Mathematical Results Concerning Kernel Techniques

Robert Schaback

schaback@math.uni-goettingen.de

Georg–August–Universität Göttingen

Overview

Overview

Black–Box systems

Overview

Black–Box systems

Learning Machines

Overview

- Identification of Black–Box systems
  Learning Machines

Overview

- Identification of Black–Box systems is Training of Learning Machines

Overview

- Identification of Black–Box systems is Training of Learning Machines

Kernel Techniques

Overview

- Identification of Black–Box systems is Training of Learning Machines

If Kernel Techniques are used:

Overview

- Identification of Black–Box systems
  is Training of Learning Machines

If Kernel Techniques are used:

- Kernel selection
  allows optimal Model selection

Overview

- Identification of Black–Box systems
  is Training of Learning Machines

If Kernel Techniques are used:

- Kernel selection
  allows optimal Model selection
- Inexact reproduction
  is Complexity reduction

Overview

- Identification of Black–Box systems
  is Training of Learning Machines

If Kernel Techniques are used:

- Kernel selection
  allows optimal Model selection
- Inexact reproduction
  is Complexity reduction

Connection to Support Vector Machines

Systems

Systems

$$Input \;\mapsto\; Output$$

Systems

$$\begin{array}{ccc} Input & \mapsto & Output \\ \mathbb{R}^M & \xrightarrow{F} & \mathbb{R}^N \end{array}$$

Systems

$$
\begin{array}{ccc}
Input & \mapsto & Output \\
\mathbb{R}^M & \xrightarrow{F} & \mathbb{R}^N
\end{array}
$$

Unknown Transfer function F

Systems

$$Input \mapsto Output$$
$$I\!R^M \xrightarrow{F} I\!R^N$$

Unknown Transfer function F

Known Observations $(x_j, y_j) \in I\!R^M \times I\!R^N$

Systems

$$\begin{array}{ccc} Input & \mapsto & Output \\ \mathbb{R}^M & \stackrel{F}{\longrightarrow} & \mathbb{R}^N \end{array}$$

Unknown	Transfer function F
Known	Observations $(x_j, y_j) \in \mathbb{R}^M \times \mathbb{R}^N$
Reproduction	$y_j = F(x_j)$

Black–Box–Models

Black–Box–Models

$$Input \ \mapsto \ Output$$

Black–Box–Models

$$
\begin{array}{ccc}
Input & \mapsto & Output \\
I\!R^M & \xrightarrow{F_p} & I\!R^N
\end{array}
$$

Black–Box–Models

$$Input \mapsto Output$$
$$\mathbb{R}^M \xrightarrow{F_p} \mathbb{R}^N$$

Unknown Model Transfer function F_p

Black–Box–Models

$$Input \mapsto Output$$
$$\mathbb{R}^M \xrightarrow{F_p} \mathbb{R}^N$$

Unknown Model Transfer function F_p

Known Observations $(x_j, y_j) \in \mathbb{R}^M \times \mathbb{R}^N$

Black–Box–Models

$$Input \mapsto Output$$
$$I\!R^M \xrightarrow{F_p} I\!R^N$$

Unknown Model Transfer function F_p

Known Observations $(x_j, y_j) \in I\!R^M \times I\!R^N$

Reproduction $y_j \approx F_p(x_j)$

Black–Box–Models

$$Input \mapsto Output$$
$$\mathbb{R}^M \xrightarrow{F_p} \mathbb{R}^N$$

Unknown	Model Transfer function F_p
Known	Observations $(x_j, y_j) \in \mathbb{R}^M \times \mathbb{R}^N$
Reproduction	$y_j \approx F_p(x_j)$
Unknown	Model Parameters $p \in P \subseteq \mathbb{R}^P$

Black–Box–Models

$$Input \mapsto Output$$
$$\mathbb{R}^M \xrightarrow{F_p} \mathbb{R}^N$$

Unknown	Model Transfer function F_p
Known	Observations $(x_j, y_j) \in \mathbb{R}^M \times \mathbb{R}^N$
Reproduction	$y_j \approx F_p(x_j)$
Unknown	Model Parameters $p \in P \subseteq \mathbb{R}^P$
Simulation:	Evaluate $y = F_p(x)$ for new inputs x

