Syllabus for the Lecture course

Introduction to machine learning, pattern recognition and statistical data modelling

Coryn A.L. Bailer-Jones
Max Planck Institute for Astronomy

SS2007, University of Heidelberg

12 lectures of 1hr 30 min.

1. Introduction

- course objectives, overview, schedule, bibliography, website etc.
- examples of problems and multivariate data
- goals of machine learning / data modelling (e.g. prediction, interpretation)
- simple methods: nearest neighbours, regression. As an illustration of the issues and how these things can be developed
- parametric vs. non-parametric data modelling
- overview of supervised learning: least squares, nearest neighbours, curse of dimensionality, regularization and generalization
- introduction to R

2. Data exploration, feature extraction, dimensionality reduction

- curse of dimensionality
- maximum likelihood method
- density estimation
- Principal Components Analysis

3. Linear methods part 1: regression

- importance of pre-processing and feature extraction.
- sphering the data
- linear regression and least squares
- regression shrinkage methods: ridge regression, lasso, etc.
- Singular Value Decomposition

4. Linear methods part 2: classification

- classification via density estimation: parametric models
- Linear Discriminant Analysis (LDA)

- Quadratic Discriminant Analysis (QDA)
- logistic regression
- Challenger disaster example

5. Basis expansions, splines

- bias-variance decomposition/trade-off
- basis expansions, splines
- generalized cross validation
- multidimensional splines

6. Kernel methods and additive models

- Local methods and kernels
- for regression: loess, locpoly
- variations on nearest neighbours: k-nn, local regression
- general additive models
- confidence intervals in splines

7. Neural networks, search and optimization

- MLPs with one and two layers; linear outputs
- cross entropy error function
- error functions: sum-of-squares, cross entropy etc.
- optimization methods
- Simulated annealing (actually, at end of lecture 8)
- Example: genetic algorithm for designing a filter system

8. More nonlinear stuff

- ANN application
- Radial Basis Function neural networks
- MARS
- Naive Bayes' classifier
- mixture models
- EM algorithm

9. Support vector machines

- separating hyperplanes
- linearly separable case
- non-separable case
- nonlinear classification via kernels
- support vector machines
- multiclass SVMs

10. Model selection and combination

- SVMs for regression
- AIC
- BIC
- model selection with Mclust using BIC
- combining weak learners: boosting
- boosting classification trees
- VC dimension

11. Unsupervised learning and clustering

- idea of an unsupervised algorithm
- quick summary of unsupervised algorithms we're already seen: mixture modelling, PCA
- K-means clustering
- K-medoids clustering
- vector quantization
- hierarchical clustering: agglomerative, divisive, hybrid
- distance metrics
- self-organizing maps

12. The final lecture

- Summary of course, in particular looking at the major/unifying concepts and common issues