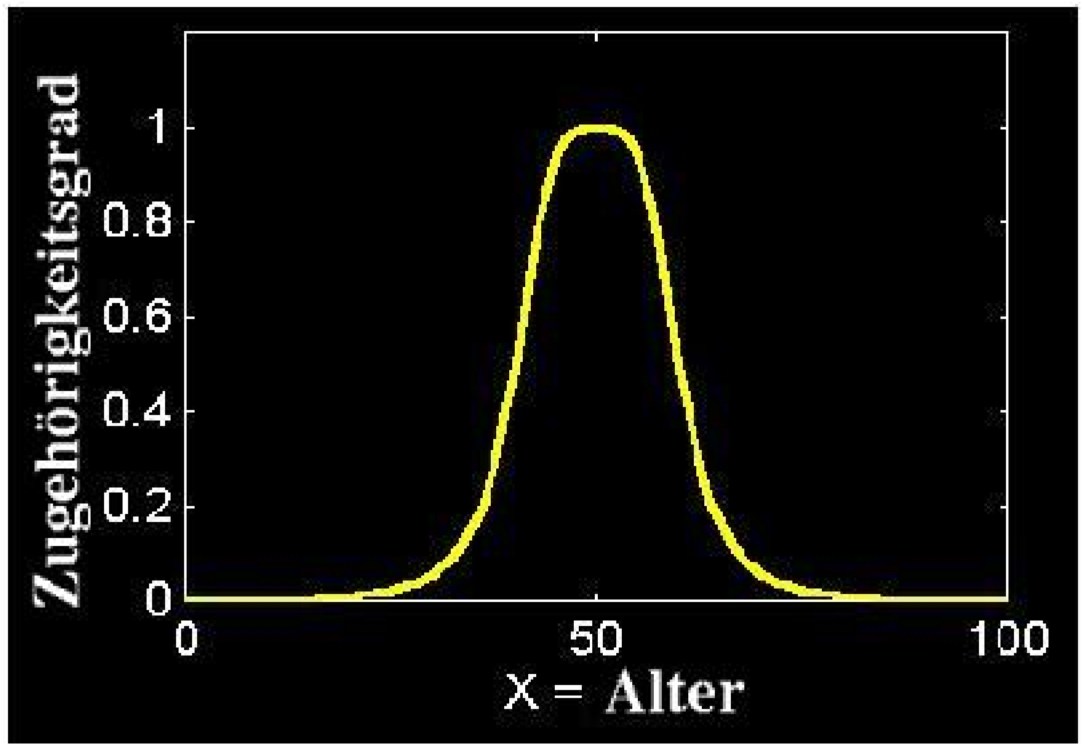
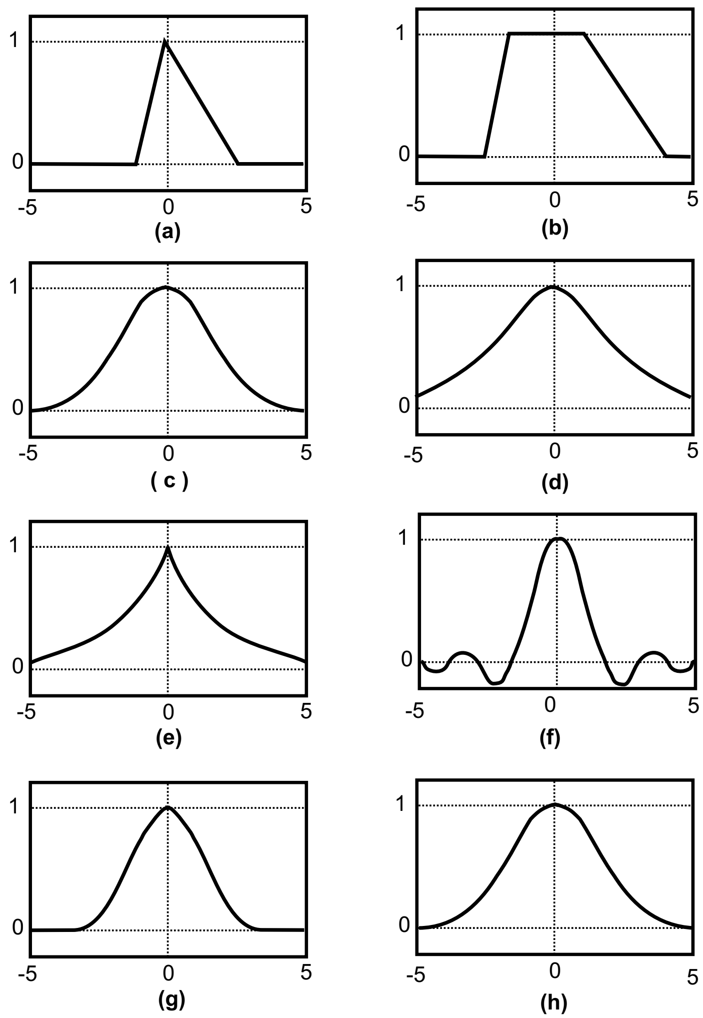
Symbolumwandlung der Kernfunktionen

Positiv definierte, konvexe Kernfunktionen können als Fuzzy-Mengen betrachtet werden. z.B.:

$$\mu_B(x) = \frac{1}{1+\left(\frac{x-50}{10}\right)^2}$$



ZF-Funktionen



Einführung in die Fuzzy-Mengen

Die Auswertung der Regeln eines Fuzzy-Reglers ist prinzipiell ein Interpolationsprozeß.

- ungenaue natürsprachliche Abstufungen von Begriffen wie “groß, “schön”, “stark” ...
- menschliche Denk- und Verhaltensmodelle mit der einstufigen Logik:

$\boxed{\text{Autofahren: “Wenn-Dann”-Regeln}}$

$\boxed{\text{Autoparken: genau auf den Millimeter?}}$

- Verwendung unscharfer Sprache statt numerischer Beschreibung:
  $\boxed{\text{bremse 2.52 m vor der Kurve}}$

  $\longrightarrow$ nur in Maschinesystemen
  $\boxed{\text{bremse kurz vor der Kurve}}$

  $\longrightarrow$ in natürlicher Sprache

Fuzzy-Mengen / Fuzzy-Logik als Mechanismus für

- ungenaue natursprachliche Abstufungen von Begriffen wie “groß, “schön”, “stark” ...

- Verwendung unscharfer Sprache statt numerischer Beschreibung.

- Abstraktion von unnötigen/zu komplexen Details.

- Modellierung menschlicher Denk- und Verhaltensprozesse mit der einstufigen Logik.