Black–Box–Models

$$Input \mapsto Output$$

$$\mathbb{R}^M \xrightarrow{F_p} \mathbb{R}^N$$

Unknown	Model Transfer function F_p
Known	Observations $(x_j, y_j) \in \mathbb{R}^M \times \mathbb{R}^N$
Reproduction	$y_j \approx F_p(x_j)$
Unknown	Model Parameters $p \in P \subseteq \mathbb{R}^P$
Simulation:	Evaluate $y = F_p(x)$ for new inputs x
Identification:	Find parameters $p \in P \subseteq \mathbb{R}^P$

Black–Box–Models

$$\begin{array}{ccc} Input & \mapsto & Output \\ \mathbb{R}^M & \xrightarrow{F_p} & \mathbb{R}^N \end{array}$$

Unknown	Model Transfer function F_p
Known	Observations $(x_j, y_j) \in \mathbb{R}^M \times \mathbb{R}^N$
Reproduction	$y_j \approx F_p(x_j)$
Unknown	Model Parameters $p \in P \subseteq \mathbb{R}^P$
Simulation:	Evaluate $y = F_p(x)$ for new inputs x
Identification:	Find parameters $p \in P \subseteq \mathbb{R}^P$
Learning:	same

Learning

Learning

$$Input \mapsto Output$$

Learning

$$\begin{array}{ccc} Input & \mapsto & Output \\ \mathbb{R}^M & \xrightarrow{F_p} & \mathbb{R}^N \end{array}$$

Learning

$$\begin{array}{ccc} Input & \mapsto & Output \\ I\!R^M & \xrightarrow{F_p} & I\!R^N \end{array}$$

Unknown Model behaviorr function F_p

Learning

$$\begin{array}{ccc} Input & \mapsto & Output \\ \mathbb{R}^M & \xrightarrow{F_p} & \mathbb{R}^N \end{array}$$

Unknown Model behaviorr function F_p

Known Training samples $(x_j, y_j) \in \mathbb{R}^M \times \mathbb{R}^N$

Learning

$$\begin{array}{ccc} Input & \mapsto & Output \\ I\!R^M & \xrightarrow{F_p} & I\!R^N \end{array}$$

Unknown	Model behaviorr function F_p
Known	Training samples $(x_j, y_j) \in I\!R^M \times I\!R^N$
Reproduction	$y_j \approx F_p(x_j)$

Learning

$$Input \mapsto Output$$
$$I\!R^M \xrightarrow{F_p} I\!R^N$$

Unknown	Model behaviorr function F_p
Known	Training samples $(x_j, y_j) \in I\!R^M \times I\!R^N$
Reproduction	$y_j \approx F_p(x_j)$
Unknown	Model Parameters $p \in P \subseteq I\!R^P$

Learning

$$\begin{array}{ccc} Input & \mapsto & Output \\ \mathbb{R}^M & \xrightarrow{F_p} & \mathbb{R}^N \end{array}$$

Unknown	Model behaviorr function F_p
Known	Training samples $(x_j, y_j) \in \mathbb{R}^M \times \mathbb{R}^N$
Reproduction	$y_j \approx F_p(x_j)$
Unknown	Model Parameters $p \in P \subseteq \mathbb{R}^P$
Behavior:	Evaluate $y = F_p(x)$ for new inputs x

Learning

$$\begin{array}{ccc} Input & \mapsto & Output \\ I\!R^M & \xrightarrow{F_p} & I\!R^N \end{array}$$

Unknown	Model behaviorr function F_p
Known	Training samples $(x_j, y_j) \in I\!R^M \times I\!R^N$
Reproduction	$y_j \approx F_p(x_j)$
Unknown	Model Parameters $p \in P \subseteq I\!R^P$
Behavior:	Evaluate $y = F_p(x)$ for new inputs x
Learning:	Find parameters $p \in P \subseteq I\!R^P$

Consequence

Consequence

- Identification in black–box systems
  is Training in Learning

Consequence

- Identification in black–box systems is Training in Learning

Both problems:

Consequence

- Identification in black–box systems
  is Training in Learning

Both problems:

- Reconstruction of an unknown multivariate function from scattered data

Feature Map

Feature Map

$$\mathbb{R}^M \overset{\Phi}{\longrightarrow} \mathcal{F}$$

Feature Map

$$\mathbb{R}^M \xrightarrow{\Phi} \mathcal{F}$$

$$\text{Input/Stimulus} \mapsto \text{features}$$

Feature Map

$$\mathbb{R}^M \xrightarrow{\Phi} \mathcal{F}$$

Input/Stimulus $\mapsto$ features

$\mathbb{R}^M \xrightarrow{\Phi} \mathcal{F}$: Reproducing Kernel Hilbert Space

Feature Map

$$\mathbb{R}^M \xrightarrow{\Phi} \mathcal{F}$$

Input/Stimulus $\mapsto$ features

$\mathbb{R}^M \xrightarrow{\Phi} \mathcal{F}$: Reproducing Kernel Hilbert Space

$\mathcal{F} \subset \{v : \mathbb{R}^M \to \mathbb{R}\}$ (scalar case)

Kernel Trick

Kernel Trick

$\mathbb{R}^M \xrightarrow{\Phi} \mathcal{F}$: Reproducing Kernel Hilbert Space

Kernel Trick

$\mathbb{R}^M \xrightarrow{\Phi} \mathcal{F}$: Reproducing Kernel Hilbert Space

$\mathcal{F} \subset \{v : \mathbb{R}^M \rightarrow \mathbb{R}\}$ (scalar case)

Kernel Trick

$\mathbb{R}^M \xrightarrow{\Phi} \mathcal{F}$: Reproducing Kernel Hilbert Space

$\mathcal{F} \subset \{v : \mathbb{R}^M \rightarrow \mathbb{R}\}$ (scalar case)

K : Kernel $\mathbb{R}^M \times \mathbb{R}^M \rightarrow \mathbb{R}$

Kernel Trick

$\mathbb{R}^M \stackrel{\Phi}{\longrightarrow} \mathcal{F}$: Reproducing Kernel Hilbert Space

$\mathcal{F} \subset \{v : \mathbb{R}^M \to \mathbb{R}\}$ (scalar case)

K : Kernel $\mathbb{R}^M \times \mathbb{R}^M \to \mathbb{R}$

$$
\begin{aligned}
\Phi(x) &= K(x, \cdot) \\
(\Phi(x), \Phi(y))_{\mathcal{F}} &= (K(x,\cdot), K(y,\cdot))_{\mathcal{F}} \\
&= K(x,y) \\
(\Phi(x), v)_{\mathcal{F}} &= (K(x,\cdot), v)_{\mathcal{F}} \\
&= v(x) \quad \text{(Reproduction)}
\end{aligned}
$$

Optimal Models from Kernels

Optimal Models from Kernels

$$\min\{\|F\|_{\mathcal{F}} \;:\; F \in \mathcal{F},\; F(x_k) = y_k\}$$

Optimal Models from Kernels

$$\min\{\|F\|_{\mathcal{F}} \;:\; F \in \mathcal{F},\; F(x_k) = y_k\}$$

is solved by models of the form

$$F_\alpha(x) = \sum_j \alpha_j K(x_j, x)$$

with coefficients $\alpha_j \in I\!R^N$.

Optimal Models from Kernels

$$\min\{\|F\|_{\mathcal{F}} \ : \ F \in \mathcal{F},\ F(x_k) = y_k\}$$

is solved by models of the form

$$F_\alpha(x) = \sum_j \alpha_j K(x_j, x)$$

with coefficients $\alpha_j \in I\!R^N$.

$$\left.\begin{array}{rl} \text{Feature map} & \Phi \\ \text{Kernel} & K \\ \text{Space} & \mathcal{F} \\ \text{Data} & (x_k, y_k) \end{array}\right\} \Rightarrow \text{Optimal Model}$$

Consequence

Consequence

- Kernel selection

Consequence

- Kernel selection
  allows optimal Model selection

Consequence

- Kernel selection
  allows optimal Model selection

The optimal kernel model is determined by

Consequence

- Kernel selection
  allows optimal Model selection

The optimal kernel model is determined by

- Features (determine Φ and K)

Consequence

- Kernel selection
  allows optimal Model selection

The optimal kernel model is determined by

- Features (determine Φ and K)
- Norm in feature space (selection criterion)

Consequence

- Kernel selection
  allows optimal Model selection

The optimal kernel model is determined by

- Features (determine Φ and K)
- Norm in feature space (selection criterion)
- Data

Consequence

- Kernel selection
  allows optimal Model selection

The optimal kernel model is determined by

- Features (determine Φ and K)
- Norm in feature space (selection criterion)
- Data

