

Algorithm Check

Jiixin Hu

Email: jhu267@wisc.edu

Date: 2019.4.25

1 Semi-supervised tensor decomposition

Setting Consider binary response tensor $Y \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ and covariate matrix $X \in \mathbb{R}^{d_1 \times p}$. Let $d_1 = d_2 = d_3 = 20, p = 5$ and set rank of *tucker* decomposition be $r_1 = r_2 = r_3 = 3$. Model is :

$$\begin{aligned} \text{logit}(\mathbb{E}Y) &= B \times_1 X \\ B &= C \times_1 N_1 \times_2 N_2 \times N_3 \end{aligned}$$

Generate data Set seed *set.seed(0423)*. Generate N_1, N_2, N_3 with orthogonal columns by *randortho()*; generate C by random gaussian entries *rmnorm(mean = 0, sd = 20)*; X can be generated by gaussian entries or binomial entries; Y can be generated by binomial distribution with probability related to B .

Initialization To get initial B , consider a glm model with vectorized Y as response and X as predictor. Then use *tucker* to get the initial C, N_1, N_2, N_3 .

Verify

- If X are from binomial entries through *rbinom(size = 1, prob = 0.5)*.
 1. The true likelihood calculated by true parameters is -3384.002; the algorithm gave likelihood convergent to -3310.752.
 2. The algorithm seems gave a good fit. See graphs(Figure 1,2):
- If X are from Gaussian entries through *rmnorm(sd = 15)*.
 1. The true likelihood calculated by true parameters is -238.7939; the algorithm gave likelihood convergent to -126.9359.
 2. When using Gaussian covariate matrix, may meet over-fitting problem. Graphs to show that(Figure 3, 4):

The scatter plots show that there are still positive relationship between the true parameter and estimate. But the estimate tends to be extremely large when true value has a relatively large value.

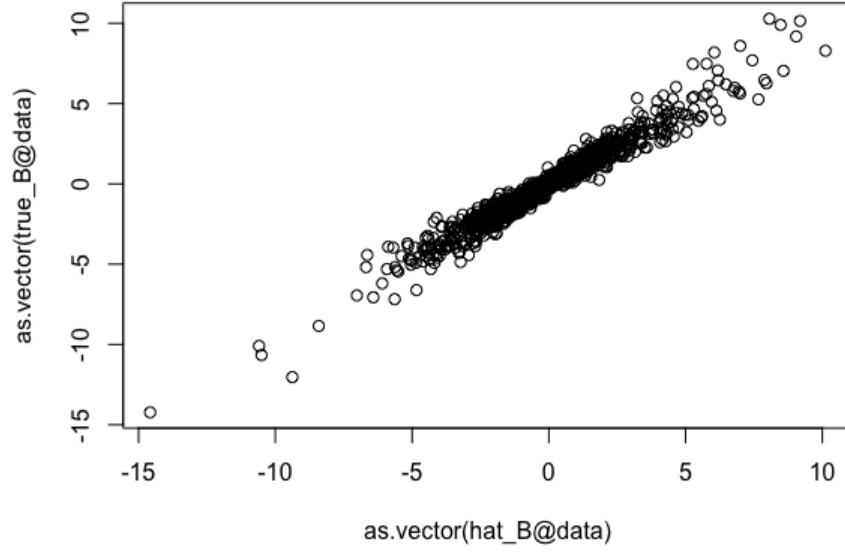


Figure 1: Scatterplot between vectorized true B and estimate B when use Binomial entries as covariate. Almost all the points are on the diagonal and that implies a good fit.

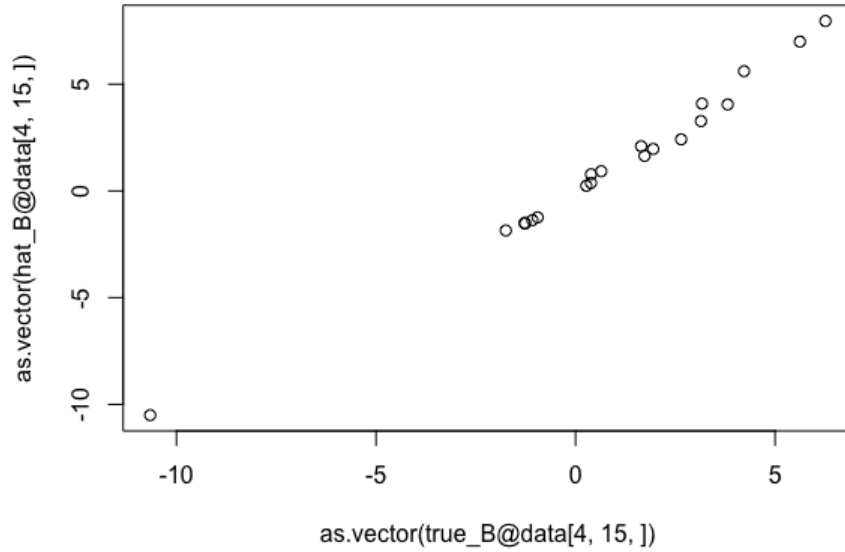


Figure 2: Scatterplot between slide[4,15,] of true B and estimate B when use Binomial entries as covariate. 20 points are nearly on the diagonal and that implies a good fit.

