Sanity check: Matrix response regression

Jiaxin Hu

May 13, 2021

1 Data generated from Mreg model

Here we consider the case $d_1 = d_2 = 20, d_3 = 50, p = 5$, the sparsity, s, varies in $(0.1, 0.5, 0.9)$ and the rank of Θ, R, varies in $(3, 6)$. The binary observations are generated by Mreg model. We fit the Mreg model with true parameters, and fit STD model with $r_3 = p$ and choose $r_1 = r_2$ in $(3, 6, 9)$ with best performance. See Figure 1 and 2.

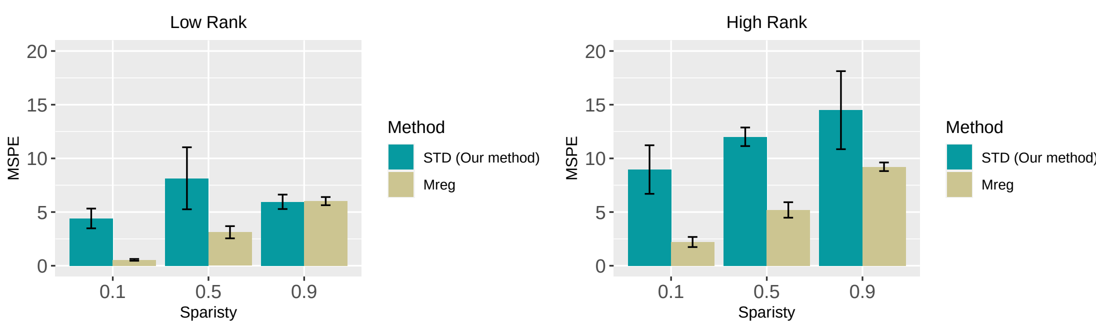

Figure 1: omparison between STD method and Mreg method. The y-axis is the PMSE and x-axis is the sparsity. Consider rank $R = 3$(low) and $R = 6$(high).

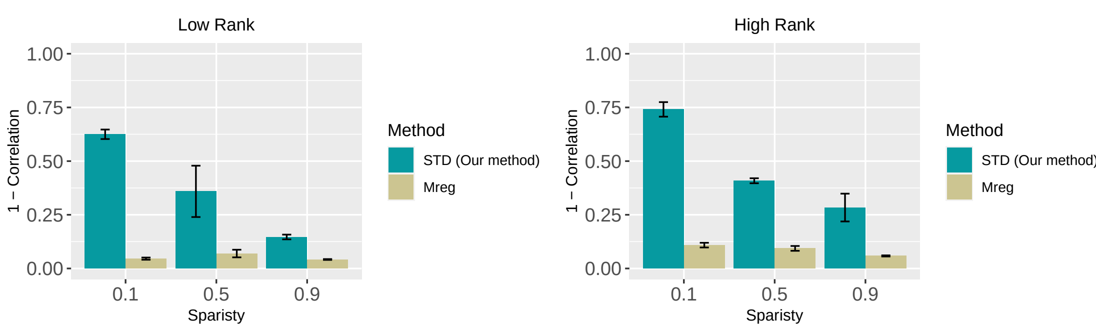

Figure 2: omparison between STD method and Mreg method. The y-axis is the 1 - correlation and x-axis is the sparsity. Consider rank $R = 3$(low) and $R = 6$(high).

Note that in most cases, Mreg outperforms than our STD method. However, the PMSE and correlation trends do not align as the sparsity goes larger (s increases). I believe the trend of PMSE makes sense for Mreg, since the model complexity goes larger as s increases. The scatter plot also confirm this.

