1 Discussion

We have developed the learning framework for the relationship between a binary label response and a high-dimensional matrix-valued predictor. Our method respects the matrix structure of the predictors and provide interpretable prediction via a nonparametric approach. The theoretical and numerical results demonstrate the competitive performance of our method.

~~In this section,~~ we discuss some extensions of our current framework and possible applications. While our focus is on matrix-valued predictors and binary responses, the developed results can be generalized to tensor-valued predictors and continuous responses. This extension will be helpful in a broader variety of applications. In addition, we show that matrix completion problem seemingly unrelevant to our method can be viewed as a special case in our framework. We believe that further exploration of the benefits of learning reduction approach in many tasks would be ~~necessary.~~ important

Extension from matrix-valued predictors to tensor-valued one: ~~Although we have presented the work in the context of matrix-valued predictors,~~ we may consider extensions to other nonconventional predictors, such as images, graphs, tensors, and functional time-series. Here we provide one way to extend matrix-valued predictors to ~~tensor valued one~~. ~~In the case~~ when $\mathcal{X} \in \mathbb{R}^{d_1 \times d_2 \times \cdots d_K}$, we generalize the function family $\mathcal{F}(r, s_1, s_2)$ to tensors

$$\mathcal{F}(\boldsymbol{r}, \boldsymbol{s}) = \{f \colon \mathcal{X} \mapsto \langle \mathcal{X}, \mathcal{B} \rangle + b | \text{rank}(\mathcal{B}) \leq \boldsymbol{r}, \text{supp}(\mathcal{B}) \leq \boldsymbol{s}, \mathcal{B} \in \mathbb{R}^{d_1 \times d_2 \times \cdots \times d_K}, b \in \mathbb{R}\},$$

where the rank of tensor $\boldsymbol{r} = (r_1, \ldots, r_K)$ is defined by Tucker decomposition and $\text{supp}(\mathcal{B}) = (s_1, \ldots, s_K)$ denotes the K-way sparsity parameter with s_k ~~meaning~~ the number of non-zero columns in k-th mode. The Tucker low-rankness is popular~~ly imposed~~ in tensor analysis. Unlike matrices, there are various notions of tensor low-rankness, such as CP rank (**?**) and train rank (**?**). We utilize Tucker decomposition here but one can choose the low-rank structure on tensors depending on applications. With this extension, we expect to handle regression problems with tensor-valued predictors such as image recognition, context-based recommendation systems, and so on.

Extension from binary valued responses to continuous valued one: We may also ask whether the results here, provided in the setting of binary regression with $y \in \{-1, 1\}$, may be extended to a continuous response $y \in \mathbb{R}$. The answer is affirmative if we assume the response y is bounded, e.g. $y \in [-L, L]$, for $L > 0$. One possible solution is to use response-dependent weight in place of response-dependent weight in the classification. Specifically, for a fixed target level $\pi \in [-L, L]$, define a new binary response $\tilde{y} = \text{sign}(y - \pi)$, a response-dependent weight $\tilde{w}_\pi(y) = |y - \pi|$, and a general weighted classification risk

$$\tilde{R}_\pi(S) \stackrel{\text{def}}{=} \mathbb{E}\left[\tilde{w}_\pi(y) \mathbb{1}(\tilde{y} \neq \text{sign}(\boldsymbol{X} \in S))\right].$$

The risk (**??**) extends the π-weighted classification risk (**??**) for a continuous response, where the weight $\tilde{w}_\pi(y) = |y - \pi|$ is the distance from the response y to the target level π. Importantly, the level set $S(\pi) = \{\boldsymbol{X} \in \mathcal{X} \colon \mu(\boldsymbol{X}) \geq \pi\}$ is the global minimum of (**??**) under conditional model of the type $y|\boldsymbol{X} = \mu(\boldsymbol{X}) + \varepsilon$, where the noise ε is a mean-zero random variable whose distribution is allowed to depend on $\boldsymbol{X}$ (**??**). With this statistical characterization, our main result on excess regression risk bound in Section **??** still holds. Therefore, our learning reduction approach equally applies to a continuous response by using $\tilde{w}_\pi(y)$ and $\tilde{y}$ in place of $w_\pi(y)$ and y, respectively. For an unbounded response, it is unclear whether the level set approach still achieves accuracy guarantees. We leave these questions for future work.

Application to nonparametric matrix completion problem: Our nonparametric regression can be ~~utilized~~ to solve matrix completion problem. Consider the ~~specific~~ problem to predict the unobserved entries of a binary matrix $\boldsymbol{Y} \in \{0,1\}^{d_1 \times d_2}$, where observed entries are labeled as $\Omega \subset [d_1] \times [d_2]$. We ~~can~~ re-express the matrix completion problem as classification problem with the training set $\{\boldsymbol{E}_{ij}, \boldsymbol{Y}_{ij}\}_{(i,j)\in\Omega}$ where $\boldsymbol{E}_{ij}$ is an indicator matrix whose (i,j)-th entry is 1 and the other entries are 0. This representation ~~makes the matrix completion problem~~ is a special case of the regression problem with indicator matrices as predictors . We ~~can successfully~~ estimate the missing entries by allow predicting the binary responses of unobserved indicator matrices. In addition, we ~~can successfully handle~~ not only ~~completion problem of~~ binary valued matrices but also continuous valued matrices or tensors ~~based on above mentioned extensions.~~ This promising application brings the nonparametric advantages of flexibility and robustness to the completion problem.