Review on boosting methods

Vu Tuan Hung and DO Quoc Khanh

Télécom ParisTech

May 11, 2012

Table of contents

1. Two-class problem
2. Multi-class problem
3. Experiments

Common model

- The goal: give a common model that can explain boosting algorithms, and that can help to construct other similar methods.
- View boosting as iterative algorithms of optimization in numerical/function space.
- Loss function $L(y, F)$.
- Choose d_m direction of descent (gradient descent, steppest descent, Newton etc.)
- Look for $f_m \approx d_m$.
- $\mathcal{F} = \sum_{m=1}^{M} \mathcal{Q}_m$.

Summary

	$L(y, F)$	d_m/β_m	pop-ver
Discrete AB	e^{-yF}	steppest descent	No
Real AB	e^{-yF}	direct optim	Yes
Gentle AB	e^{-yF}	Newton step	Yes
LS Boost	$(y-F)^2$	steppest descent	No
LAD TB	$\lvert y-F \rvert$	grad/tree	No
M TB	Huber	grad/tree	No
Logit B	$\log(1+e^{2F})-2y^*F$	Newton step	Yes
L2 TB	$\log(1+e^{-2yF})$	grad/tree	No
Lk TB	- Binomial Likelihood	grad/tree	No

Table 1: Summary on boosting algorithms.

Traditional approach

- J-class problem.
- Transformation of y into $N \times J$-binary matrix.
- Divise the problem into J binary problems.
- Two-class algorithms $\rightarrow F_j, j = 1 : J$.
- Output for each $\mathbf{x}$: $arg \max\limits_{j=1:J} F_j(\mathbf{x})$

Traditional approach

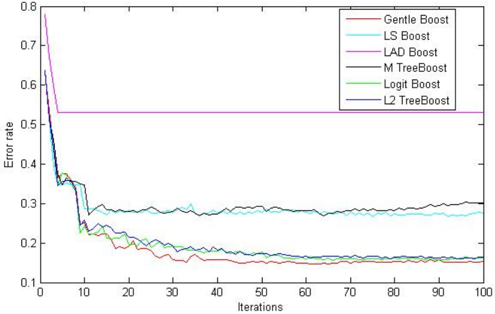

Figure 1: Comparison of the performances of *AdaBoost.MH* using 6 two-class boosting methods, experiments on simulated data with 0% Bayes error.

Other generalizations

- Logit Boost for J classes and Lk TreeBoost.
- Can use our model for derivation.

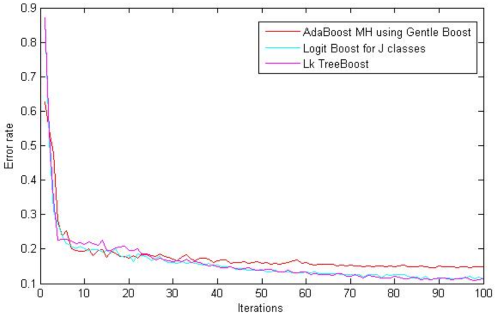

Figure 2: Multi-class classification on simulated data, $J = 4$, Bayes error 0%.

Questions

?