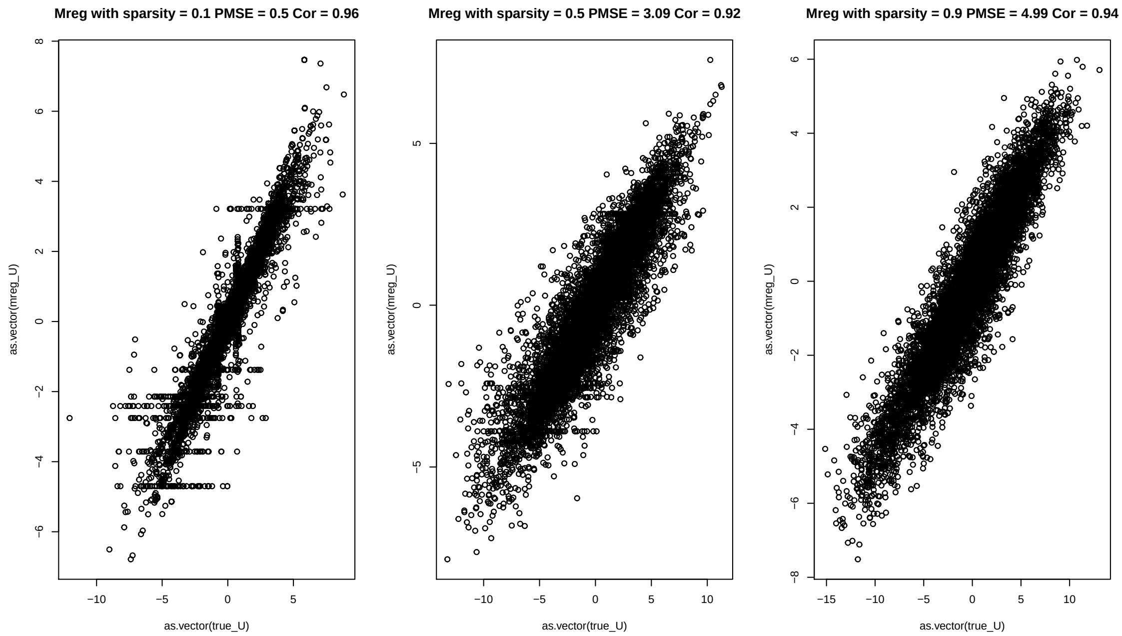

Figure 3: Scatter plot between estimated linear predictor and the true linear predictor $\Theta + \mathcal{B} \times_3 \boldsymbol{X}$.

Overall, this sanity check makes sense since Mreg has good performance when the data is generated from Mreg model. Here remains a question: how to explain the STD performance as the sparsity varies?

2 Data generated by STD model

In previous simulations, in the low signal settings, we know that Mreg outperforms than our model, which is counterintuitive because the data is generated by our STD model. By accident, I found the ratio between feature matrix and the tensor dimension affects the performance of Mreg and our STD model.

Here we consider the case $d_1 = d_2 = d_3 = 20, \boldsymbol{r} = c(2, 2, 2)$ and the signal $\alpha = 3$. Let the ratio $\gamma = p/d_3$ vary in $(0.1, 0.3, 0.5, 0.7)$. Data is generated by STD model. We fit the STD model with true parameters, and fit the Mreg model with $R = 2$ and various sparsities $s = (0.1, 0.5, 1)$. See Figure 4.

By the Figure 4, in terms of PMSE, our STD only outperforms when $\gamma \leq 0.3$. This consists with the previous weird conclusion, since we set $\gamma = 0.4$ in previous simulations. Meanwhile, in terms of correlation, our STD always outperforms than Mreg, which implies our model reflects the true linear predictors better.

Since our coefficient tensor generated by STD is not sparse, Mreg with $s = 1$ should have better performance than the other cases. The right panel in Figure 4 confirms that. However the right panel in Figure 4 also implies that Mreg with no sparsity still can not estimate the coefficient tensor very well compared with our STD. I believe this situation is caused by the model misspecification.