Charakteristische Funktion vs. Zugehörigkeitsfunktion

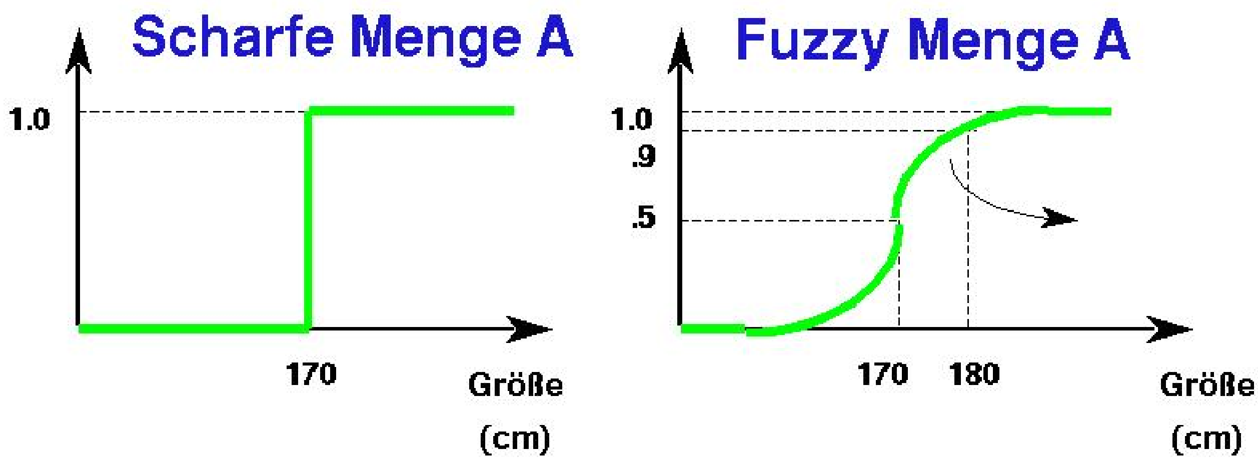
Für **Fuzzy-Mengen** A verwendet man eine verallgemeinerte charakteristische Funktion μ_A, die jedem Element $x \in X$ eine reelle Zahl aus $[0, 1]$ zuordnet — den "Grad", zu dem das Element x zur beschriebenen unscharfen Menge A gehört:

$$\mu_A : X \to [0, 1]$$

μ_A wird als Zugehörigkeitsfunktion (ZF) bezeichnet.

$$A = \{(x, \mu_A(x)|x \in X\}$$

Zugehörigkeitsfunktion



Charakteristik der kontinuierlichen ZF:

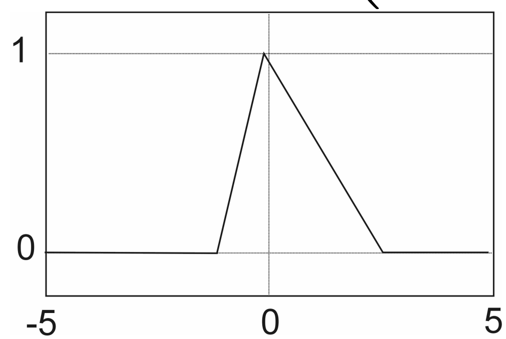
- Positiv definierte, konvexe Funktionen (einige wichtige Kernfunktionen).

- Subjektive Perzeption

- keine probalistische Funktionen

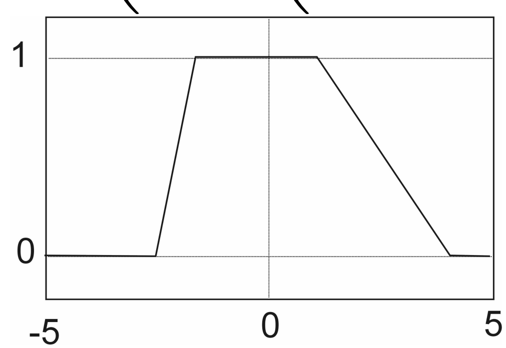
ZF-Formulierung - I

Triangulare: $trimf(x;a,b,c) = \max\left(\min\left(\frac{x-a}{b-a}, \frac{c-x}{c-b}\right), 0\right)$



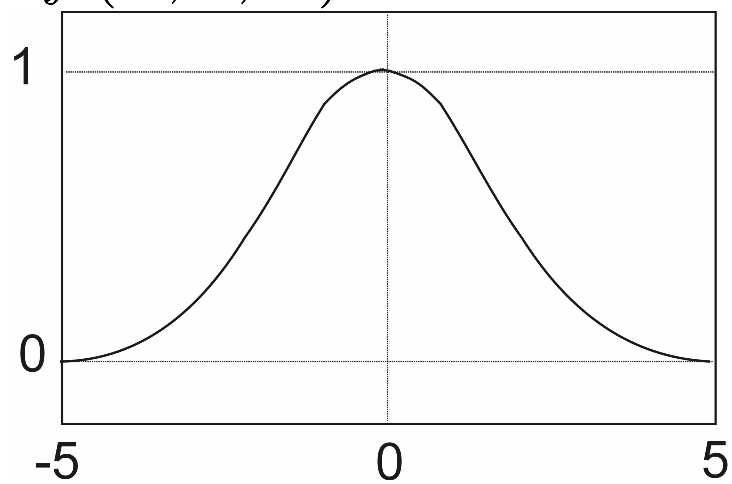
Trapezoidale:

$$trapmf(x;a,b,c,d) = \max\left(\min\left(\frac{x-a}{b-a}, 1, \frac{d-x}{d-c}\right), 0\right)$$

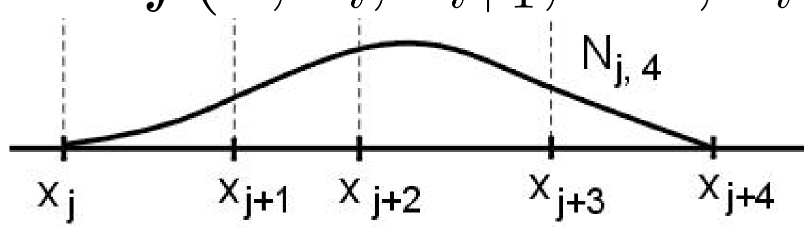


ZF-Formulierung - II

Gaussian: $gaussmf(x; c, \sigma) = e^{-\frac{1}{2}\left(\frac{x-c}{\sigma}\right)^2}$



B-Splines: $bsplinemf(x, x_i, x_{i+1}, \cdots, x_{i+k})$



Linguistische Variable

Eine numerische Variable nimmt numerische Werte an:

$$Das\_Alter = 25$$

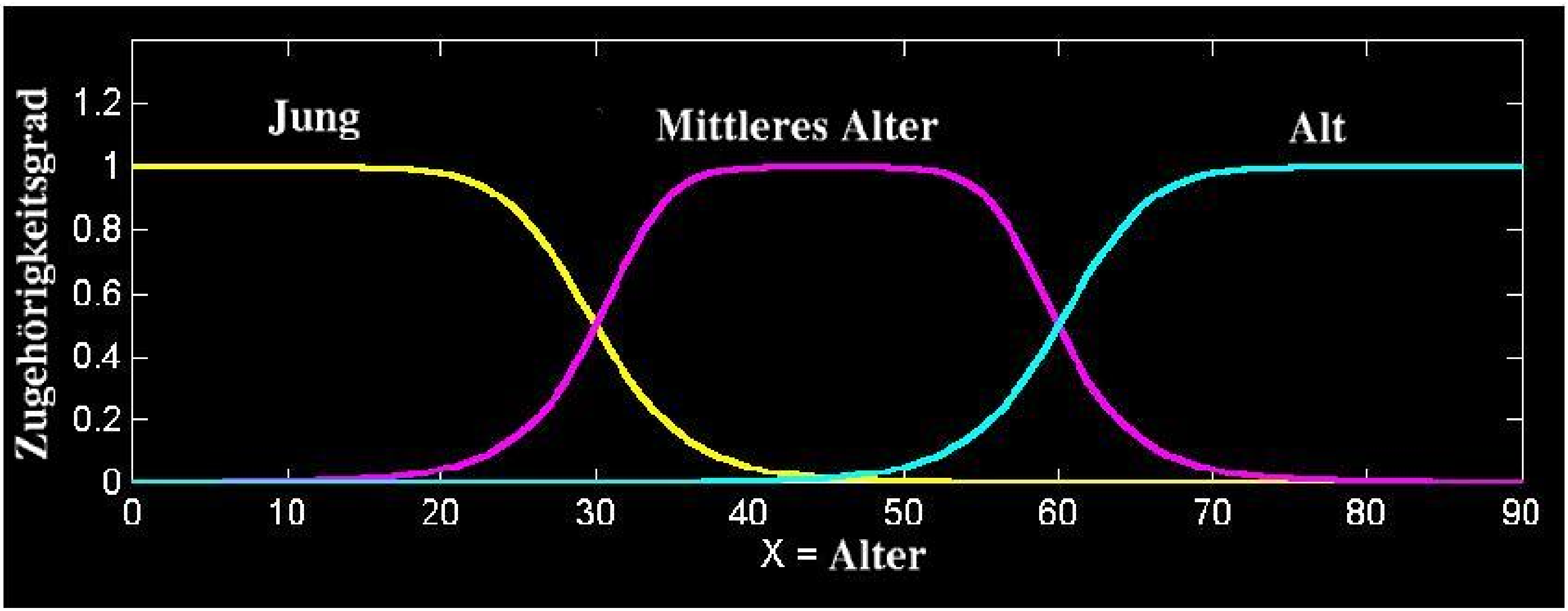
Eine linguistische Variable nimmt linguistische Werte (Terme) an:

$$Das\_Alter\ ist\ jung$$

Ein linguistischer Wert ist eine Fuzzy-Menge.

