Regularization for uplift regression supplementary material

Krzysztof Rudaś[1,2] and Szymon Jaroszewicz[1,2]

[1] Institute of Computer Science, Polish Academy of Sciences
[2] Faculty of Mathematics and Information Science, Warsaw University of Technology

1 Additional plots of regularization regions

Figure 1 shows regions analogous to Figure 1 in the main paper for the squared L_2 norm regularization. For easier comparison, regularization regions in L_1 and L_2 norm for different types of interaction models for $\lambda_1 = \lambda_2 = 1$ and $p = 1$ are superimposed in Figure 2.

2 Proof of Theorem 2

Proof. $\lambda_2 \to \infty$ implies that the optimal estimator must satisfy $\beta^S = 0$. The form of the model thus simplifies to

$$y_i = \left(t_i - \tfrac{1}{2}\right) x_i \beta^U + \varepsilon_i,$$

which leads to the estimator

$$\begin{aligned}
\hat{\beta^U} &= \arg\min_{\beta^U} \sum_{i=1}^{n} \left(y_i - \left(t_i - \tfrac{1}{2}\right) x_i \beta^U\right)^2 + \lambda_1 \|\beta^U\|_q^q \\
&= \arg\min_{\beta^U} \frac{1}{4} \sum_{i=1}^{n} \left(\frac{y_i}{t_i - \frac{1}{2}} - x_i \beta^U\right)^2 + \lambda_1 \|\beta^U\|_q^q \\
&= \arg\min_{\beta^U} \sum_{i=1}^{n} \left(\bar{y}_i - x_i \beta^U\right)^2 + 4\lambda_1 \|\beta^U\|_q^q,
\end{aligned}$$

where the second equality is obtained by factoring out $t_i - \frac{1}{2}$ and noting that $(t_i - \frac{1}{2})^2 = \frac{1}{4}$. □

3 Coefficients used to simulate synthetic data

Values of β^C, β^U, and β^T for different scenarios used in experiments are presented in the table below

	β^C	β^U	β^T
Scenario 1	$(-4, -4, -4, 0, ..., 0)$	$(7.8, 7.8, 7.8, 0, ..., 0)$	$(3.8, 3.8, 3.8, 0, ..., 0)$
Scenario 2	$(-4, -4, -4, 0, ..., 0)$	$(3.8, 3.8, 3.8, 0, ..., 0)$	$(-0.2, -0.2, -0.2, 0, ..., 0)$
Scenario 3	$(-0.2, -0.2, -0.2, 0, ..., 0)$	$(-3.8, -3.8, -3.8, 0, 0, 0)$	$(-4, -4, -4, 0, ..., 0)$
Scenario 4	$(-4, -4, -4, 0, ..., 0)$	$(-0.2, -0.2, -0.2, 0, 0, 0)$	$(-3.8, -3.8, -3.8, 0, ..., 0)$

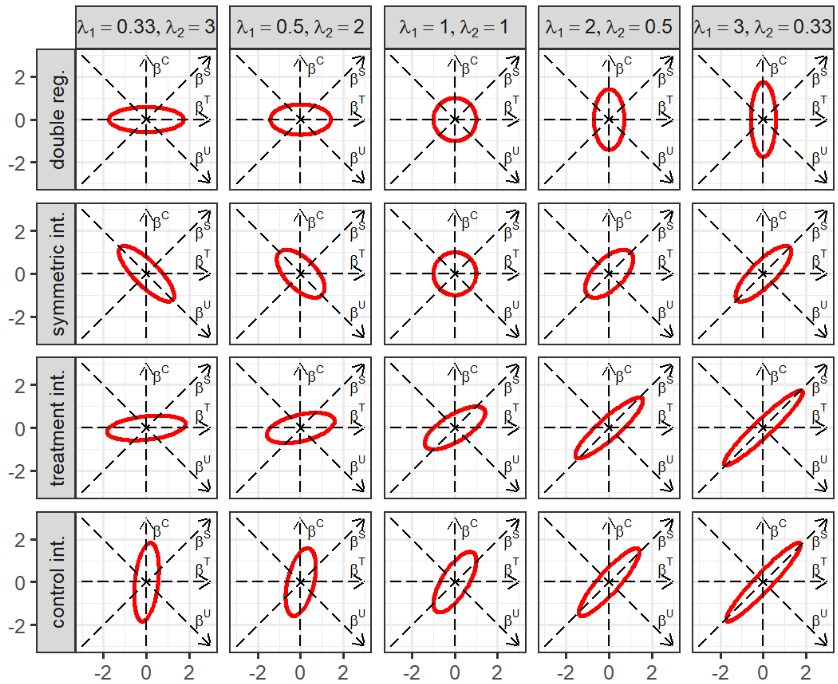

Fig. 1: Regularization regions in L_2 norm for different types of estimators and different parameters λ_1, λ_2 for $p = 1$

4 Experimental results for L_2 norm regularization on synthetic data

Figure 3 presents the MSE of L_2 regularized models for the four scenarios of synthetic data generation. Comparing with analogous results for the L_1 norm in Figure 2 of the main paper, the main difference is the relatively good performance of the unregularized method. Only when β^U is small regularization works significantly better. Another difference is that when $\beta^T \approx 0$ or $\beta^C \approx 0$ the double regularized model does not work as well as it did for L_1 regularization.

5 Results on Lalonde dataset

The second dataset we consider is the well known Lalonde dataset [1] describing the effects of a job training program which addressed a population of low skilled adults. A randomly selected sample of the population was invited to take part in a job training program. Their income in the third year *after* randomization is the target variable. Our goal is to build a model predicting how effective will the program be for a given individual. There are a total of 297 treatment records and 425 controls.