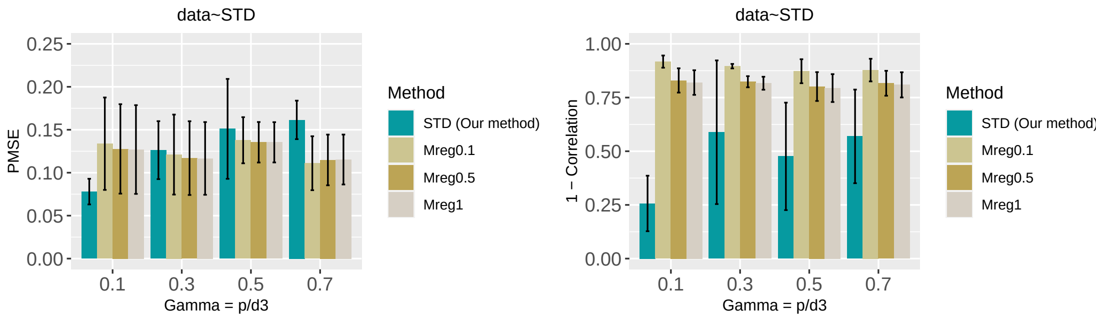

Figure 4: Comparison between STD method and Mreg methods with sparsity $s = 0.1, 0.5, 1$. The left panel records the PMSE performance vs the γ, and the right panel records the $1 - \text{Correlation}$ vs the γ.

3 Invariance of feature encoding

Here we claim that Mreg is not invariant to the feature encoding. Note that our STD algorithm will give different estimations even with the same data. So the following results with STD will change if we use different seeds.

3.1 Gender

Assume there is only one feature about the Gender. There are two ways to encoding the gender information: (1) $X_1 = (1, ..., 1, 0, ..., 0)^T$, and (2) $X_2 = (0, ..., 0, 1, ..., 1)^T$. In X_1 we encodes female as 1 and male as 0 while X_2 switches the encoding. We generate data from Mreg model with $d_1 = d_2 = d_3 = 10, s = 0.1, R = 2$ with X_1. We fit Mreg and STD with X_1 and X_2 respectively, and then compare the estimate matrices for the difference between female and male. See Figures 5, 6, and 7. For clearer visualization, we let the left panels subtract the right panels and calculate the MSE between these two panels. See Figure 8.

3.2 Age

In gender case, if we standardize two encodings, Mreg will give the same estimation. Now, we claim that standardization not always work to address the lack of invariance to feature encoding.

Assume there are two features about the Age: Age 1, Age 2, and Age 3. There are two possible encodings.

$$X_1 = \begin{bmatrix} 1_5 & 0 \\ 0 & 1_5 \\ 0 & 0 \end{bmatrix}, \quad X_2 = \begin{bmatrix} 1_5 & 0 \\ 0 & 1_5 \\ -1_5 & -1_5 \end{bmatrix}.$$

X_1 encodes Age 3 with $(0, 0)$ and X_2 encodes Age 3 with $(-1, -1)$. We standardize X_1 and X_2 to make their columns with mean 0 and variance 1, denoted as $\tilde{X}_1, \tilde{X}_2$, respectively. Note that the $\tilde{X}_1 \neq \tilde{X}_2$. We generated the data from Mreg model with $d_1 = d_2 = 10, d_3 = 15, s = 1, R = 2$ with $\tilde{X}_1$. We fit Mreg and STD with X_1 and X_2 respectively, and then compare the estimate matrices for the difference between Age 1 and Age 2. See Figures 9, 10, and 11. For clearer visualization, we let the left panels subtract the right panels and calculate the MSE between these two panels. See Figure 12.