Fuzzy-Partition

Fuzzy-Partitionen über die linguistischen Werte “Jung”, “Mittleres Alter” und “Alt”:



Fuzzy-Logik: Ableitungsmechanismen

Eine Fuzzy-Regel wird folgendermaßen formuliert:

"IF A THEN B"

mit Fuzzy-Mengen A, B über den Universen X, Y.

Einer der wichtigsten Ableitungsmechanismen ist der verallgemeinerte Modus-Ponens (GMP):

Implikation:	IF x is A THEN y is B
Prämisse:	x is A'
Konklusion:	y is B'

Fuzzy-Systeme zur Funktionsapproximation

Grundidee:

- Beschreibung des gewünschten Reglerverhaltens mit Hilfe umgangssprachlicher, qualitativer Regeln.
- Quantifizierung linguistischer Werte durch Fuzzy-Mengen.
- Regel–Auswertung durch Verfahren der Fuzzy-Logik bzw. der Interpolation.

Fuzzy-Regeln:

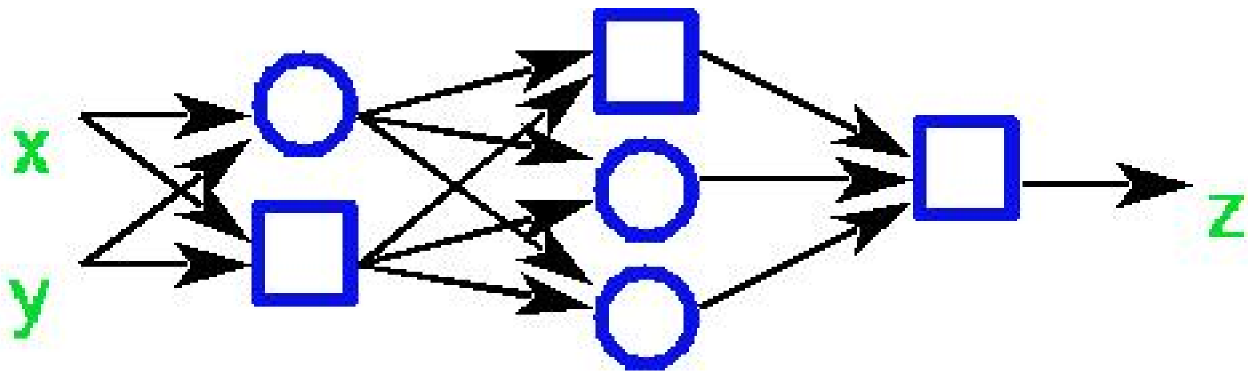
„**IF** (eine Menge Konditionen werden erfüllt)
THEN
(eine Menge Konsequenzen können bestimmt werden)"

In den Prämissen (Antecedenten) vom IF-Teil:
linguistische Variablen aus der Domäne der Prozeßzustände;

In den Konklusionen (Konsequenten) vom THEN-Teil:

linguistische Variablen aus der Regelungsdomäne.

Adaptive Netzwerke

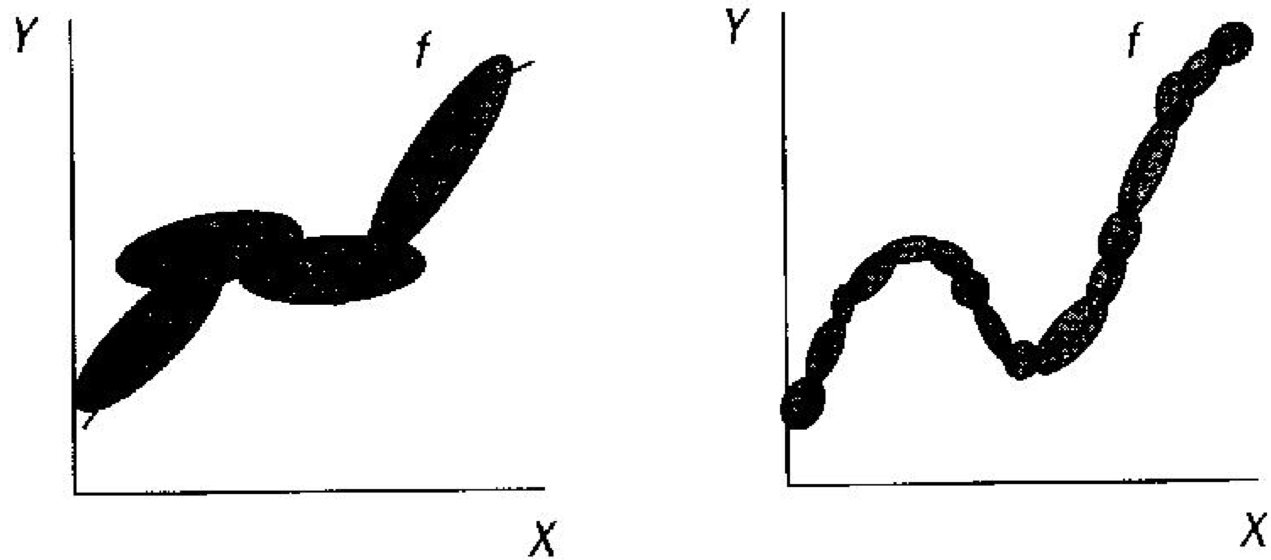


Architektur:

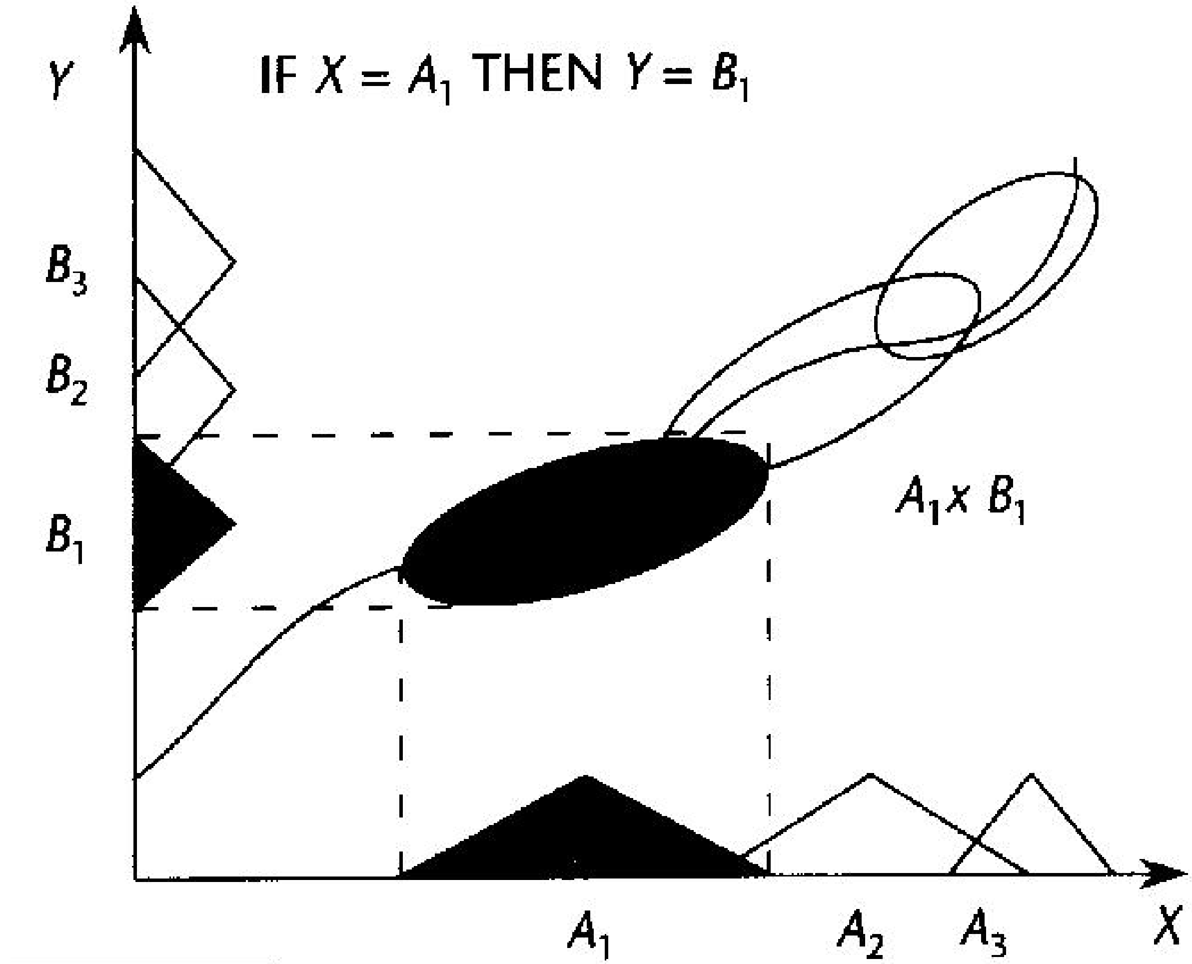
Feedforward Netzwerke mit verschiedenen Knoten-Funktionen

Rule Extraction

Die Fuzzy-Patchs (Kosko):



Ein Fuzzy-Regel-Patch:



Additive Systeme

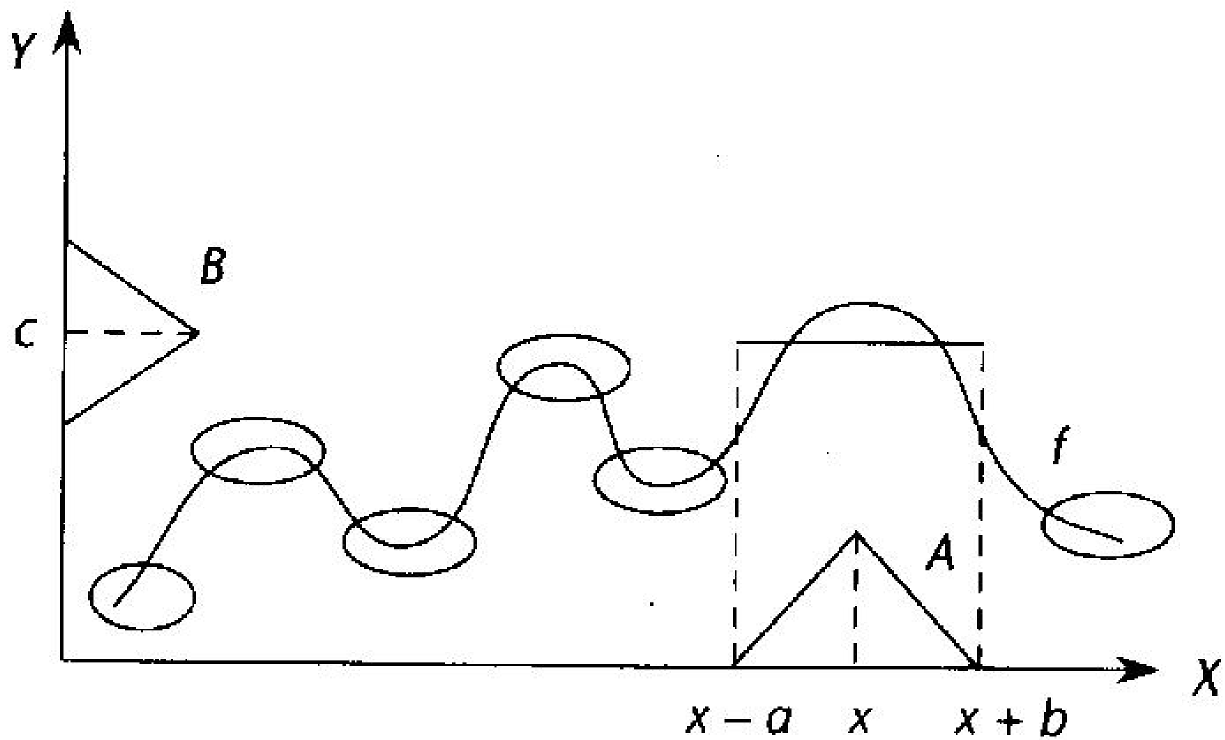
Ein additiver Fuzzy-Controller addiert die “THEN”-Teile der gefeuerten Regeln.

Fuzzy-Approximations-Satz:

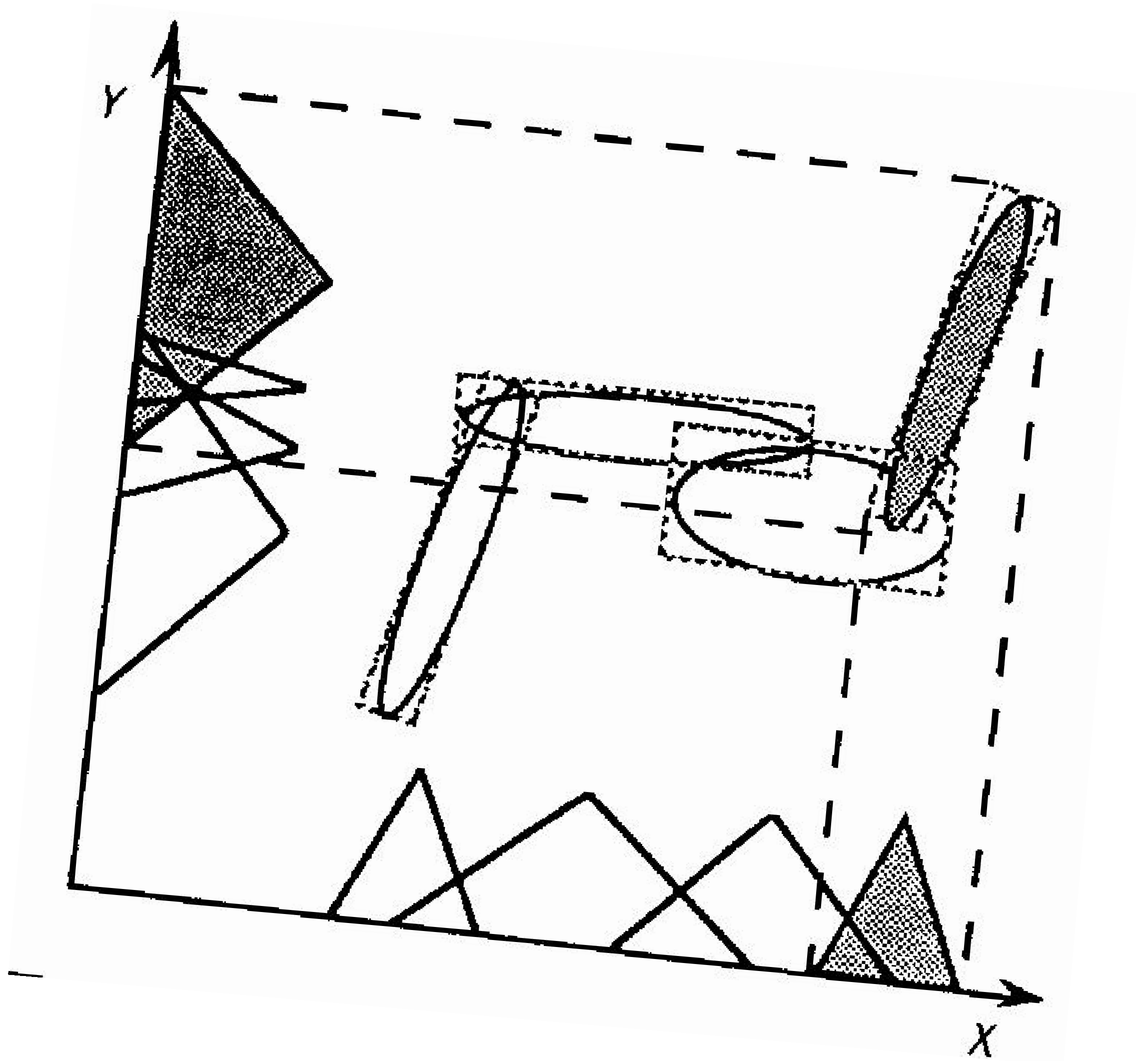
Ein additiver Fuzzy-Controller kann eine beliebige stetige Funktion $f : X \to Y$ approximieren wenn X kompakt ist.

