
Ensembles for Discovery of Compact Structures and Learning Back-propagation Forests*

Madalina Fiterau

Abstract

In many practical scenarios, complex high-dimensional data contains low-dimensional structures that could be informative of the analytic problems at hand. I will present a method that detects such structures if they exist, and uses them to construct compact interpretable models for different machine learning tasks that can benefit practical applications.

To start with, I will formalize Informative Projection Recovery, the problem of extracting a small set of low-dimensional projections of data that jointly support an accurate model for a given learning task. Our solution to this problem is a regression-based algorithm that identifies informative projections by optimizing over a matrix of point-wise loss estimators. It generalizes to multiple types of machine learning problems, offering solutions to classification, clustering, regression, and active learning tasks. Experiments show that our method can discover and leverage low-dimensional structures in data, yielding accurate and compact models. Our method is particularly useful in applications in which expert assessment of the results is of the essence, such as classification tasks in the healthcare domain.

In the second part of the talk, I will describe back-propagation forests, a new type of ensemble that achieves improved accuracy over existing ensemble classifiers such as random forests classifiers or alternating decision forests. This research was performed under the mentorship of Dr. Peter Kotschieder and in collaboration with Dr. Samuel Rota-Bul (FBK Trento, IT). Back-propagation (BP) trees use soft splits, such that a sample is probabilistically assigned to all the leaves. Also, the leaves assign a distribution across the labels. The splitting parameters are obtained through SGD by optimizing the log loss over the entire tree, which is a non-convex objective. The probability distribution over the leaves is computed exactly by maximizing a log concave procedure. In addition, I will present several proposed approaches for the use of BP forests within the context of compact informative structure discovery.

*Machine Learning External Seminar, Gatsby Unit, March 9, 2015.