ARTICLE SUMMARY ON
"Optimization problems for machine learning"

MOHAMMAD HEYDARI
mohammadheydari.edu@gmail.com

This paper surveys the machine learning literature and presents in an optimization framework several commonly used machine learning approaches. Particularly, mathematical optimization models are presented for regression, classification, clustering, deep learning, and adversarial learning, as well as new emerging applications in machine teaching, empirical model learning, and Bayesian network structure learning. Such models can benefit from the advancement of numerical optimization techniques which have already played a distinctive role in several machine learning settings. The strengths and the shortcomings of these models are discussed and potential research directions and open problems are highlighted.

The pursuit to create intelligent machines that can match and potentially rival humans in reasoning and making intelligent decisions goes back to at least the early days of the development of digital computing in the late 1950s (Solomonoff, 1957). The goal is to enable machines to perform cognitive functions by learning from past experiences and then solving complex problems under conditions that are varying from past observations. Fueled by the exponential growth in computing power and data collection coupled with the widespread of practical applications, machine learning is nowadays a field of strategic importance.

Mathematical programming constitutes a fundamental aspect of many machine learning models where the training of these models is a large scale optimization problem. This paper surveyed a wide range of machine learning models namely regression, classification, clustering, and deep learning as well as the new emerging paradigms of machine teaching and empirical model learning. The important mathematical optimization models for expressing these machine learning models are presented and discussed. Exploiting the large scale optimization formulations and devising model specific solution approaches is an important line of research particularly benefiting from the maturity of commercial optimization software to solve the problems to optimality or to devise effective heuristics. However, as highlighted in Liang, Poggio, Rakhlin, and Stokes (2019) and Poggio et al. (2017), providing quantitative performance bounds remains an open problem. The nonlinearity of the models, the associated uncertainty of the data, as well as the scale of the problems represent some of the very important and compelling challenges to the mathematical optimization community. Furthermore, bilevel formulations play a big role in adversarial learning (Hamm & Noh, 2018), including adversarial training, data poisoning and neural network robustness.
Based on this survey, we summarize the distinctive features and the potential open machine learning problems that may benefit from the advances in computational optimization.

• **Regression**. The typical approaches to avoid overfitting and to handle uncertainty in the data include shrinkage methods and dimension reduction. These approaches can all be posed as mathematical programming models. General non-convex regularization to enforce sparsity without incurring shrinkage and bias (such as in lasso and ridge regularization) remain computationally challenging to solve to optimality. Investigating tighter relaxations and exact solution approaches continue to be an active line of research (Atamturk & Gomez, 2019).

• Classification. Classification problems can also be naturally formulated as optimization problems. Support vector machines in particular have been well studied in the optimization literature. Similar to regression, classifier sparsity is one important approach to avoid overfitting. Additionally, exploiting the kernel tricks is key as nonlinear separators are obtained without additional complexity. However, when posed as an optimization problem, it is still unclear how to exploit kernel tricks in sparse SVM optimization models. Another advantage to express machine learning problems as optimization problems and in particular classification problems is to account for inaccuracies in the data. Handling data uncertainty is a deeply explored field in the optimization literature and several practical approaches have been presented to handle uncertainty through robust and stochastic optimization. Such advances in the optimization literature are currently being investigated to improve over the standard approaches (Bertsimas et al., 2019).

- **Clustering.** Clustering problems are in general formulated as MINLPs that are hard to solve to optimality. The challenges include handling the non-convexity as well as the large scale instances which is a challenge even for linear variants such as the capacitated centered clustering (formulated as a binary linear model). Especially for large-scale instances, heuristics are typically devised. Exact approaches for clustering received less attention in the literature.

- **DNNs architectures as MIPs.** The advantage of mathematical programming approaches to model DNNs has only been showcased for relatively small size data sets due to the scale of the underlying optimization model. Furthermore, expressing misclassification conditions for adversarial examples in a nonrestrictive manner, and handling the uncertainty in the training data are open problems in this context.

- **Adversarial** learning and adversarial robustness. Optimization models for the search for adversarial examples are important to identify and subsequently protect against novel sets of attacks. The complexity of the mathematical models in this context is highly dependent on the classifier function. Untargeted attacks received less attention in the literature, and the mathematical programming formulation (110)– (114) has been introduced in Section 7.2. Furthermore, designing models robust to adversarial attacks is a two-player game, which can be cast as a bi-level optimization problem. The loss function adopted by the learner is one main complexity for the resulting mathematical model and solution approaches remain to be investigated.

- **Data poisoning:** Similar to adversarial robustness, defending against the poisoning of the training data is a two-player game. The case of online data retrieval is especially challenging for gradient-based algorithms as the KKT conditions do not hold.

- **Activation ensembles.** Activation ensembles seek a trade-off between the classifier accuracy and computational feasibility of training with a mathematical programming approach. Adopting activation ensembles to train large DNNs have not been investigated yet.

- **Machine teaching.** Posed as a bi-level optimization problem, one of the challenges in machine teaching is to devise computationally tractable single-level formulations that model the learner, the teaching risk, and the teaching cost. Machine teaching also generalizes a number of two-player games that are important in practice including data poisoning and adversarial training.

- **Empirical model learning.** This emerging paradigm can be seen as the bridge combining machine learning for parameter estimation and operations research for optimization. As such, theoretical and practical challenges remain to be investigated to propose prescriptive analytics models jointly combining learning and optimization in practical applications.

Mohammad Heydari
810197494
Jan 2022

Reference:

1. Claudio Gambella, Bissan Ghaddar, Joe Naoum-Sawaya. “Optimization problems in machine learning: A survey”. Available: **https://en.x-mol.com/paper/article/1346547675631341568**