Bilevel Programming and Machine Learning

Kristin P. Bennett
Gautam Kunapuli, Jing Hu, Jong-Shi Pang
Dept. of Mathematical Sciences
Rensselaer Polytechnic Institute
Troy, NY 12180

Great progress has been made in machine learning by cleverly reducing machine learning problems to convex optimization problems with one or more hyper-parameters. But the convexity of the learning problem is dependent on defining the learning task very narrowly. Many practical aspects of learning tasks, such as data scaling, model selection, feature selection, dealing with missing values, and kernel selection, typically fall outside of the convex programming paradigm. To more fully incorporate the learning task into the optimization problem, we propose framing the learning problem as a Stackelberg game. The resulting bilevel optimization problems allow for efficient systematic search of large numbers of hyper-parameters. The upper level problem seeks to optimize generalization error by manipulating the hyper-parameters. The lower level problems optimize the model weights with the hyper-parameters fixed and return the models to the upper level problem for evaluation.

Consider the prototypical problem of training support vector machines with hyper-parameters using cross validation to select the hyper-parameters. While cross validation is a commonly employed and widely accepted method for selecting these parameters, its implementation by a grid search procedure in the parameter space effectively limits the desirable number of hyper-parameters in a model, due to the combinatorial explosion of grid points in high-dimensions. Explicitly tackling model selection by cross validation as a bilevel optimization problem allows for efficient systematic search of large numbers of hyper-parameters, raising the intriguing possibility of novel machine learning models enabled by bilevel programming. One could then expand the hyper-parameters of the models beyond the usual loss and kernel parameters to include hyper-parameters addressing other aspects of modeling such as feature selection, scaling, missing values, kernel selection, and inductive transfer [1,2] .

Bilinear programming research offers theory and algorithms for understanding and solving the challenging underlying bilinear optimization problem. The optimality conditions of bilinear programs are quite different from the usual KKT conditions. The bilinear algorithms are typically quite distinct from existing “gradient” hyper-parameter methods for k-fold cross validation such as [3]. At each iteration, the gradient methods solve k optimization problems in order to calculate the gradient of the hyper-parameters and then a gradient step is taken in the hyper-parameter space. The bilinear programming algorithms optimize both the weights and hyper-parameters simultaneously, which ultimately should lead to more efficient general purpose algorithms. We are investigating model selection of SVMs with extended sets of hyper-parameters using several algorithms for bilinear programs based on nonlinear programming relaxations, successive linear programming, branch and bound, and non-convex bundle methods [1,2,4,5].

References:

[1] K. P. Bennett, J. Hu, G. Kunapuli, and J.-S. Pang, "Model Selection via Bilevel Optimization", Proceedings of the International Joint Conference in Neural Networks, Vancouver, 2006.

[2] G. Kunapuli, K. P. Bennett, J. Hu, and J. S. Pang, “Bilevel Model Selection for Support Vector Machines,” *Data Mining and Mathematical Programming*, P. M. Pardalos and P. Hansen, CRM Proceedings and Lecture Notes, American Mathematical Society, volume : 45, 2008, in press.

[3] S. Sathiya Keerthi, Vikas Sindhwani, Olivier Chapelle, “An Efficient Method for Gradient-Based Adaptation of Hyperparameters in SVM Models,” in Advances in Neural Information Processing Systems 19, 673-680, 2007.

[4] J. Hu, J. Mitchell, J.-S. Pang, K. P. Bennett, G. Kunapuli, “On the Global solution of Linear of Linear Programs with Linear Complementarity Constraints,” SIAM Journal on Optimization, 2008, to appear.

[5] G. Kunapuli, K. Bennett, J. Hu, and J.-S. Pang, “Classification Model Selection via Bilevel Programming,” Computational Optimization and Applications, 2008, to appear.

This work was supported by the Office of Naval Research Grant no: N00014-06-1-0014.

Topic: learning algorithms
Preference: oral