

# Statistical inference with reproducing kernel Hilbert space

Kenji Fukumizu

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

April 18, 2008 / Statistical Learning Theory II

1 Outline of this course

2 Information on this course

# Outline I

- 1 Introduction: overview of kernel methods
  - Basic idea of kernel method
  - Examples of kernel methods
- 2 Elements of positive definite kernel and reproducing kernel Hilbert space
  - Definition and properties of positive definite kernel
  - Quick introduction to Hilbert spaces
  - Reproducing kernel Hilbert spaces
- 3 Methods with kernels
  - Kernel CCA, Kernel FDA, Kernel XXX
  - Representer theorem

# Outline II

- ④ Elements of support vector machine
  - Large margin classifier
  - Optimization in dual form
  - Computational aspect and SMO
  - Software and simulations
- ⑤ Generalization ability of SVM
  - Uniform bound of generalization
  - Methods in computational learning theory
- ⑥ Extension and relatives of SVM
  - Multiclass classification
  - Structured output
  - Various optimizations

## Outline III

- ⑦ 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
  - Preliminaries from operator theory
  - Mercer's theorem
- ⑧ Kernel methods for probabilities
  - Random element in RKHS
  - Mean in RKHS
  - Characteristic kernel
  - Homogeneity test with kernels

# Outline IV

- 9 Independence and kernels
  - Covariance operator and independence
  - Independence test with kernels
  - kernel ICA
- 10 Conditional independence and kernels
  - Conditional covariance operator
  - Conditional independence test
  - Application to dimension reduction

1 Outline of this course

2 Information on this course

# 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" 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/H20\_kernel/`

**Internal server:**

`http://w3.ism.ac.jp/~fukumizu/H20\_kernel/`

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

# Schedule

In April, May, June, and July,

- Class: Apr. 18, 25, May 2, 9, 23, 30, June 13, 20, July 25,  
...
- **No class: May 16, (June 6), June 27, July 4, 11, 18.**