



The Lynne and William Frankel Center
for Computer Science

Department of Computer Science
Ben Gurion University of the Negev

Tel: 08-6428032

fradmin@cs.bgu.ac.il



Distinguished Lecturer Series

Departments of Computer Science and Electrical & Computer Engineering



Supported by Jeffrey & Holly Ullman

Abraham J. Wyner

Professor

The Wharton School, The University of Pennsylvania

Explaining the Success of AdaBoost and Random Forests as Interpolating Classifiers

Abstract: There is a large literature explaining why AdaBoost is a successful classifier. The literature on AdaBoost focuses on classifier margins and boosting's interpretation as the optimization of an exponential likelihood function. These existing explanations, however, have been pointed out to be incomplete. A random forest is another popular ensemble method for which there is substantially less explanation in the literature. We introduce a novel perspective on AdaBoost and random forests that proposes that the two algorithms work for essentially similar reasons. While both classifiers achieve similar predictive accuracy, random forests cannot be conceived as a direct optimization procedure. Rather, random forests is a self-averaging, interpolating algorithm which fits training data without error but is nevertheless somewhat smooth. We show that AdaBoost has the same property. We conjecture that both AdaBoost and random forests succeed because of this mechanism. We provide a number of examples and some theoretical justification to support this explanation. In the process, we question the conventional wisdom that suggests that boosting algorithms for classification require regularization or early stopping and should be limited to low complexity classes of learners, such as decision stumps. We conclude that boosting should be used like random forests: with large decision trees and without direct regularization or early stopping.

Bio: Abraham (Adi) Wyner is Professor and Chair of the Undergrad Program in Statistics at the University of Pennsylvania's Wharton School. Before arriving at University of Pennsylvania in 1999, he was Assistant Professor of Statistics at University of California, Berkeley. His research is in machine learning, discrete time series, Information Theory, and the application of Statistics to Environmental Sciences, Neuroscience, Information Theory and Sports.

12:00 on Tuesday December 12, 2017 — Saal Auditorium (202), Alon Building for Hi-Tech (37)
12:00 יום ג' דצמבר 2017 – באודיטוריום סאל (202), בניין אלון לטכנולוגיה עילית (37)