A Multiplicative Multitask Lasso approach with task descriptor variables.

Víctor Bellón VICTOR.BELLON@MINES-PARISTECH.FR
Chloé-Agathe Azencott CHLOE-AGATHE.AZENCOTT@MINES-PARISTECH.FR
Véronique Stoven VERONIQUE.STOVEN@MINES-PARISTECH.FR
MINES ParisTech, PSL-Research University, CBIO-Centre for Computational Biology, 35 rue St Honor 77300 Fontainebleau, France

Institut Curie, 75248 Paris Cedex, France

INSERM U900, 75248 Paris Cedex, France

Abstract

The principle of multitask learning framework is that, instead of learning different but related tasks independently, they are learned jointly by sharing information between these tasks. For example, when a robot learns to walk over different grounds, the algorithm is similar for walking on ice or on the pavement, but requires a specific adaptation for each type of ground.

Different regularized approaches have been proposed for features selection in multitask learning such as the Multitask Lasso (Obozinski et al., 2006). This approach employs a block regularization over the weights, imposing to select the same features for all the tasks. However, many problems require more flexibility to allow taking into account some specificities for each task.

In (Lozano & Swirszcz, 2012) the authors propose avoid this limitation by formulating each regression coefficient as the product of two factors. One factor controls the overall sparsity and captures features common to all tasks. The second factor reflects the tasks specificities. Based on the intuition that the second factor should be similar for similar tasks, we suggest to characterize each task by a set of descriptor variables and rewrite the second factor as a linear combination of these descriptor variables. Hence we can account more finely for the structure between the tasks.

Given K tasks and their corresponding datasets (X^k, Y^k) with $X^k \in \mathbb{R}^{n^k \times p}$ and a vector of descriptor variables $D^k \in \mathbb{R}^L$ for each task k we minimize the following function:

$$\min_{\theta>0,\alpha} \frac{1}{2}\sum_{k=1}^{K}\frac{1}{n^k}\sum_{i=1}^{n^k}\left\| Y_i^k - \sum_{j=1}^{p}\theta_j\left(\sum_{l=1}^{L}\alpha_{j,l}D_l^k\right)X_{i,j}^k\right\|_2^2 + \lambda_1\sum_{j=1}^{p}|\theta_j| + \lambda_2\sum_{j=1}^{p}\sum_{l=1}^{L}|\alpha_{j,l}|$$

where θ controls the overall sparsity of the model and α are the weights of the second factor that reflects task specificity.

An advantage of our method is that it allows prediction for tasks that have not been learnt, as long as this new task can be represented using the same descriptor variables. Note that in the case where $L = k$ and $D_l^k = \delta_k(l)$ and where δ_k is the Dirac delta function, the model is equivalent to (Lozano & Swirszcz, 2012).

Preliminary results on simulated data varying the number of features, tasks and samples show comparable results, according to mean squared error, to those obtained by the methods in (Lozano & Swirszcz, 2012; Obozinski et al., 2006). They also show more robustness in scenarios prone to overfitting as those where p is larger than n .

Proceedings of the 32^{nd} International Conference on Machine Learning, Lille, France, 2015. JMLR: W&CP volume 37.

References

Lozano, Aurelie and Swirszcz, Grzegorz. Multi-level Lasso for Sparse Multi-task Regression. *Proceedings of the 29th International Conference on Machine Learning (ICML-12)*, pp. 361–368, 2012.

Obozinski, Guillaume, Taskar, B, and Jordan, Mi. Multi-task feature selection. *Statistics Department, UC Berkeley, . . .* , pp. 1–15, 2006.