Statistical Data Analysis with Positive Definite Kernels

Kenji Fukumizu

Institute of Statistical Mathematics, ROIS
Department of Statistical Science, Graduate University for Advanced Studies

October 6-10, 2008, Kyushu University

Outline of this course

Information on this course

Outline I

6 (Mon) Introduction: overview of kernel methods

- Basic idea of kernel method
- Examples of kernel methods

Basics on positive definite kernels

- Positive definite kernels
- Reproducing kernel Hilbert spaces

7 (Tue) Methods with kernels (I)

- Converting data with kernel
- Kernel PCA, kernel CCA

Methods with kernels (II)

- introduction to SVM
- Representer theorem
- Structured data

Outline II

8 (Wed) Support vector machine (I)

- Basics on convex analysis
- Optimization of SVM and its dual form
- Computational aspect and SMO

Support vector machine (II)

- Extension to multiclass and structured output
- Generalization of SVM

Seminar: Dependence analysis with positive definite kernel and its application

Outline III

9 (Thu) Theory of positive definite kernel and reproducing kernel Hilbert space

- Negative definite kernel and Schönberg's theorem
- Various examples of positive definite kernels
- Bochner's theorem, Mercer's theorem

Statistical inference with positive definite kernels (I)

- Mean on RKHS and Characteristic kernel
- Covariance on RKHS and independence

10 (Fri) Statistical inference with positive definite kernels (II)

- Measuring conditional independence with kernels
- Relation to other measures

Relation to other statistical methods

- Relation to functional data analysis, Gaussian process, and spline

Outline of this course

Information on this course

Comments on Terminology

- "Kernel" is a general word for a function of the form

$$k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}.$$

  But, "kernel" is often used to mean "positive definite kernel" in the methodology discussed in this course.

- Traditionally in statistics, "kernel method" often implies the method of kernel density estimation or Parzen window approach:

$$p(x) = \frac{1}{N} \sum_{i=1}^{N} k(x, X_i).$$

- In this course, "kernel method" is used for "the method with positive definite kernels".

Tips

Web page: `http://www.ism.ac.jp/~fukumizu/Kyushu2008/`

The information and the slides for this course will be put on the web page.

Time Table

	6 (Mon)	7 (Tue)	8 (Wed)	9 (Thu)	10 (Fri)
AM 10:30-12:00		Methods (I)	SVM (I)	Theory	Statistical inference (II)
PM(1) 14:00-	Intro.	Methods (II)	SVM (II)	Statistical inference (I)	Relation to other methods
PM(2) -16:30	Basics on pos. def. kernels	Methods (II)	(Seminar. 16:00-)	Statistical inference (I)	