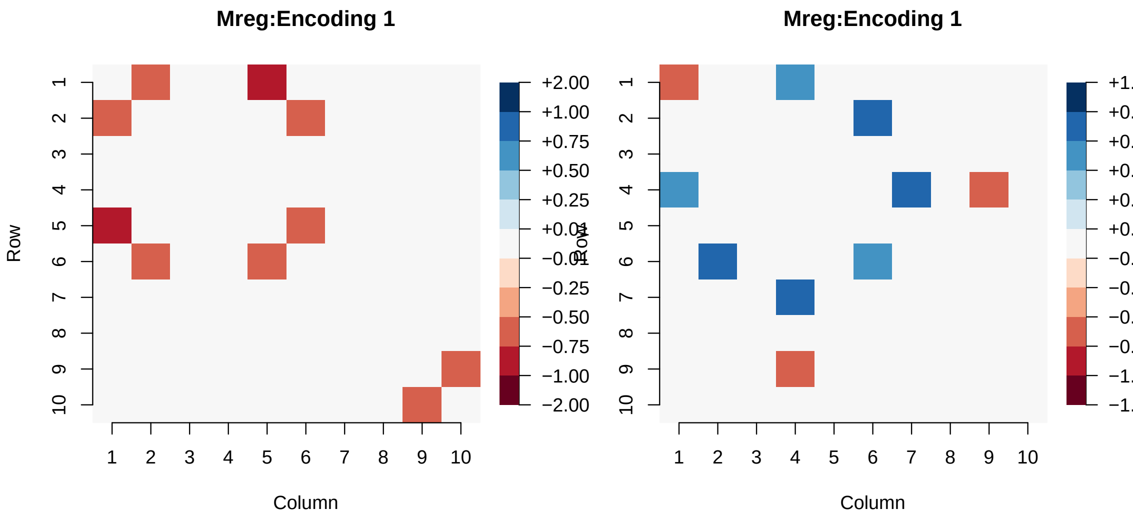

Figure 5: Estimate matrices of Mreg for the difference between female and male with difference encodings.

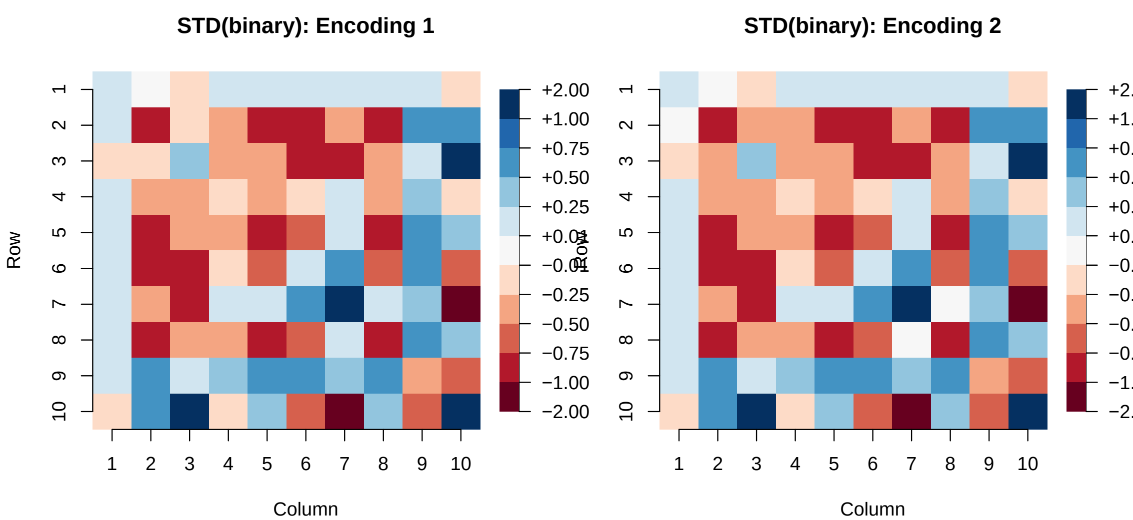

Figure 6: Estimate matrices of binary STD for the difference between female and male with difference encodings.

