Kernel Machines and Additive Fuzzy Systems:
Classification and Function Approximation

Yixin Chen and *James Z. Wang*

The Pennsylvania State University

http://www.cse.psu.edu/~yixchen

Outline

- Introduction
- VC theory and Support Vector Machines
- Additive fuzzy systems and kernel machines
- Support vector learning for a class of additive fuzzy systems
- Experimental results
- Conclusions and future work

Introduction

- Building a fuzzy system
  - Structure identification
  - Parameter estimation
  - Model validation
- Do we get a good fuzzy model?
  - How capable can a fuzzy model be?
  - How well can the model generalize?

Introduction (continued)

- Several types of fuzzy models are "universal approximators"
- Generalization performance
  - Structural risk minimization
  - Bias variance dilemma
  - Overfitting phenomena

  ⇒A "right" tradeoff between training accuracy and model complexity

Introduction (continued)

- Two approaches to find a “right” tradeoff
  - Cross-validation for model selection
  - Model reduction to simplify the model
- Vapnik-Chervonenkis (VC) theory
  - A general measure of model set complexity
  - Bounds on generalization
  - Support Vector Machines (SVM)

Outline

- Introduction
- VC theory and Support Vector Machines
- Additive fuzzy systems and kernel machines
- Support vector learning for a class of additive fuzzy systems
- Experimental results
- Conclusions and future work

VC Theory and Support Vector Machines

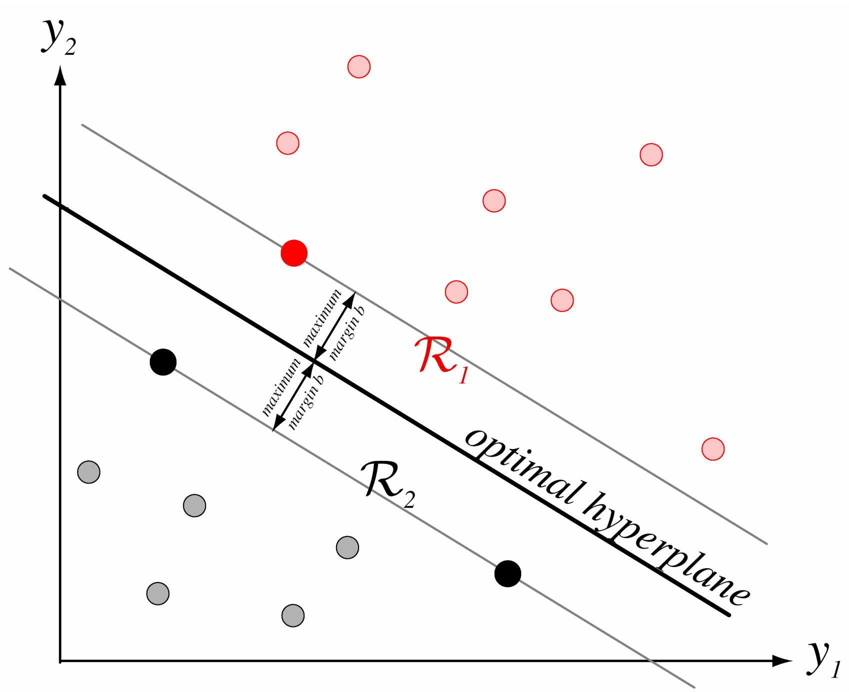
- One result from VC Theory

  Binary classification: given a set of training samples $\{(\vec{x}_1, y_1), \cdots, (\vec{x}_l, y_l)\} \subset \mathbb{R}^n \times \{+1, -1\}$ drawn independently from some unknown distribution $P(\vec{x}, y)$, with probability $1 - \eta$, the probability of misclassification for any decision function $f \in \mathbb{H}$ is bounded above by

$$R_{P(\vec{x},y)}(f) \leq R_{emp}(f) + \sqrt{\frac{h(1 + \ln\frac{2l}{h}) - \ln\frac{\eta}{4}}{l}}$$

VC Theory and Support Vector Machines (continued)

