



THE UNIVERSITY OF
CHICAGO

Department of Statistics
DISSERTATION PROPOSAL

TAKINTAYO AKINBIYI

Department of Statistics
The University of Chicago

**Incorporating Localization Into PAC-Bayes
Generalization Bounds**

TUESDAY, April 23, 2013, at 1:30 PM
110 Eckhart Hall, 5734 S. University Avenue

ABSTRACT

We consider the situation where one has an infinite set of classifiers and needs to select one which has good average performance relative to an unknown data distribution. Each classifier inputs a data point and produces a real valued output. We suppose there is a loss function which for any data point and classifier produces a bounded real valued loss. Furthermore, we have a sample of possibly dependent data points which are drawn from the unknown data distribution. The PAC-Bayes framework suggests algorithms which input the sample and output a “posterior” distribution on classifiers from which a classifier is randomly selected. These algorithms choose a posterior which minimizes an estimated upper bound on the expected loss when a classifier is drawn from the posterior and applied to a fresh data from the unknown data distribution. Various PAC-Bayes theorems provide such upper bounds which are correct with high probability. For each posterior, the upper bound is typically a weighted sum of the average loss when a classifier is drawn randomly from the posterior and a measure of the “complexity” of the posterior defined as the KL divergence of the posterior from a reference distribution chosen before observing the training sample. Following the ideas of Olivier Catoni we improve these upper bounds by (1) providing an empirical upper bound on the KL divergence when the posterior and reference distribution are chosen to have a particular form and (2) incorporating the sample second moment of the loss for each posterior. The use of reference distributions of this particular form is called localization. Furthermore, while many PAC-Bayes theorems exist for the case when the training data are i.i.d. we attempt to provide results applicable in the dependent case.