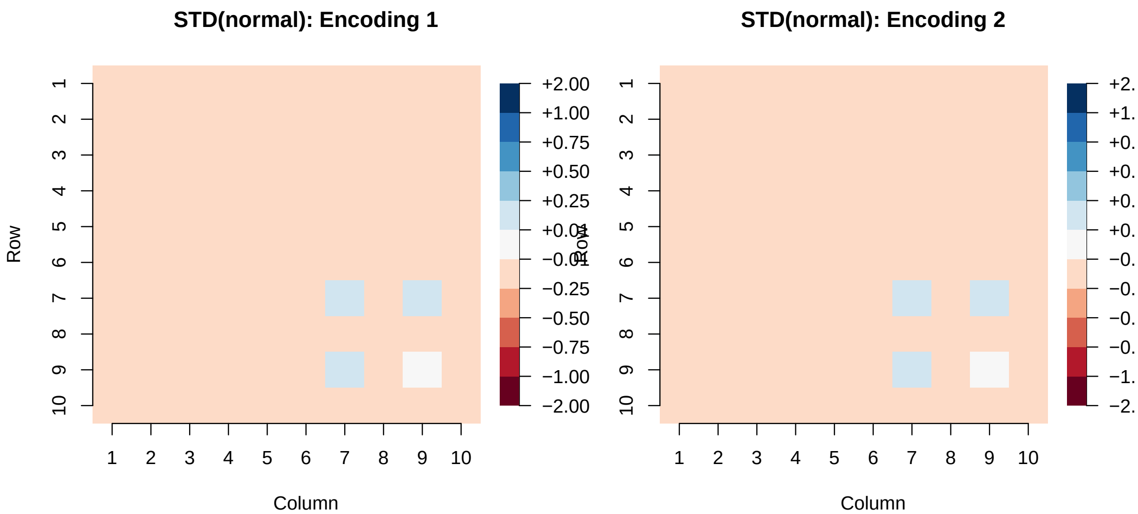

Figure 7: Estimate matrices of normal STD for the difference between female and male with difference encodings.

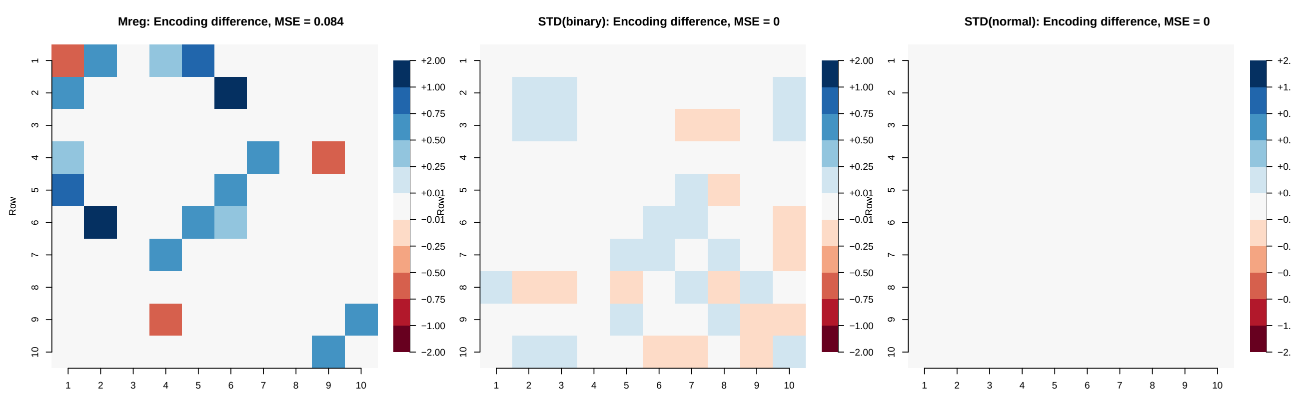

Figure 8: Subtractions between left panel and right panel in Figures 5, 6, and 7.

