

# Survey of Boosting from an Optimization Perspective

ICML 2009 Tutorial

Manfred K. Warmuth and S.V.N. Vishwanathan

[\[tutorial description\]](#) [\[presenters\]](#) [\[outline\]](#) [\[slides\]](#) [\[code\]](#)

## Tutorial description

Boosting has become a well known ensemble method. The algorithm maintains a distribution on the  $\pm$ -labeled examples and a new base learner is added in a greedy fashion. The goal is to obtain a small linear combination of base learners that clearly separates the examples. We focus on a recent view of Boosting where the update algorithm for distribution on the examples is characterized by a minimization problem that uses a relative entropy as a regularization.

The most well known boosting algorithm is AdaBoost. This algorithm approximately maximizes the hard margin, when the data is separable. We focus on recent algorithms that provably maximize the soft margin when the data is noisy. We will teach the new algorithms, give a unified and versatile view of Boosting in terms of relative entropy regularization, and show how to solve large scale problems based on state of the art optimization techniques.

Our goal is to motivate people to mimic the recent successes of the SVM community for scaling up the solvable problem size. This goal is challenging because in Boosting the regularization (relative entropy) is more complicated than the one used for SVMs (squared Euclidean distance). Nevertheless we can solve dense problems with 200K examples in less than a minute on a laptop.

## Presenters

**Manfred K. Warmuth**, Professor of Computer Science, University of California Santa Cruz, [manfred@cse.ucsc.edu](mailto:manfred@cse.ucsc.edu), <http://www.cse.ucsc.edu/~manfred/>.

Manfred's background: Theoretical machine learning, PAC learning, on-line learning, compression schemes, generalization bounds, developed the Exponentiated Gradient and the Weighted Majority algorithms, boosting, relative entropy regularization, Bregman divergences, matrix updates based on the quantum relative entropy regularization.

**S.V.N. Vishwanathan**, Assistant Professor of Statistics and Computer Science, Purdue University, [vishy@stat.purdue.edu](mailto:vishy@stat.purdue.edu), <http://www.stat.purdue.edu/~vishy>.

Vishy's background: kernel methods, non-smooth optimization, structured prediction.

## Outline

Part one (by Manfred):

- We will start with LPBoost, a simple Boosting algorithm that maximizes the soft margin based on Linear Programming.
- We discuss simple artificial examples for which LPBoost converges slowly to the maximum margin. If a relative entropy regularization is added then  $O(\frac{\log n}{\epsilon^2})$  iteration bound can be proven, where  $n$  is the number of examples and  $\epsilon$  a precision parameter.
- We survey most recent Boosting algorithms culminating in algorithms that explicitly maximize the one norm margin of the produced linear combination of base learners. Starting with AdaBoost, we will characterize each algorithm by the underlying primal/dual optimization problems they aim to solve.
- Stopping criteria.
- High level discussion of methods for proving the  $O(\frac{\log n}{\epsilon^2})$  iteration bounds and generalization error bounds.
- Weakness of the  $O(\frac{\log n}{\epsilon^2})$  iteration bound.

In part two, Vishy will discuss up to date optimization techniques for solving large instances of the optimization problems.

- Introduce relevant topics from convex analysis and numerical optimization.
- Discuss the high level difference between a corrective and a totally corrective algorithm.
- Sketch various strategies for solving both variants.
- In particular, talk about two approaches for solving the totally corrective variant that work in the primal or dual domain.
- Revisit the gap between practice and theory (the looseness of the bounds, how practical algorithms behave, and why the bounds we can prove are rather weak).
- Show on natural data that LPBoost can converge slowly and that adding a relative entropy regularization speeds up convergence.
- Show on natural data that the margin can be a weak proxy for generalization.

## Final slides

The slides for this tutorial can be found at

[\[Part I: Entropy Regularized LPBoost - by Manfred Warmuth\]](#)  
[\[Part II: An Optimization View of Boosting - by Vishy Vishwanathan\]](#)

## Code

An efficient C++ implementation of various state of the art boosting algorithms can be found [\[here\]](#). All plots in the tutorial were generated using this code. Comments and feedback welcome!

## Acknowledgments

Karen Glocer provided help with the experiments.