Projects

2008

Projects

1. Exploit fast algoritm for NN (Indyk's algorithms, improved fast Gauss transforms,...) [RIF]
2. Implement and test a "large-scale" nonlinear RLS algorithm, using and expanding the ideas discussed in class. [RIF]
3. Preconditioning RLS using clustering algorithms [RIF, ROSS].
4. Why Reproducing Kernel Hilbert Spaces are a natural set of hypothesis spaces for supervised learning. Draw ideas from embedding theorems, extension to Banach spaces having in mind sparsity based regularization.[TP, L]
5. Discuss ideas for algorithms based on maximizing stability of the algorithm at the predicted point and minimizing empirical error. [TP]
6. Comparison of Ivanov and Tikhonov regularization: stability and approximation properties. [TP, L, SR]

Projects

1. Boosting and compressed sensing. Connection between ensemble methods and sparsity based regularization. [L]
2. Dimensionality reduction. Find a general way to pose the problem in view of a comparison of existing techniques such as Laplacian-eigenmaps, PCA, LDA, IR, ICA. How is overfitting avoided?[L]
3. Computational comparison of algorithms for sparsity based regularization. [L]
4. Theoretical and empirical comparison between sparsity based regularizartion and RLS.[L]
5. Stability of sparsity based regularization.[TP, L]

Projects

1. (Non-linear) Regularized least-squares with noise on the data. Unifying review of (kernel) Partial Least-Squares/Total Least-Squares literature. Connections to Gaussian Processes formulation of problems involving noisy data and noisy observations/labels. [JB]
2. Intelligently choosing small datasets that are "representative" of larger datasets in the interest of working with a computationally more tractible problem. Analysis of Myrvoll's greedy functional gradient technique for selecting data from the full training set. (Speech, finance, vision applications) [JB]
3. Cross-validation with sequential (and non-iid) data (e.g. time series): does it make sense to include predictions on the past given data from the future in a LOOE calculation? When can you, in some sense, get away with it? When would it be "bad"? [JB]

Other projects (review)

1. Random Projections.
2. Learning the kernel: recent developments.
3. How to deal with unbalanced training sets.
4. Learning from non i.i.d. data.
5. Regularization parameter choice.
6. Graph regularization.
7. Learning Invariances.

Computational Neuroscience-type projects

1. Various projects involving the CBCL visual cortex model are available. These projects will typically require more time but could also lead to a publication [Thomas Serre].
2. Derived distance applied to text, with domain-specific transformations.
3. Parametric study of derived distance (layers, template selection, ...).