

**ACTIVE LEARNING OF GAUSSIAN MIXTURE
MODELS USING DIRECT ESTIMATION OF ERROR
REDUCTION**

By

Jeffry Gaston

An Abstract of a Thesis Submitted to the Graduate Faculty
of Rensselaer Polytechnic Institute
in Partial Fulfillment of the
Requirements for the Degree of
MASTER OF SCIENCE

Major Subject: COMPUTER SCIENCE

The original of the complete thesis is on file
in the Rensselaer Polytechnic Institute Library

Examining Committee:

Sanmay Das, Thesis Adviser
Malik Magdon-Ismail, Member
Mohammed Zaki, Member

Rensselaer Polytechnic Institute
Troy, New York

April 2012
(For Graduation May 2012)

Traditional supervised learning for training a classifier involves a static set of fully-labeled data. However, labels can be costly to obtain. Active learning can provide lower misclassification rates for an equivalent number of labeled datapoints by allowing the learner to choose particularly important points for labeling. Standard paradigms for choosing points to label may be to request points near the decision boundary, or points where uncertainty is maximum. We show how to directly optimize the target criterion by estimating error reduction directly in a probabilistic framework (the Gaussian Mixture Model) and show that in several cases this can reduce the need for as many labeled training examples as would be required by the maximum-uncertainty approach.