

Kernel Method for Bayesian Inference

Kenji Fukumizu

The Institute of Statistical Mathematics.
Graduate University of Advanced Studies /
Tokyo Institute of Technology

Nov. 17-26, 2010

Intensive Course at Tokyo Institute of Technology



Outline I

17 (Wed) Introduction: overview of kernel methods

- Basic ideas of kernel method
- Examples of kernel methods

Basics on positive definite kernels

- Positive definite kernels
- Reproducing kernel Hilbert spaces

18 (Thu) Methods with kernels (I)

- Kernel PCA, kernel CCA, kernel FDA, Basics of SVM

Methods with kernels (II)

- Principle of kernel methods
- Representer theorem, etc.

Outline II

19 (Fri) Support vector machine and related topics

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

Support vector machine and related topics

- Extension of SVM
- Generalization ability of SVM: computational learning theory

Outline III

24 (Wed) Theory of positive definite kernel and RKHS (I)

- Positive and negative definite kernels
- Various examples of positive definite kernels

Theory of positive definite kernel and RKHS (II)

- Bochner's theorem, Mercer's theorem
- Explicit expression of RKHS

25 (Thu) Kernel methods for structured data

- Kernels for strings and graphs

Nonparametric inference with kernels (I)

- Mean on covariance on RKHS
- Characteristic property

Outline IV

26 (Fri) Nonparametric inference with kernels (II)

- Homogeneity and independence test
- Conditional independence with kernels

Relation to other statistical methods

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

Comments on Terminology

- “Kernel” is a general word for a function of the form

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

But, “kernel” is often used to mean “positive definite kernel” for the methodology in this course in machine learning community.

- In statistics, the word “kernel” is often used for the method of kernel density estimation or Parzen window approach, *e.g.*,

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

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

Tips

Web page:

`http://www.ism.ac.jp/~fukumizu/TITECH2010/`

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

Time Table

1st week	17 (Wed)	18 (Tue)	19 (Wed)
15:00-16:30	Intro	Methods (I)	SVM (I)
16:40-18:10	Basics on pos. def. kernels	Methods (II)	SVM (II)
2nd week	24 (Wed)	25 (Thu)	26 (Fri)
15:00-16:30	Theory on kernel and RKHS (I)	Structured data	Nonparametric inference (II)
16:40-18:10	Theory on kernel and RKHS (II)	Nonparametric inference (I)	Relation to other methods