An Idiot's guide to Support vector machines (SVMs)

R. Berwick, Village Idiot

SVMs: A New Generation of Learning Algorithms

- Pre 1980:
  - Almost all learning methods learned linear decision surfaces.
  - Linear learning methods have nice theoretical properties
- 1980's
  - Decision trees and NNs allowed efficient learning of non-linear decision surfaces
  - Little theoretical basis and all suffer from local minima
- 1990's
  - Efficient learning algorithms for non-linear functions based on computational learning theory developed
  - Nice theoretical properties.

Key Ideas

- Two independent developments within last decade
  - Computational learning theory
  - New efficient separability of non-linear functions that use "kernel functions"
- The resulting learning algorithm is an optimization algorithm rather than a greedy search.

Statistical Learning Theory

- Systems can be mathematically described as a system that
  - Receives data (observations) as input and
  - Outputs a function that can be used to predict some features of future data.
- Statistical learning theory models this as a function estimation problem
- Generalization Performance (accuracy in labeling test data) is measured

Organization

- Basic idea of support vector machines
  - Optimal hyperplane for linearly separable patterns
  - Extend to patterns that are not linearly separable by transformations of original data to map into new space – Kernel function
- SVM algorithm for pattern recognition

Unique Features of SVM's and Kernel Methods

- Are explicitly based on a theoretical model of learning
- Come with theoretical guarantees about their performance
- Have a modular design that allows one to separately implement and design their components
- Are not affected by local minima
- Do not suffer from the curse of dimensionality