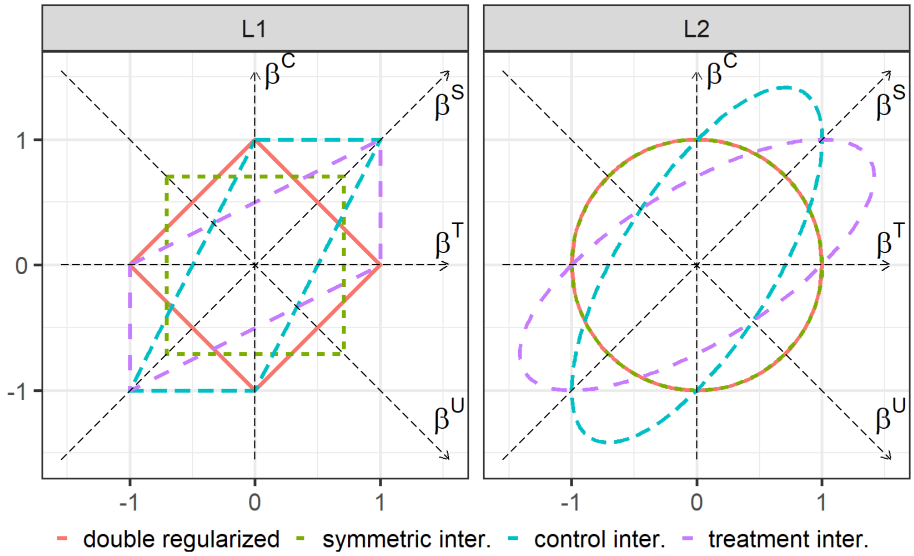

Fig. 2: Regularization regions in L_1 and L_2 norms for different types of interaction models for $\lambda_1 = \lambda_2 = 1$ and $p = 1$

Figure 4 shows the results. For this dataset, the performance of all of regularizers is comparable (for both L_1 and L_2 regularization) but we may notice that the symmetric interaction method obtains slightly better results for the L_2 case.

References

1. Lalonde, R.: Evaluating the econometric evaluations of training programs. American Economic Review **76**, 604–620 (1986)

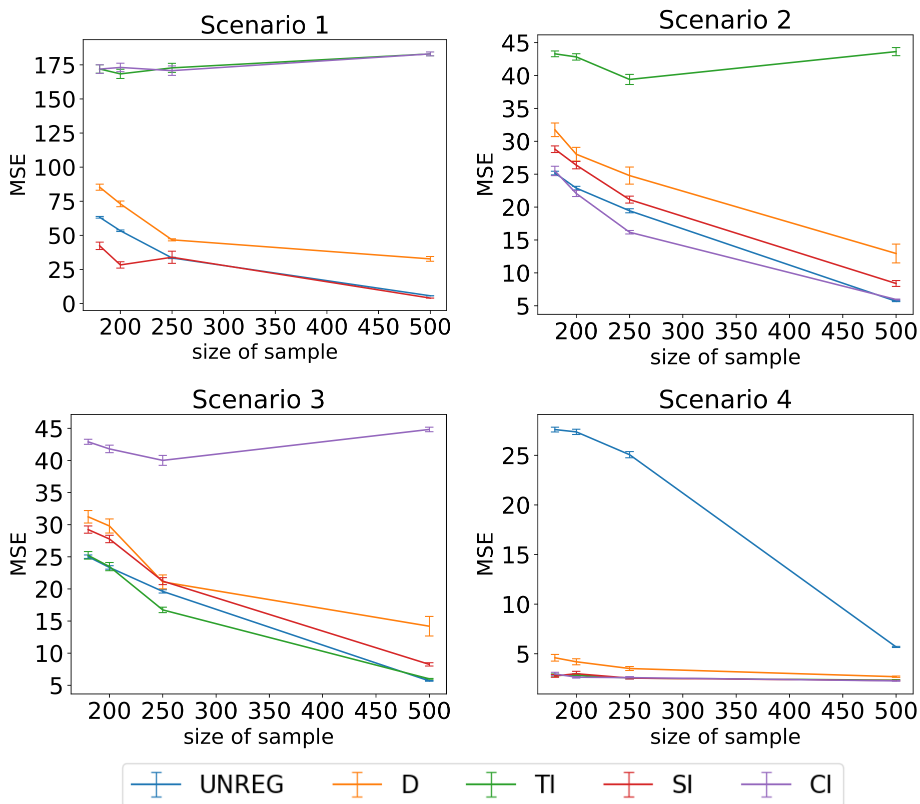

Fig. 3: Predictive MSE of estimators with L_2 penalty under different simulation scenarios

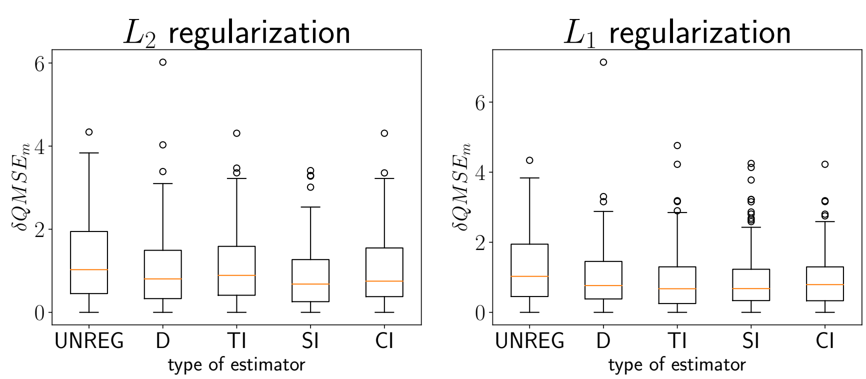

Fig. 4: Results for the Lalonde dataset