The Chinese University of Hong Kong
Department of Statistics

Seminar

DELTA BOOSTING MACHINE

By

MR. Simon Lee
Chief Data Scientist at AXA Hong Kong

Abstract

We introduce Delta Boosting(DB) as a new member of the boosting family. Similar to the popular Gradient Boosting (GB), this new member is presented as a forward stagewise additive model that attempts to reduce the loss at each iteration by sequentially fitting a simple base learner to complement the running predictions. Instead of relying on the negative gradient, as is the case for GB, DB adopts a new measure called delta as the basis. Delta is defined as the loss minimizer at an observation level. We also show that DB is the optimal boosting member for a wide range of loss functions. The optimality is a consequence of DB solving for the split and adjustment simultaneously to maximize loss reduction at each iteration. In addition, we introduce an asymptotic version of DB that works well for all twice differentiable strictly convex loss functions. This asymptotic behavior does not depend on the number of observations, but rather on a high number of iterations which can be augmented through common regularization techniques. We show that the basis in the asymptotic extension only differs from the basis in GB by a multiple of the second derivative of the log-likelihood. The multiple is considered to be a correction factor, one that corrects the bias towards the observations with high second derivatives in GB. When negative log-likelihood is used as the loss function, this correction can be interpreted as a credibility adjustment for the process variance. Simulation studies and real data application we conducted suggest that DB is a significant improvement over GB. The performance of the asymptotic version is less dramatic, but the improvement is still compelling. Like GB, DB provides a high transparency to users, and we can review the marginal influence of variables through relative importance charts and the partial dependence plots. We can also assess the overall model performance through evaluating the losses, lifts and double lifts on the holdout sample.

Date:	May 10, 2018 (Thursday)
Time:	3:00 p.m. - 4:00 p.m.
Venue:	Lady Shaw Building, Room LT2
	The Chinese University of Hong Kong