

[9] $\boldsymbol{\eta}$ = Weighted mean of $\boldsymbol{f}_0$ with weight $\boldsymbol{W}$ + $\sum_{k=1}^R \boldsymbol{q}_k \boldsymbol{t}$

[10] $\boldsymbol{W} = \{\phi \boldsymbol{V}(\boldsymbol{\mu})\}^{-1} h'[\psi(\boldsymbol{\eta})]^2$

until $\triangle \hat{\eta} < \epsilon$

[11] Define $\boldsymbol{\beta}_m^*$ = the coefficient for the Glm of $\boldsymbol{y}$ on $\boldsymbol{T} = (\boldsymbol{t}_1, \ldots, \boldsymbol{t}_K) \in \mathbb{R}^{n \times K}$

[12] Convert $\boldsymbol{\beta}_m^*$ into a matrix $\boldsymbol{B}_m^{(t+1)}$.

end for

[13] $\alpha^{(t+1)} = \arg\max_\alpha \ell(\alpha, \boldsymbol{B}_1^{(t+1)}, \boldsymbol{B}_2^{(t+1)}, \ldots, \boldsymbol{B}_M^{(t+1)})$

Discussion

Discussion

- IRPLS Estimation for tensor GLM with Tucker decomposition :
  For a response variable Y, m-dimensional tensor predictor $\boldsymbol{X} \in \mathbb{R}^{p_1 \times \cdots \times p_M}$,

$$g(\mu) = g(E(Y|\boldsymbol{X})) = \eta = \alpha + \langle \boldsymbol{B}, \boldsymbol{X} \rangle$$

- A tensor coefficient $\boldsymbol{B} \in \mathbb{R}^{p_1 \times \cdots \times p_M}$ follows a Tucker decomposition,

$$\boldsymbol{B} = \sum_{r_1=1}^{R_1} \cdots \sum_{r_M=1}^{R_M} g_{r_1,\ldots,r_M} \beta_1^{(r_1)} \circ \cdots \circ \beta_M^{(r_M)} = [\![\boldsymbol{G}; \boldsymbol{B}_1, \ldots, \boldsymbol{B}_M]\!], and$$

  where $\beta_m^{(r_m)} \in \mathbb{R}^{p_m}$ $m = 1, \ldots, M$, $r_m = 1, \ldots, R_m$ are all column vectors, $g_{r_1,\ldots,r_M}$ are constants, $\boldsymbol{G} \in \mathbb{R}^{R_1 \times \cdots \times R_M}$ and $\boldsymbol{B}_m \in \mathbb{R}^{p_m \times R_m}$.

Discussion

- When updating $\boldsymbol{B}_m \in \mathbb{R}^{p_m \times R_m}$,

$$
\begin{aligned}
\langle \boldsymbol{B}, \boldsymbol{X} \rangle &= \langle \boldsymbol{B}_{(m)}, \boldsymbol{X}_{(m)} \rangle \\
&= \langle \boldsymbol{B}_m, \boldsymbol{X}_{(m)} (\boldsymbol{B}_M \otimes \cdots \otimes \boldsymbol{B}_{m+1} \otimes \boldsymbol{B}_{m-1} \otimes \cdots \otimes \boldsymbol{B}_1) \boldsymbol{G}_{(m)}^\top \rangle \\
&= \mathrm{vec}^\top \boldsymbol{B}_m \mathrm{vec}\{ \boldsymbol{X}_{(m)} (\boldsymbol{B}_M \otimes \cdots \otimes \boldsymbol{B}_{m+1} \otimes \boldsymbol{B}_{m-1} \otimes \cdots \otimes \boldsymbol{B}_1) \boldsymbol{G}_{(m)}^\top \} \\
&= \beta^* \boldsymbol{x}^*.
\end{aligned}
$$

- It turns into a low-dimensional GLM with $\mathrm{vec}\boldsymbol{B}_m \in \mathbb{R}^{p_m R_m}$ parameters.
- When updating $\boldsymbol{G}$,

$$
\begin{aligned}
\langle \boldsymbol{B}, \boldsymbol{X} \rangle &= \langle \mathrm{vec}\boldsymbol{B}, \mathrm{vec}\boldsymbol{X} \rangle \\
&= \langle (\boldsymbol{B}_M \otimes \cdots \otimes \boldsymbol{B}_1) \mathrm{vec}\boldsymbol{G}, \mathrm{vec}\boldsymbol{X} \rangle \\
&= \langle \mathrm{vec}\boldsymbol{G}, (\boldsymbol{B}_M \otimes \cdots \otimes \boldsymbol{B}_1)^\top \mathrm{vec}\boldsymbol{X} \rangle \\
&= \beta^* \boldsymbol{x}^*.
\end{aligned}
$$

References

Marx, B, D. (1996). Iteratively reweighted partial least squares estimation for generalized linear regression. *Technometrics 38*, pages 374–381.

McCullagh, P. and Nelder, J. A. (1983). *Generalized Linear Models (Monographs on Statistics and Applied Probability)*. Chapman and Hall, London.

Wold, S. (1966). *Estimation of principle components and related models by iterative models iterative least squares*. Academic Press, NY.

Zhou, Hua., L. L. Z. H. (2013). Tensor regression with applications in neuroimaging data analysis. *Journal of the American Statistical Association 108*, pages 540–552.