Learning from Features of Sets and Probabilities*

Zoltán Szabó (Gatsby Unit, University College London)

Abstract

Many modern learning problems make inferences based on probabilities as inputs. For example, Microsoft's TrueSkill software updates a probabilistic estimate of player skill, based on input probability distributions representing the (uncertain) skill of rival players and the current player skill (similar models are developed for social recommender systems and search advertising). The probabilistic foundation of these models, known as Expectation Propagation, requires taking probabilities as inputs, and providing parameters (for instance a belief on player skill) as output. These computations can be obtained in closed form only for a narrow class of parametric models, however. We describe a more general framework, using an efficient "bag of Fourier features" representation of the input probabilities, which allows us to develop expectation propagation schemes for a much wider range of models. Our theoretical contribution is to determine how many Fourier features are needed for accurate learning, yielding a bound exponentially tighter in domain size than previous results.

A related learning setting arises when only samples from probability distributions are available as inputs: in our work, these have been sets of multi-spectral satellite measurements taken at different locations of a region, from which we want to predict the aerosol value. We represent such inputs as "bags of samples", on which we define kernel measures of similarity for the purpose of learning (these kernels are dot products between potentially infinitely many features of the sample bags). We show how many samples per bag to use, so as to give the same asymptotic performance as we would obtain given the true underlying distributions generating the samples, and the best performing algorithm.

Preprints:

1. Learning on bags-of-features: `http://arxiv.org/abs/1411.2066`.
2. Improved bounds on random Fourier features: `http://www.gatsby.ucl.ac.uk/~szabo/publications/sriperumbudurszabo15optimal.pdf`.

*Imperial College London, Department of Computing, UK, 9 March 2016; abstract.