

Abstract :UCSD ITA Workshop 2/9/06

## THE IMPORTANCE OF REGULARIZATION

Regularization in statistics comes from the corresponding term in applied mathematics, where smooth approximate solutions to ill conditioned integral equations were obtained by introducing a smoothness penalty (Tikhonov(1963)). In statistics the issue arose with avoiding overfitting in high dimensional parametric or nonparametric regression. Methods which have been developed include smoothness or more generally complexity penalties, which can be interpreted as replacing maximum likelihood procedures by Bayes posterior modes, cross validation, and Monte Carlo linked procedures such as the bootstrap. In machine learning, regularization has appeared, for instance, in the guise of slack variables in SVM and early stopping in boosting.

We will discuss three examples which illustrate the importance of regularization:

1. Boosting
2. Estimating a high dimensional covariance matrix for classification or other purposes
3. Setting confidence bounds on extreme percentiles of a distribution, using the  $m$  out of  $n$  bootstrap.