# 290C - Advanced Machine Learning
# On-line Learning
# UC Santa Cruz

Manfred K. Warmuth

April 2, 2007

The class will cover the application and analysis of on-line learning algorithms. Application areas include: on-line tuning of operating system parameters, adaptive query processing for databases, on-line tracking for vision applications, on-line ranking of webpages, ...

I welcome auditors and visitors from other universities in the Bay area. However you need to come to all the classes to make it worth your while, since the classes will build on each other.

**Classtime:** T,Th 4-5:45 Earth and Marine Sciences, B210

**First class:** Tu April 3rd

**Last class:** Th June 7th

**Class web page:** http://www.soe.ucsc.edu/classes/cmps290c/Spring07/

In our approach, we typically make no statistical assumptions about how the data is generated. Instead we prove *regret* bounds that hold for arbitrary sequences of data points. Here regret is the total loss of the on-line algorithm minus the total loss of the best model chosen in hindsight after all the data is known.

An on-line algorithm must be designed designed for the case when the data is changing over time. It must constantly hedge its bets in various ways: Should it choose a model that does best on the recent data or stick to a model that worked well in the past. The algorithm must keep some uncertainty information about which model it should choose akin to a Bayesian posterior.

We begin with the simplest algorithms and analysis techniques and slowly work ourselves up to the fancier more recent work that makes use of some Quantum Physics machinery.

We will cover a large number of model or state spaces and how to keep uncertainty information over them. Classical Bayesian analysis will be one special case but we will learn how to keep uncertainty information over sets of a fixed size, permutations, paths in graphs, and directions. For the last case, the uncertainty information is summarized as a density matrix we use techniques akin to Quantum Physics and Quantum Information Theory in our analysis. You will learn lots of new linear algebra, information theory, Bregman projection methods and optimization.

There will be a number of theoretical and practical homework assignments and a project. I will give a large number of open problems in this class. Hopefully some of the projects will lead to research papers.

Here are some topics that will be covered:

- Halving algorithm, Weighted Majority algorithm, Expert framework. Analysis using a potential function or the relative entropy
- Shifting experts, long-term memory, measuring the on-lineness of the data
- Applications to the Disk Spindown Problem, Caching, ...
- Motivation of the relative entropy. Bregman divergences and the use of the Generalized Pythagorean Theorem
- How does Bayesian analysis fit into the expert framework
- Predicting the stock market
  comparison of various on-line algorithms including the Antecore algorithm
- Learning disjunctions - the Winnow algorithm
- Additive and multiplicative updates for linear regression
- Derivations of the updates
- Use of Bregman divergences as parameter divergence as well as matching loss functions

- The Blessing and the Curse of the Multiplicative updates
- Playing games with multiplicative updates.
  Boosting and its analysis using relative entropies
- Learning as well as be best set of experts
  On-line Principal Component Analysis
- Keeping uncertainty information about directions with density matrices
  The probability calculus for density matrices
- Support vector machines and kernel methods for the additive family
- Kernel methods for the multiplicative family
- Winnowing subspaces
- PAC-Bayes bounds
- Minimax and last-step minimax