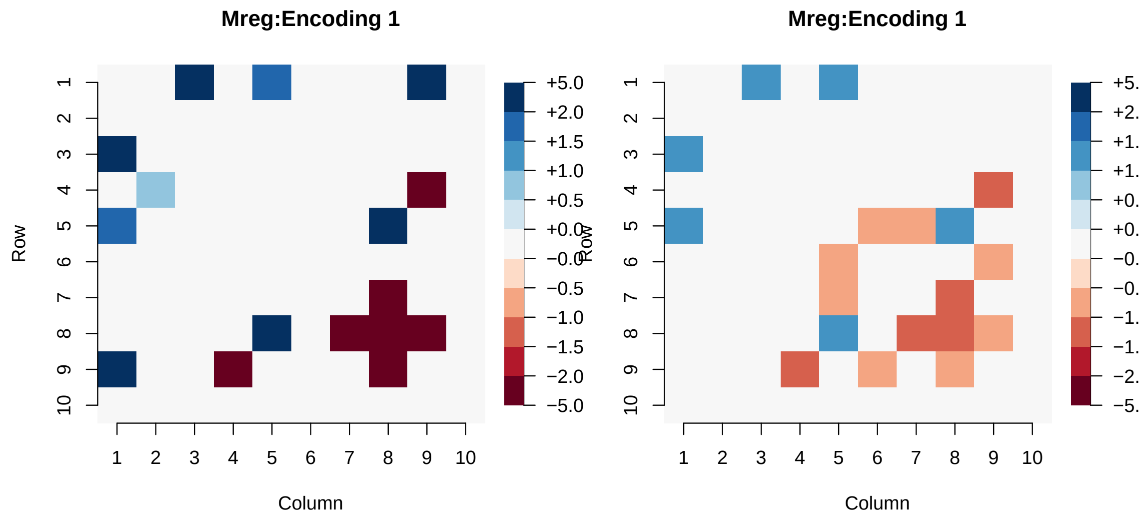

Figure 9: Estimate matrices of Mreg for the difference between Age 1 and Age 2. with difference encodings.

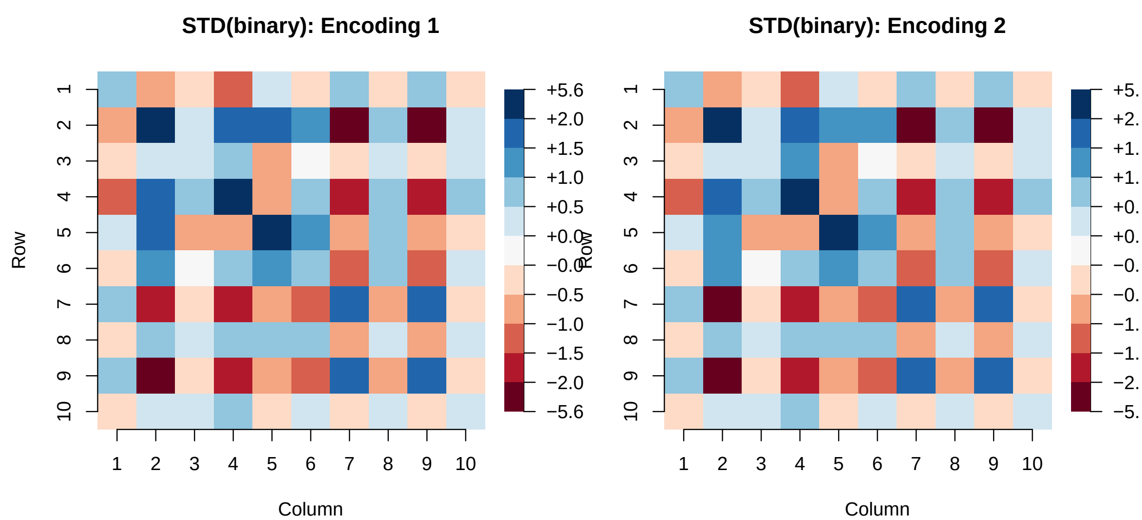

Figure 10: Estimate matrices of Mreg for the difference between Age 1 and Age 2. with difference encodings.

