

## Ideas for 2002 Projects

### 1. Hypothesis Testing with Small Sets (Dradulov...)

Given past observations from a distribution predict whether a new sample comes from the distribution. Extension KS-statistic to multidimensional distributions.

### 2. Connection between MED and Regularization

The following paper “Maximum entropy discrimination” by T. Jaakkola, M. Meila, and T. Jebara introduces a way of using generative models in classification and regression problems by using entropy rather than a RKHS norm as the regularization functional.

### 3. Feature Selection for SVMs: Theory and Experiments

Often only a small number of a large set of input features is relevant in a classification task. Algorithms exist for selecting features when the classifier is an SVM. Possible projects consist of developing a better theoretical understanding of some properties of these algorithms. “Choosing Multiple Parameters for Support Vector Machines” by O. Chapelle, V. Vapnik, O. Bousquet, and S. Mukherjee.

#### 4. Bayes Classification Rule and SVMs

For a two class problems one if the class conditional probabilities are know then one can use the following rule to classify

$$y = \text{sign} \left( \ln \frac{P(c = 1|x)}{P(c = -1|x)} - \ln \frac{P(c = -1)}{P(c = 1)} \right).$$

When classes have equal probabilities the SVM can be placed in such a framework. What happens when the classes are not equiprobable ?

One paper looked at will be Lin, Y., Lee, Y., and Wahba, G. " Support Vector Machines for Classification in Nonstandard Situations " TR 1016, March 2000.

5. IOHMMs: Evaluation of HMMs for classification vs direct classification

HMMs and generative models offer an alternative to classifiers like SVM. A comparison of IOHMMs vs. SVMs. The relevant paper is Y. Bengio and P. Frasconi, "Input/Output HMMs for Sequence Processing," IEEE Transactions on Neural Networks, vol. 7, no. 5, pp. 1231–1249, 1996.

6. Reusing the Test Set: Data Mining Bounds (2 projects:  
1. Littlestone, 2. Evgeniou)

How much does one pay when one reuses data (Evgeniou) ? Get VC or confidence bounds on deviation between empirical and expected when data is reused.

One can get generalization bounds from on-line analysis in a very interesting way. For classification these bounds are the tightest around. Can this be extended and how general is this technique ? The relevant paper is N. Littlestone. From on-line to batch learning. In Proceedings of the Second Annual Workshop on Computational Learning Theory, pages 269–284, San Mateo, CA, 1989.

7. Large-Scale Nonlinear Least Square Regularization [Gene Y.]

8. View-based classification
  
9. Local vs Global classifiers: experiments and theory (conjecture: unlike Vapnik's statement are local and global subsumed under the same formulation?)  
  
Vapnik talks about local and global structures. The relevant paper is Local Learning Algorithms (1992) Leon Bottou, Vladimir Vapnik, Neural Computation.
  
10. RKHS invariance to measure: historical math
  
11. Concentration experiments (dot product vs. square distance)

12. Decorrelating classifiers (Niyogi): experiments about generalization using a tree of stumps

The paper to look at is P. Niyogi, J. B. Pierrot, and O. Siohan, "Multiple Classifiers by Constrained Minimization," to appear, Proceedings of International Conference on Acoustics, Speech, and Signal Processing, Istanbul, Turkey, June, 2000.

13. Kernel synthesis and selection

The papers to look at are Poggio, T. and F. Girosi. A Sparse Representation for Function Approximation, Neural Computation, Vol. 10, No. 6, 1445-1454, 1998 and On Optimizing Kernel Alignment (Nello Cristianini, Jaz Kandola, Andre Elisseeff, John Shawe-Taylor).

14. Bayesian Interpretation of regularization and in particular of SVMs: limits theorem and integral approximations and Gaussian Processes.
15. Philosophy project: history of induction from Kant to Popper and current state
16. “Religious” project: Bayesian Priorhood