Bayesian optimisation of approximateness in the trade-off between statistical and computational efficiency

An MPhil project proposal

J. Soendermann (*jjes2*), Selwyn College

Project Supervisor: Prof. Zoubin Ghahramani

Abstract

Given a fixed amount of time and the choice between running a computationally complex learning algorithm on a limited amount of data and running a simpler or approximate algorithm on a larger amount of data, the latter can sometimes be the better choice. More generally, there is a trade-off between statistical and computational efficiency. Finding the balance between the two that produces optimal results is currently done by human machine learning experts. My project aims to automate this process.

1 Introduction, approach and outcomes

Before the advent of computers, statistical methods had to be simple enough to be calculated by humans. With the introduction of computers to the field of statistics and the steady increase in their computational power, many new statistical methods have been devised that weren't feasible before. At the same time, statisticians have developed new, potentially useful methods that are computationally intractable.

Currently, using these methods requires an expert in machine learning to make high level decisions about using approximate inference methods or simpler methods that can process more data than more complex models could process given the same amount of time. These decisions are often made ad hoc and it would be very desirable to automate this process.

Specifically, a trade-off between computational and statistical efficiency has to be made. Reducing the complexity of the statistical method being used with the purpose of using more data is an idea that has not traditionally been considered in statistics. Recent research that does try to address this problem, such as [1], tries to devise complex proofs about characteristics of approximate versions of existing algorithms. In this project, however, I plan to approach the problem of finding optimal trade-offs from a more empirical, computer science based perspective.

Classical optimisation methods may not be suitable to solve this problem because gathering data can take a very long time – training an entire deep neural net may only be one data point. One promising way to approach this problem is Bayesian optimisation [3], as Bayesian optimisation methods have been shown to are very efficient in their data usage. In [2], the authors describe a way of using Bayesian optimisation for a related problem, namely selecting both a machine learning algorithm and a set of hyperparameters for it from the space of algorithms and hyperparameters. They also give a list of standard datasets commonly used for research in the field which I intend to draw upon to find suitable data to evaluate my own system. The paper furthermore contains a list of classification algorithms that could be used as a starting point to find a minimal subset of algorithms with which to evaluate the idea.

I plan build a system that can automatically choose among a set of different machine learning algorithms and approximations. While determining a suitable candidate among this set of algorithms, it will run approximate versions of the algorithms, running less approximate versions only for promising candidates. Building this system will involve theoretical work, potentially in the form of a paper. It will also require me to implement my ideas to test their validity.

2 Workplan

8 December 2014	Start date
late December 2014	Identify a minimally interesting set of machine learning algorithms for classification
January 2015	Identify generic methods for performing approximate inference for these algorithms
late January 2015	Investigate how performance varies as different approximation parameters are varied
February 2015	Develop a method that can predict the variation in performance as approximation parameters are varied
March 2015	Create a system that uses this method in building classification models
April 2015	Demonstrate that this results in improved accuracy and superior computation time trade-offs compared to relevant alternatives
May 2015	Write report and test system
12 June 2015	Deadline

References

[1] Michael I. Jordan, *On statistics, computation and scalability*. Bernoulli 2013, Vol. 19, No. 4, 1378-1390

[2] Chris Thornton, Frank Hutter, Holger H. Hoos, Kevin Leyton-Brown, *Auto-WEKA: Combined Selection and Hyperparameter Optimization of Classification Algorithms*. Proc. of KDD 2013, 2013

[3] Jasper Snoek, Hugo Larochelle, Ryan P. Adams, *Practical Bayesian Optimization of Machine Learning Algorithms*. Advances in Neural Information Processing Systems, 2012