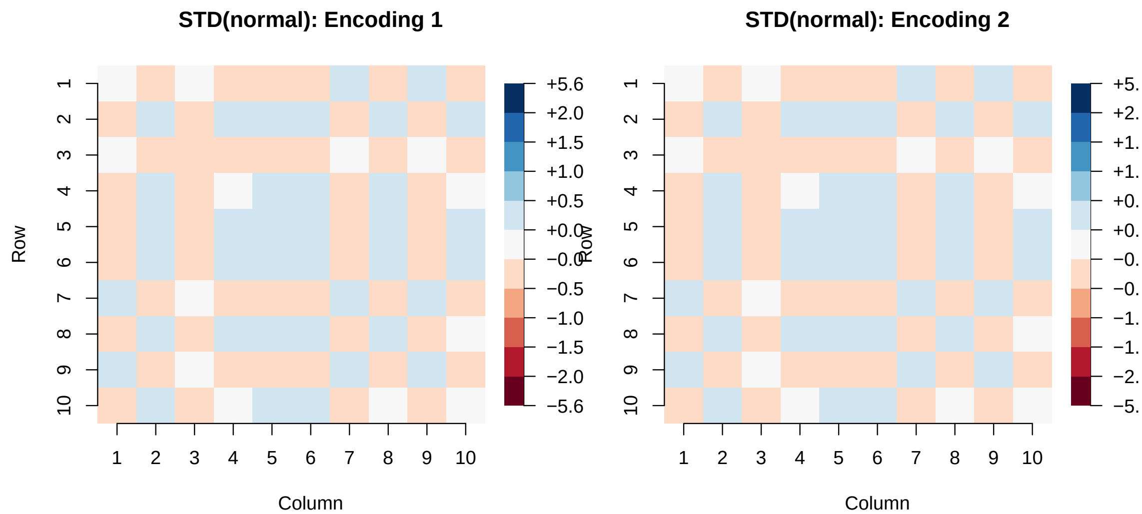

Figure 11: Estimate matrices of Mreg for the difference between Age 1 and Age 2. with difference encodings.

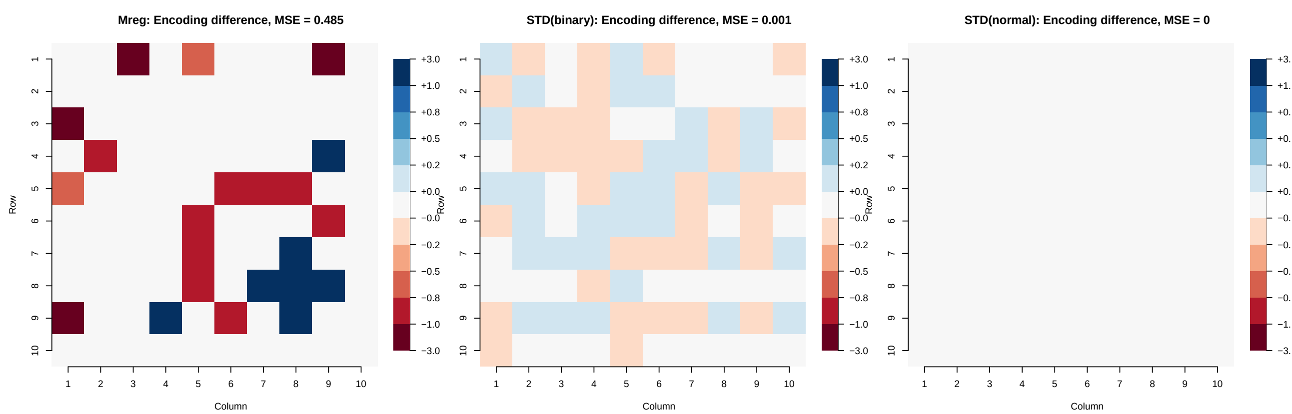

Figure 12: Subtractions between left panel and right panel in Figures 9, 10, and 11.