

Some CMPS 242 Project Ideas - on the research side

1. boosting with noise, in e.g. micro-array data. Add noise to "spread-out" examples getting something closer to maximum margin classification
2. Learn the learning rate: run WM or experts algorithm with different β values at the same time. The problem is that the different weights are incomparable, after several mistakes the low-beta values are dominated by the high-beta ones. The solution is to normalize by beta so that this imbalance is corrected. Consider the weight each time the master alg makes a mistake – it goes down to $1/2 + \beta/2$ of the previous weight. Thus it might make sense to re-normalize by this factor. In the continuous case, the weight of an expert with loss ℓ after the master alg has made m mistakes would be

$$\int_{\beta} \frac{\beta^{\ell}}{(1/2 + \beta/2)^m}$$

Unfortunately, I don't see a closed form for these. (Look at Cover portfolio paper?)

3. learning weights to more closely predict expert moves in Go
4. Find relationships between Boltzmann distribution and on-line algs? relation between on-line algs and Narendra & Thathachar, 1989? (reinf learn tutorial) (This one is a little speculative)
5. Semantic similarity between words - use LSA-type methods/clustering to get a word hierarchy. See refs in Zigoris W2005 ML report
6. Recursively use AdaBoost for segmentation: classify based on features and then on derived features which are the (previous iteration's) classifications of nearby points.
7. Experiment with an annealed version of AdaBoost/LogitBoost where alphas decrease with iteration (like AveBoost).
8. Needle-in-a-haystack boosting – consider delta-examples (good minus each bad and visa-versa) how to weight the delta examples?
9. Compare iterative vs. LP all-at-once boosting (see Manfred's papers)
10. Evaluate effectiveness of PCA - how many components to choose and how effective is it?
11. Finding the k in k-clustering - survey and evaluate the various approaches.
12. Adding a confidence parameter to each example to set its margin goal in SVM classification (i.e. minimizing $\|w\|_2$ such that each $y_i(w\dot{x} + b) \geq c_i$).