New discipline: Kernel Engineering

Identification or Training

Identification or Training

Model $$F_\alpha(x) = \sum_j \alpha_j K(x_j, x)$$

Identification or Training

$$\text{Model} \qquad F_\alpha(x) = \sum_j \alpha_j K(x_j, x)$$

$$\text{Data} \qquad y_k = F_\alpha(x_k)$$

Identification or Training

Model	$F_\alpha(x)$	$=$	$\sum_j \alpha_j K(x_j, x)$
Data	y_k	$=$	$F_\alpha(x_k)$
System	y_k	$=$	$\sum_j \alpha_j K(x_j, x_k)$

Identification or Training

$$\begin{array}{lrcl} \text{Model} & F_\alpha(x) & = & \sum_j \alpha_j K(x_j, x) \\ \text{Data} & y_k & = & F_\alpha(x_k) \\ \text{System} & y_k & = & \sum_j \alpha_j K(x_j, x_k) \end{array}$$

Large positive definite symmetric system

Inexact Reproduction

Inexact Reproduction

Given $\epsilon \geq 0$.

Inexact Reproduction

Given $\epsilon \geq 0$.

$$\min\{\|F\|_{\mathcal{F}} \; : \; F \in \mathcal{F}, \; |F(x_k) - y_k| \leq \epsilon\}$$

Inexact Reproduction

Given $\epsilon \geq 0$.

$$\min\{\|F\|_{\mathcal{F}} \;:\; F \in \mathcal{F},\, |F(x_k) - y_k| \leq \epsilon\}$$

is solved by models of the form

$$F_\alpha(x) = \sum_j \alpha_j K(x_j, x)$$

Inexact Reproduction

Given $\epsilon \geq 0$.

$$\min\{\|F\|_{\mathcal{F}} \ : \ F \in \mathcal{F},\ |F(x_k) - y_k| \leq \epsilon\}$$

is solved by models of the form

$$F_\alpha(x) = \sum_j \alpha_j K(x_j, x)$$

with

$$|F(x_k) - y_k| < \epsilon \Rightarrow \alpha_k = 0$$

Inexact Reproduction

Given $\epsilon \geq 0$.

$$\min\{\|F\|_{\mathcal{F}} \ : \ F \in \mathcal{F}, \ |F(x_k) - y_k| \leq \epsilon\}$$

is solved by models of the form

$$F_\alpha(x) = \sum_j \alpha_j K(x_j, x)$$

with

$$|F(x_k) - y_k| < \epsilon \Rightarrow \alpha_k = 0$$

Support points (vectors) :

(x_k, y_k) with $|F(x_k) - y_k| = \epsilon$

Consequence

Consequence

- Inexact reproduction
  is Complexity reduction

Consequence

- Inexact reproduction
  is Complexity reduction

In the Learning Machines context:

Consequence

- Inexact reproduction
  is Complexity reduction

In the Learning Machines context:
Support Vector Machines

Consequence

- Inexact reproduction
  is Complexity reduction

In the Learning Machines context:
Support Vector Machines
Consequence of Optimization Theory

Consequence

- Inexact reproduction
  is Complexity reduction

In the Learning Machines context:
Support Vector Machines
Consequence of Optimization Theory

No Statistics

Conclusion

Conclusion

Black–Box systems

Conclusion

Black–Box systems

Learning Machines

Conclusion

- Identification of Black–Box systems
  Learning Machines

Conclusion

- Identification of Black–Box systems is Training of Learning Machines

Conclusion

- Identification of Black–Box systems is Training of Learning Machines

Kernel Techniques

Conclusion

- Identification of Black–Box systems is Training of Learning Machines

If Kernel Techniques are used:

Conclusion

- Identification of Black–Box systems is Training of Learning Machines

If Kernel Techniques are used:

- Kernel selection allows optimal Model selection

Conclusion

- Identification of Black–Box systems
  is Training of Learning Machines

If Kernel Techniques are used:

- Kernel selection
  allows optimal Model selection
- Inexact reproduction
  is Complexity reduction

Conclusion

- Identification of Black–Box systems is Training of Learning Machines

If Kernel Techniques are used:

- Kernel selection allows optimal Model selection
- Inexact reproduction is Complexity reduction

Connection to Support Vector Machines

Conclusion

- Identification of Black–Box systems
  is Training of Learning Machines

If Kernel Techniques are used:

- Kernel selection
  allows optimal Model selection
- Inexact reproduction
  is Complexity reduction

Connection to Support Vector Machines

http://www.num.math.uni-goettingen.de/schaback