Gradient Descent for Two-way Fixed Effects

Yiqing Xu, Minsheng Liu

March 10, 2017

Assume we have a linear system with m parameters and n observations which has two-way fixed effects, one with k and another l group values. We have the following linear system

$$\begin{aligned} \phi(\vec{\beta}, \vec{\alpha}, \vec{\chi}, \vec{\mu}) &= X\vec{\beta} + D_1\vec{\alpha} + D_2\vec{\chi} + \vec{\mu}\,, \\ \vec{y} &= \phi(\vec{\beta}, \vec{\alpha}, \vec{\chi}, \vec{\mu}) + \vec{\epsilon}\,, \end{aligned}$$

where

1. $\vec{y}, \vec{\mu}, \vec{\epsilon} \in \mathbb{R}^n$ denote the responses, intercepts[1], and errors;
2. $X \in \mathbb{R}^n \times \mathbb{R}^m$ represents the observed variables; $\vec{\beta} \in \mathbb{R}^m$ is the parameter vector to compute;
3. $D_1 : \mathbb{R}^k \to \mathbb{R}^n$ is the linear transformation(i.e. the dummy matrix) associating the k groups of the first fixed effect with the n observations; $\vec{\alpha} \in \mathbb{R}^k$ is for the first fixed effect;
4. $D_2 : \mathbb{R}^l \to \mathbb{R}^n$ and $\vec{\chi} \in \mathbb{R}^l$.

[1] Note that $\mu_i = \mu_j$ for $i \neq j$.

The function to minimize:

$$f(\vec{\beta}, \vec{\alpha}, \vec{\chi}, \vec{\mu}) = (\phi(\vec{\beta}, \vec{\alpha}, \vec{\chi}, \vec{\mu}) - \vec{y})^2\,.$$

We approach the solution by gradient descent. Given $\tilde{\beta}, \tilde{\alpha}, \tilde{\chi}, \tilde{\mu}$ of the last iteration, define

$$\begin{aligned} \bar{\beta} &= \phi(\tilde{\beta}, \tilde{\alpha}, \tilde{\chi}, \tilde{\mu}) - X\tilde{\beta} - \tilde{y}\,, \\ \bar{\alpha} &= \phi(\tilde{\beta}, \tilde{\alpha}, \tilde{\chi}, \tilde{\mu}) - D_1\tilde{\alpha} - \tilde{y}\,, \\ \bar{\chi} &= \phi(\tilde{\beta}, \tilde{\alpha}, \tilde{\chi}, \tilde{\mu}) - D_2\tilde{\chi} - \tilde{y}\,, \\ \bar{\mu} &= \phi(\tilde{\beta}, \tilde{\alpha}, \tilde{\chi}, \tilde{\mu}) - \tilde{\mu} - \tilde{y}\,. \end{aligned}$$

The new iteration is given as follows,

$$\begin{aligned} \hat{\beta} &= (X^\top X)^{-1}(X^\top \bar{\beta})\,, \\ \hat{\alpha}_i &= \frac{\mathsf{col}_i(D_1) \cdot \bar{\alpha}}{\mathsf{count}(\mathsf{col}_i(D_1))}\,, \\ \hat{\chi}_i &= \frac{\mathsf{col}_i(D_2) \cdot \bar{\chi}}{\mathsf{count}(\mathsf{col}_i(D_2))}\,, \\ \hat{\mu} &= \left(\frac{1}{n} \cdot \sum_i \bar{\mu}_i\right) \langle 1, \ldots, 1 \rangle\,, \end{aligned}$$

where count counts the number of non-zero entries in the vector and col is the column selector.