- Support Vector Machines (SVMs)
  - Optimal separating hyperplane



VC Theory and Support Vector Machines (continued)

- Kernel trick

A Mercer kernel is a function,

$$K : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R} \quad ,$$

satisfying

$$K(\vec{x}, \vec{z}) = \langle \Phi(\vec{x}), \Phi(\vec{z}) \rangle_{\mathbb{F}}$$

where $\Phi(\vec{x})$ is sometimes referred to as the Mercer features

VC Theory and Support Vector Machines (continued)

- Quadratic programming

$$\text{maximize } W(\vec{\alpha}) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{l} \alpha_i \alpha_j y_i y_j K(\vec{x}_i, \vec{x}_j)$$

$$\text{subject to } C \geq \alpha_i \geq 0,\ i = 1, \cdots, l,\ \text{and} \sum_{i=1}^{l} \alpha_i y_i = 0$$

- Decision function

$$f(\vec{x}) = \operatorname{sgn}\left(\sum_{i=1}^{l} y_i \alpha_i K(\vec{x}, \vec{x}_i) + b\right)$$

Outline

- Introduction
- VC theory and Support Vector Machines
- Additive fuzzy systems and kernel machines
- Support vector learning for a class of additive fuzzy systems
- Experimental results
- Conclusions and future work

Additive Fuzzy Systems and Kernel Machines

- Kernel Machines

$$f(\vec{x}) = \sum_{i=1}^{l} \alpha_i K(\vec{x}, \vec{x}_i) + b$$

- A class of additive fuzzy systems is functionally equivalent to a class of kernel machines

Additive Fuzzy Systems and Kernel Machines (continued)

- Additive Fuzzy System (AFS)
  - *m* fuzzy rules of the form

$$\text{Rule } j: \quad \text{IF } \mathbf{A}_j^1 \text{ AND } \mathbf{A}_j^2 \text{ AND } \cdots \text{ AND } \mathbf{A}_j^n \text{ THEN } b_j$$

  - Product as fuzzy conjunction operator
  - Addition for fuzzy rule aggregation
  - First order moment defuzzification
  - Reference function
  - Kernel is the product of reference functions

Additive Fuzzy Systems and Kernel Machines (continued)

- Positive definite fuzzy systems (PDFS)
  - Reference functions are positive definite functions → Mercer kernels

    Examples:

Gaussian	Symmetric triangle
Cauchy	Hyperbolic secant
Laplace	Squared sinc

Outline

- Introduction
- VC theory and Support Vector Machines
- Additive fuzzy systems and kernel machines
- Support vector learning for a class of additive fuzzy systems
- Experimental results
- Conclusions and future work

Support Vector Learning for a Class of Additive Fuzzy Systems

SVM		PDFS
Kernel	⟺	Reference functions
Support vectors	⟺	IF-part of fuzzy rules
Lagrange multiplier	⟺	THEN-part of fuzzy rules

Outline

- Introduction
- VC theory and Support Vector Machines
- Additive fuzzy systems and kernel machines
- Support vector learning for a class of additive fuzzy systems
- Experimental results
- Conclusions and future work

Experimental Results

- USPS data set
  - Training data (7291), testing data (2007)
  - 5-fold cross-validation to determine parameters

Experimental Results (continued)

Reference Function	$r \pm$STD	m
Gaussian	$95.2\% \pm 0.3\%$	573
Cauchy	$95.2\% \pm 0.3\%$	567
Laplace	$94.7\% \pm 0.4\%$	685
Symmetric Triangle	$95.0\% \pm 0.3\%$	652
Hyperbolic Secant	$95.0\% \pm 0.3\%$	468
Squared Sinc	$95.2\% \pm 0.2\%$	391

Linear SVM: 91.3%

k-nearest neighbor: 94.3%

Outline

- Introduction
- VC theory and Support Vector Machines
- Additive fuzzy systems and kernel machines
- Support vector learning for a class of additive fuzzy systems
- Experimental results
- Conclusions and future work

Conclusions and Future Work