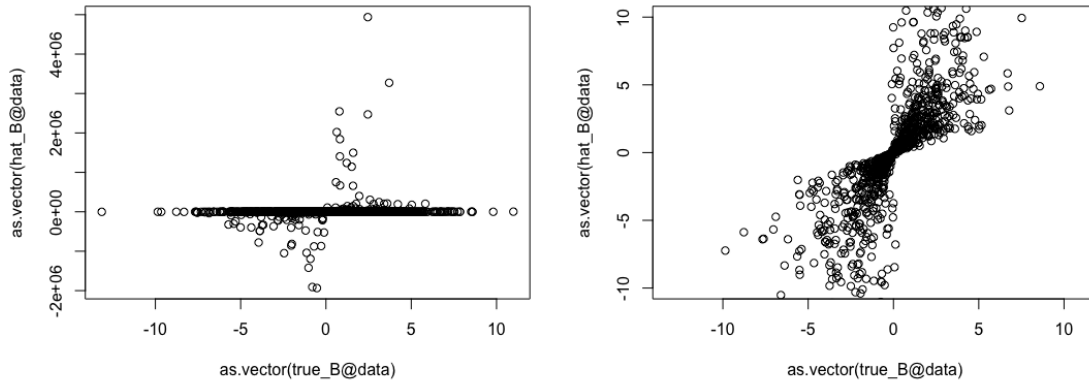


Figure 3: Scatterplot between vectorized true B and estimate B when use Gaussian entries as covariate; right plane is zoomed from left plane with constrain $ylim = c(-10, 10)$.

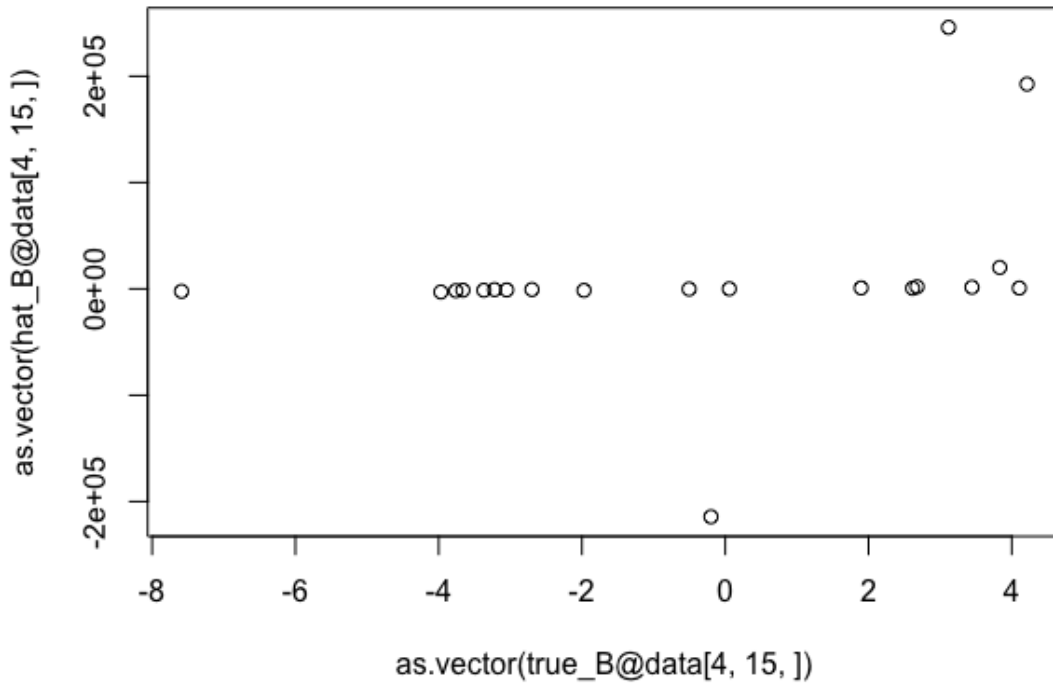


Figure 4: Scatterplot between slide[4,15,] of true B and estimate B when use Gaussian entries as covariate.