Machine Learning Group at RSISE

THE AUSTRALIAN NATIONAL UNIVERSITY

Topics

- Kernel Methods (Support Vectors)
- Boosting
- Learning Theory
- Parallel Computing and Data Mining
- Logic and Inference

People

Jyrki Kivinen (Finland), John Lloyd (Australia), Shahar Mendelson (Israel), Gunnar Rätsch (Germany), Alex Smola (Germany), Bob Williamson (Australia)

Collaborations

Microsoft Research (Redmond and Cambridge), Telstra (Melbourne), University of Wisconsin, MPI for Biological Cybernetics (Tübingen), Technion (Haifa), Fraunhofer (Berlin)

Kernel Methods

THE AUSTRALIAN NATIONAL UNIVERSITY

New Method of Learning

- Can solve problems exactly (Neural Networks have many local minima).
- Easy to use, very flexible to incorporate prior knowledge
- Works well
- We can prove things

Applications

- Handwritten Digit recognition (OCR programs)
- Computer Security (Network intrusion detection)
- Drug Discovery (finding good candidates, classifying molecules, etc.)
- Computer Vision (face detection)

Boosting

THE AUSTRALIAN NATIONAL UNIVERSITY

Basic Idea

- Lots of idiots get it right on average ...
- More precisely: combine many weak learning methods to one strong one
- Works for almost any weak learning algorithm

Applications

- Junk Mail classification
- Improving existing systems (without breaking them)
- Text retrieval and searches on web

Learning Theory

THE AUSTRALIAN NATIONAL UNIVERSITY

Basic Problem

- How much should I believe my data?
- Real world analog: memorizing vs. understanding
- When is a hypothesis too complex, when too easy
- How can we tell that we can trust our estimate?

Tools

- Statistics
- Banach Spaces and Convex Geometry
- Combinatorial Methods

Parallel Computing and Data Mining

Basic Problem

- What if my algorithm scales worse than linear with the data?
- How to get kernel methods down from cubic to linear
- Millions observations, how to extract information, make algorithms stable.
- High Performance Numerical Algebra
- Optimization

Toys

- RSISE/CSL Bunyip (Beowulf cluster of 200 Pentium III CPUs)
- APAC Supercomputer (Compaq cluster of 4x128 Alpha CPUs)
- RSISE network of 40+ Linux boxes

Logic and Inference

Basic Problem

- How to extract rules from data (if customer buys diapers and a sixpack, then will he buy playboy?)
- How to perform inference in the presence of uncertainty
- How to deal with categorical data
- Search algorithms and heuristics

Applications

- Web searching
- Drug discovery
- Knowledge discovery