Optimale Fuzzy-Regel-Patchs

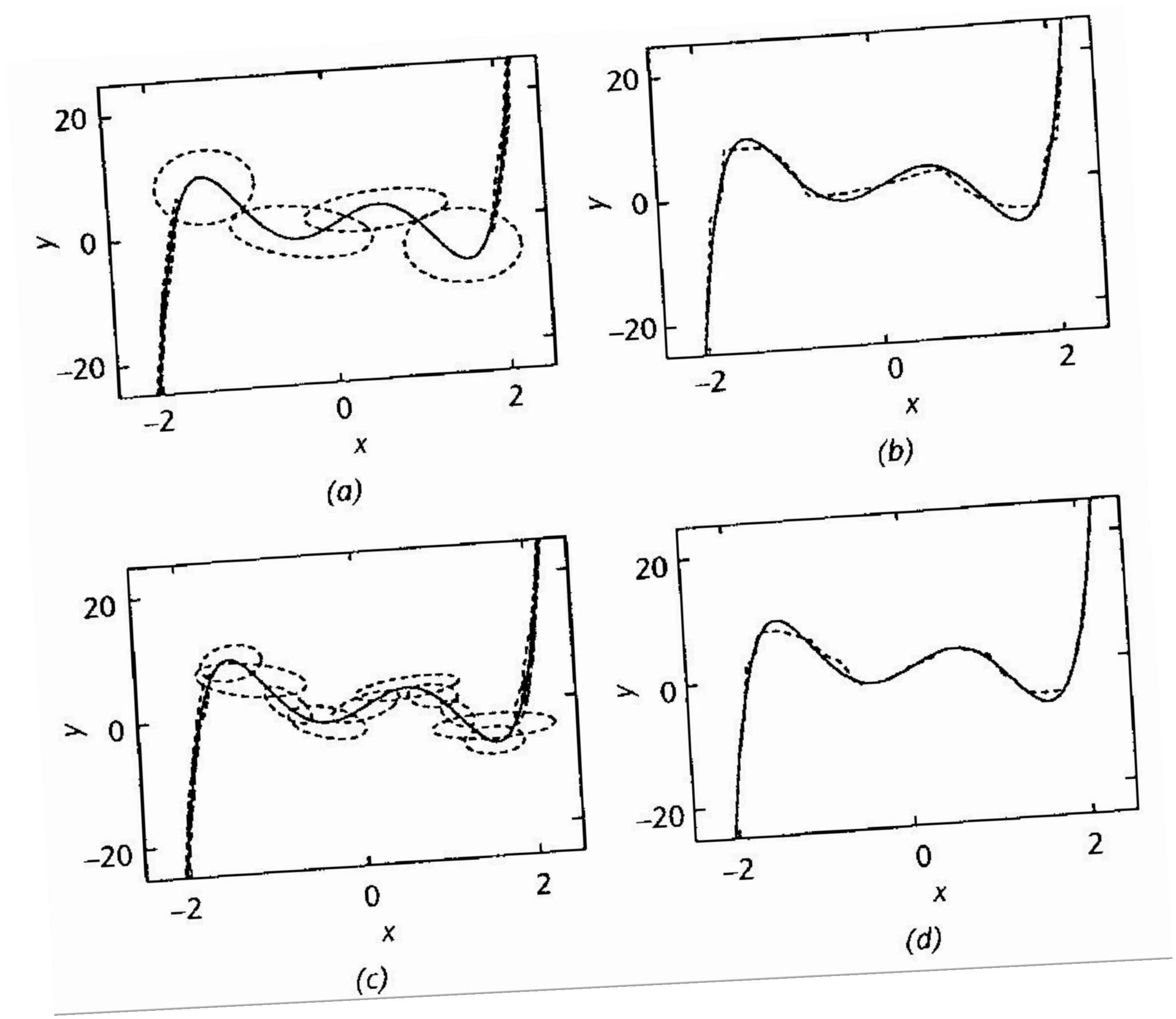
Optimale Fuzzy-Regel-Patchs decken die Extrema einer Funktion:



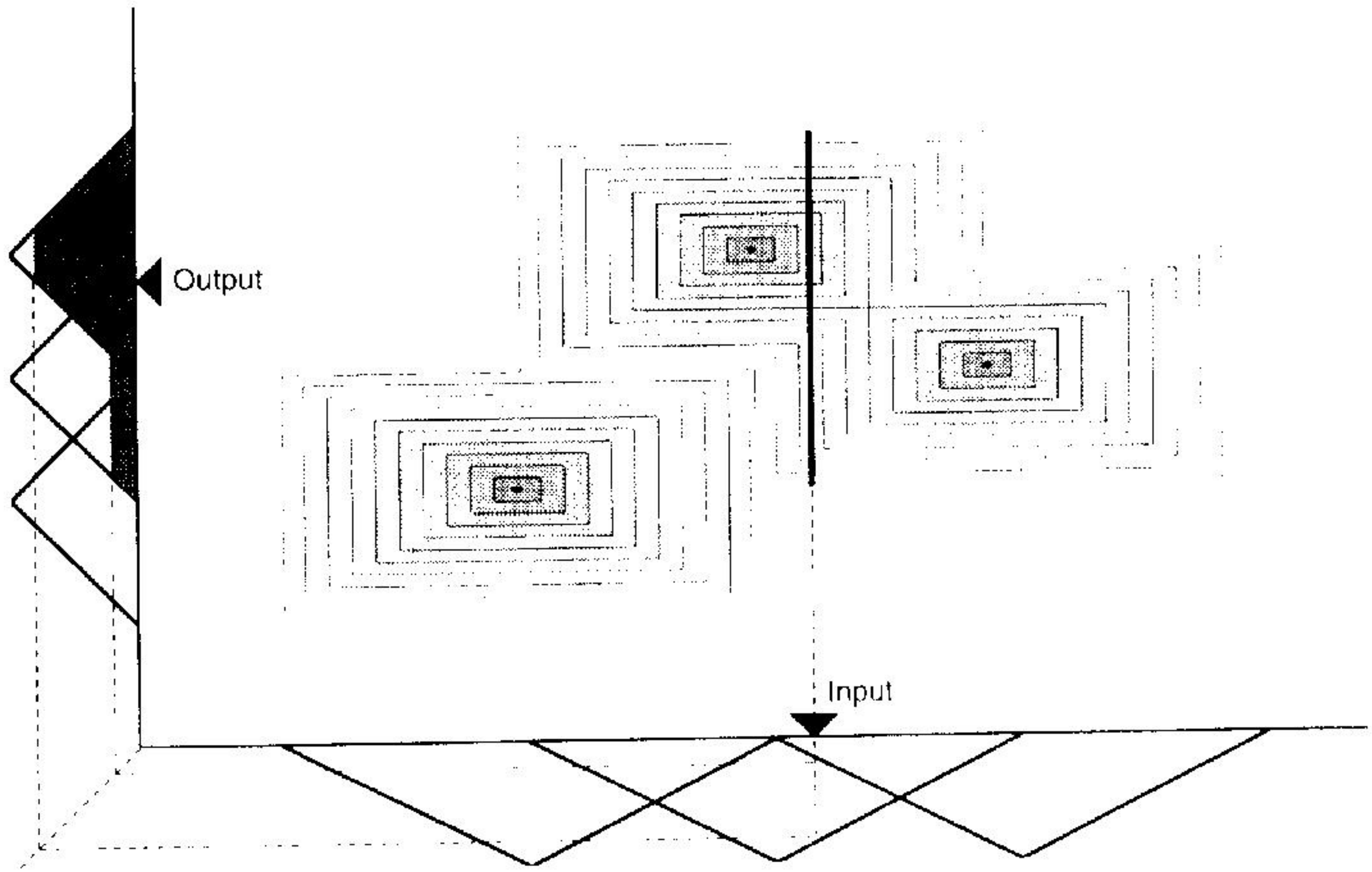
Projektion der Ellipsoide auf die Ein- und Ausgangsachse:



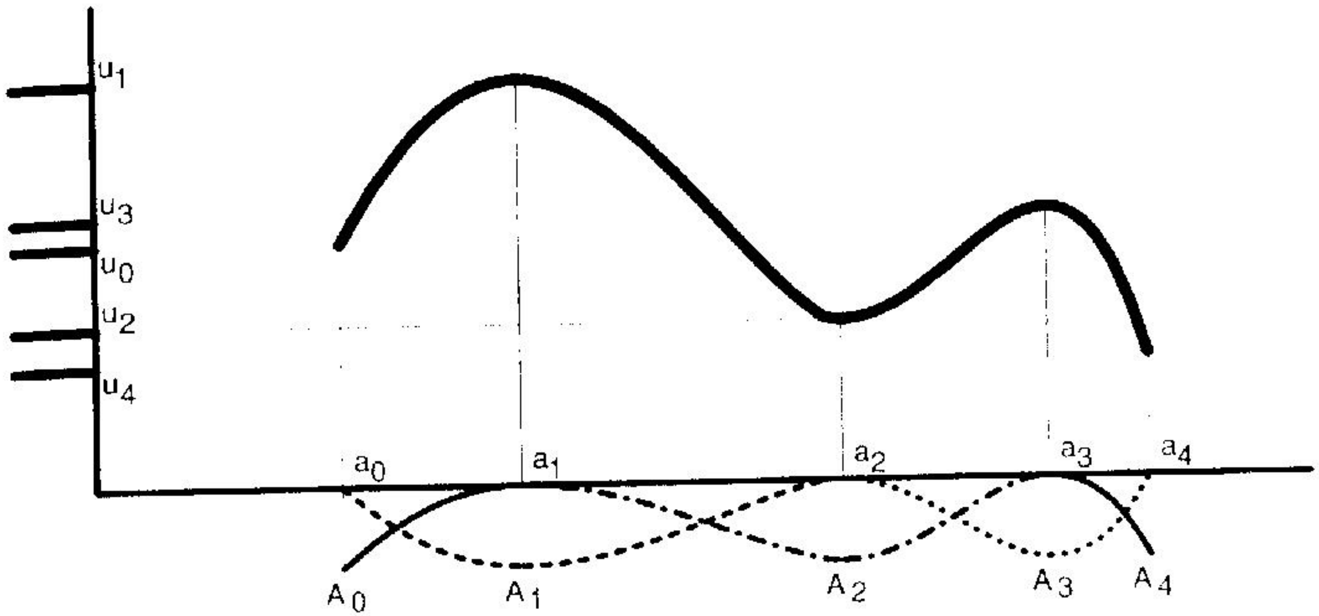
Die Größe eines Ellipsoids hängt von den Trainingsdaten ab.



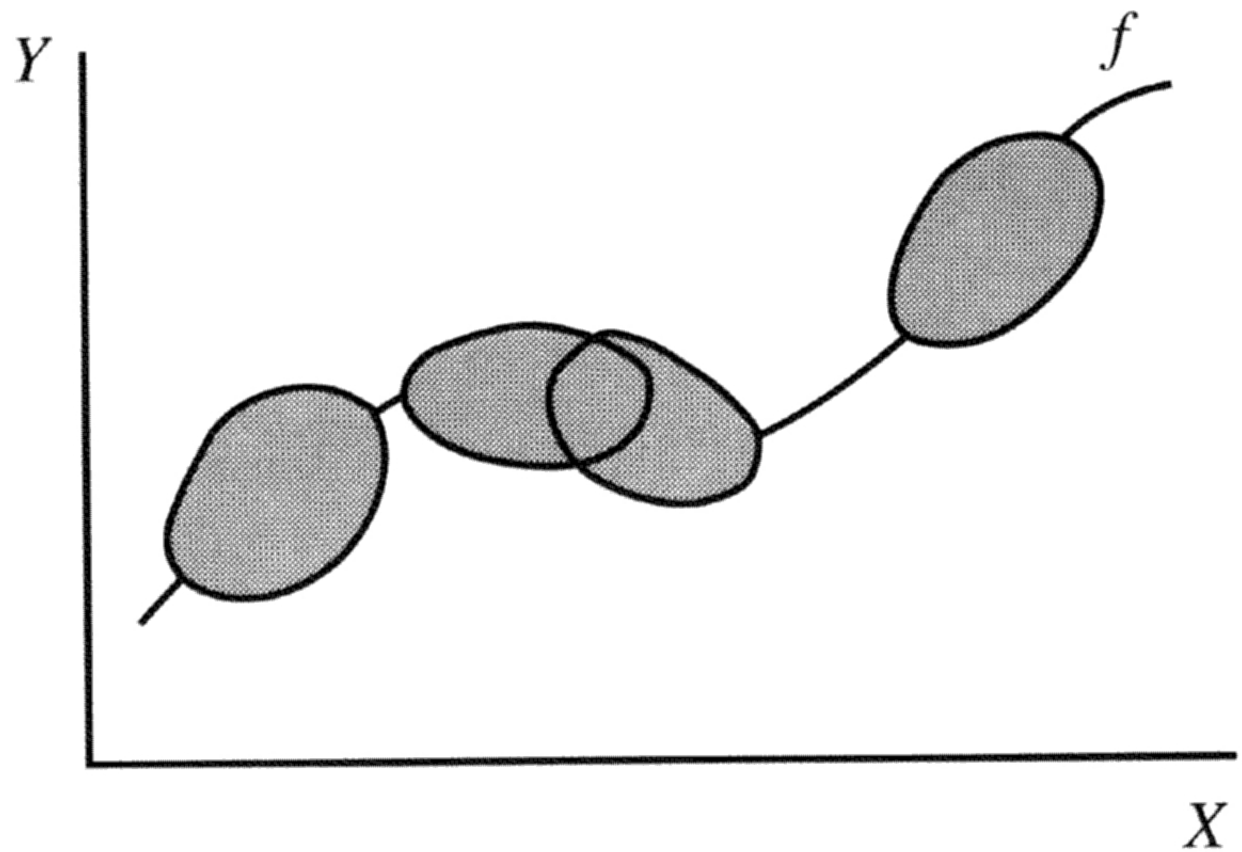
Visualisierung des Eingangs-Ausgangs-Raumes:

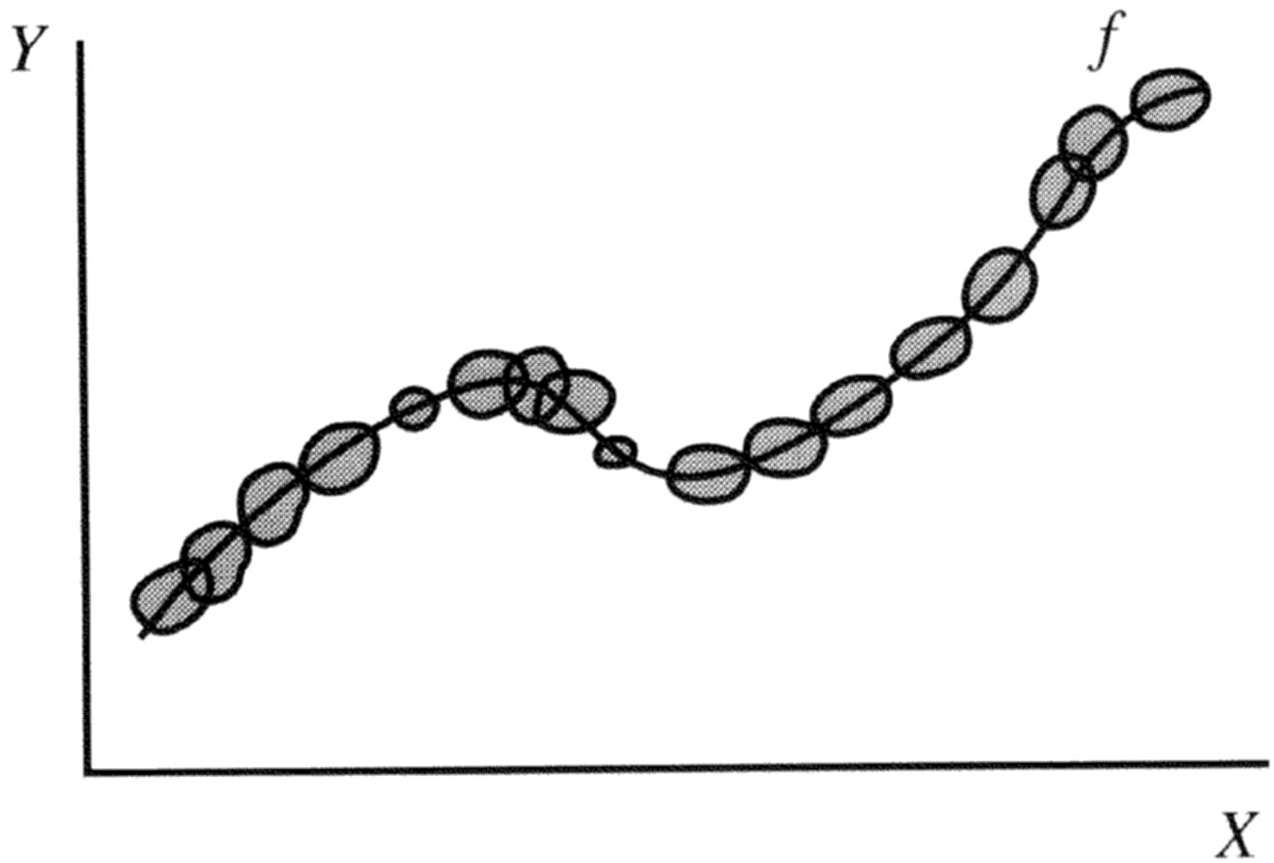


Ein Beispiel der Interpolation:

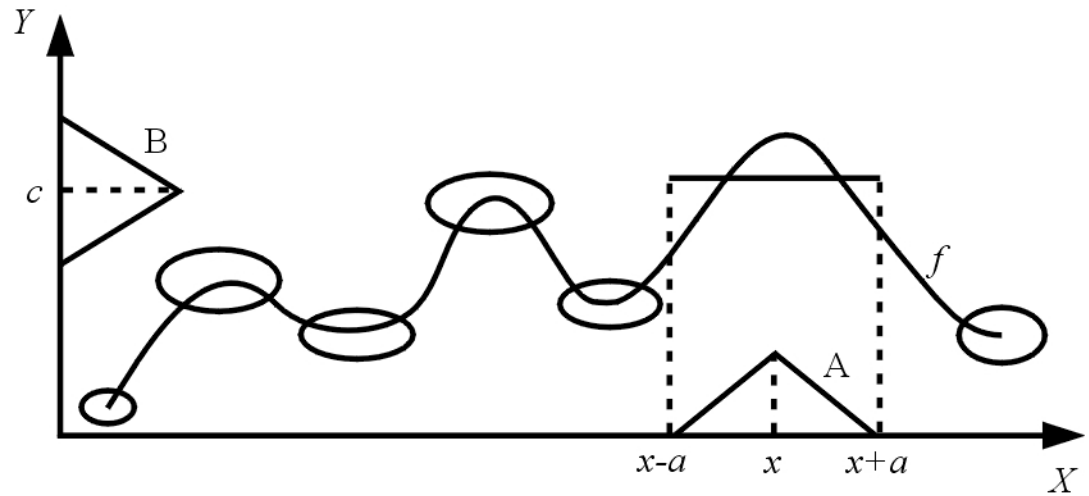


Datencluster entlang der Funktion:



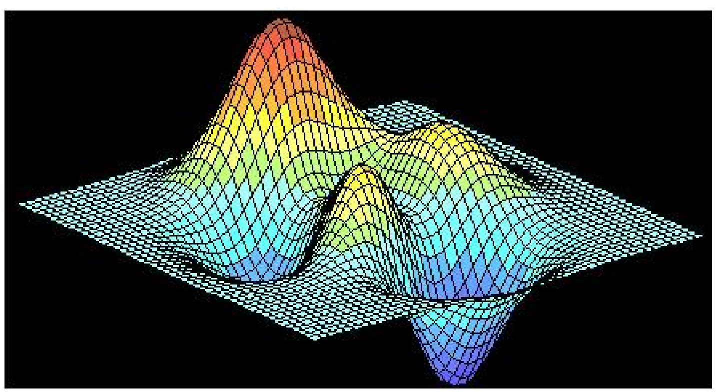

Y
f
X

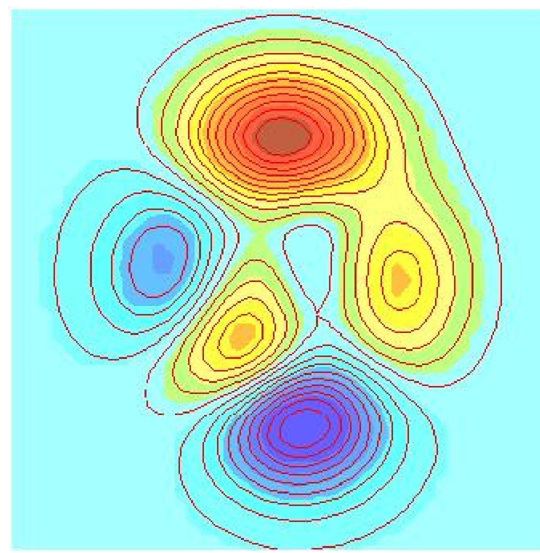
Approximation mit Fuzzy Mengen an den Projektionen der Extremapunkte:



Approximation einer zweidimensionalen Funktion

$$
\begin{aligned}
z =& f(x, y) \\
=& 3(1-x)^2 e^{-x^2-(y+1)^2} - 10(\frac{x}{5} - x^3 - y^5)e^{-x^2-y^2} - \frac{1}{3}e^{-(x+1)^2-y^2}
\end{aligned}
$$





Ableitungen der Funktion:

$$\begin{aligned}\frac{dz}{dx} =& -6(1-x)e^{-x^2-(y+1)^2} - 6(1-x)^2xe^{-x^2-(y+1)^2} \\ & -10(\frac{1}{5}-3x^2) * e^{-x^2-y^2} + 20(\frac{1}{5}x - x^3 - y^5)xe^{-x^2-y^2} \\ & -\frac{1}{3}(-2x-2)e^{-(x+1)^2-y^2}\end{aligned}$$

$$\begin{aligned}\frac{dz}{dy} =& 3(1-x)^2(-2y-2)e^{-x^2-(y+1)^2} \\ & + 50y^4e^{-x^2-y^2} + 20(\frac{1}{5}x - x^3 - y^5)ye^{-x^2-y^2} \\ & + \frac{2}{3}ye^{-(x+1)^2-y^2}\end{aligned}$$

$$
\begin{aligned}
\frac{d\frac{dz}{dx}}{dx} =& 36xe^{-x^2-(y+1)^2} - 18x^2e^{-x^2-(y+1)^2} - 24x^3e^{-x^2-(y+1)^2} + 12x^4e^{-x} \\
&+ 72xe^{-x^2-y^2} - 148x^3e^{-x^2-y^2} - 20y^5e^{-x^2-y^2} + 40x^5e^{-x^2-y^2} + 4 \\
&- \frac{2}{3}e^{-(x+1)^2-y^2} - \frac{4}{3}e^{-(x+1)^2-y^2}x^2 - \frac{8}{3}e^{-(x+1)^2-y^2}x
\end{aligned}
$$

$$\frac{d(\frac{dz}{dy})}{dy} = -6(1-x)^2e^{-x^2-y(+1)^2} + 3(1-x)^2(-2y-2)^2e^{-x^2-(y+1)^2} + 2$$
$$-200y^5e^{-x^2-y^2} + 20(\frac{1}{5}x - x^3 - y^5)e^{-x^2-y^2}$$
$$-40(\frac{1}{5}x - x^3 - y^5)y^2e^{-x^2-y^2} + \frac{2}{3}e^{-(x+1)^2-y^2}$$
$$-\frac{4}{3}y^2e^{-(x+1)^2-y^2}$$

Vereinte Betrachtung aus der statistischen Lerntheorie - I

$\{(\mathbf{x}_i, y_i)\}_{i=1}^{l}$ seien die Daten oder Beispiele. Gesucht wird eine Funktion f, die die folgende Darstellung minimiert:

$$H[f] = \frac{1}{l} \sum_{i=1}^{l} V(y_i, f(\mathbf{x}_i)) + \lambda \|f\|_K^2$$

wobei $V(\cdot, \cdot)$ eine $Verlustfunktion$ und $\|f\|_K^2$ eine Norm im Hilbert-Raum $\mathcal{H}$ ist, welche von einer positiv definierten Kernfunktion K bestimmt wird, und λ der Regularisierungsparameter ist.

Die Probleme in der Modellierung, Datenregression und Musterklassifikation beruhen jeweils auf einer Art von $V(\cdot, \cdot)$.

Vereinte Betrachtung aus der statistischen Lerntheorie - II

1. $V(y_i, f(\mathbf{x}_i)) = (y_i - f(\mathbf{x}_i))^2$

   ($y_i \in R^1$, V: Quadratische Fehlerfunktion)

   $\Rightarrow$ *Regularization Networks*, RN

2. $V(y_i, f(\mathbf{x}_i)) = |y_i - f(\mathbf{x}_i)|_\epsilon$

   ($y_i \in R^1$, $|\cdot|_\epsilon$: eine ϵ-unempfindliche Norm)

   $\Rightarrow$ *Support Vector Machines Regression*, SVMR

3. $V(y_i, f(\mathbf{x}_i)) = |1 - y_i f(\mathbf{x}_i)|_+$

   ($y_i \in \{-1, 1\}$, $|x|_+ = x$ für $x \geqq 0$, sonst $|x|_+ = 0$)

   $\Rightarrow$ *Support Vector Machines Classification*, SVMC

Vereinigte Betrachtung aus der statistischen Lerntheorie - II

1. $V(y_i, f(\mathbf{x}_i)) = (y_i - f(\mathbf{x}_i))^2$

   (y_i: eine reelle Zahl, V: Quadratische Fehlerfunktion)

   $\Rightarrow$ *Regularization Networks*, RN

2. $V(y_i, f(\mathbf{x}_i)) = |y_i - f(\mathbf{x}_i)|_\epsilon$

   (y_i: eine reelle Zahl, $|\cdot|_\epsilon$: eine ϵ-unempfindliche Norm)

   $\Rightarrow$ *Support Vector Machines Regression*, SVMR

3. $V(y_i, f(\mathbf{x}_i)) = |1 - y_i f(\mathbf{x}_i)|_+$

   (y_i: -1 oder 1, $|x|_+ = x$ für $x \geqq 0$, sonst $|x|_+ = 0$)

   $\Rightarrow$ *Support Vector Machines Classification*, SVMC

Für Modellierungs- und Regelungsaufgaben ist die 1. Definition am wichtigsten.

Universelle Funktionsapproximierung - I

Ein Regularisierungsnetzwerk kann alle glatten Funktionen beliebig gut approximieren.

Die allgemeine Lösung dieses Problems ist:

$$f(x) = \sum_{i=1}^{l} c_i K(x; x_i)$$

wobei c_i die Koeffizienten sind.

Universelle Funktionsapproximierung - II

Proposition der Approximation:

Für irgendeine kontinuierliche Funktion Y, die auf einer kompakten Untermenge R^n definiert wird und K eine der Kernfunktionen ist, gibt es eine Funktion $y^*(x) = \sum_{i=1}^{l} c_i K(x; x_i)$, so dass für alle x und irgendein ϵ gilt:

$$|Y(x) - y^*(x)| < \epsilon$$

Universelle Funktionsapproximatoren

Verwendung verschiedener Kernfunktionen führt zu unterschiedlichen Modellen:

$K(\mathbf{x} - \mathbf{x}_i) = \exp(-\lVert\mathbf{x} - \mathbf{x}_i\rVert^2)$	Gaussian RBF
$K(\mathbf{x} - \mathbf{x}_i) = (\lVert\mathbf{x} - \mathbf{x}_i\rVert^2 + c^2)^{-\frac{1}{2}}$	Inverse Multiquadratische Funktione
$K(\mathbf{x} - \mathbf{x}_i) = \tanh(\mathbf{x} \cdot \mathbf{x}_i - \theta)$	Multilagen Perzeptron
$K(\mathbf{x} - \mathbf{x}_i) = (1 + \mathbf{x} \cdot \mathbf{x}_i)^d$	Polynom vom Grad d
$K(x - x_i) = B_{2n+1}(x - x_i)$	B-Splines
$K(x - x_i) = \frac{\sin(d + \frac{1}{2})(x - x_i)}{\sin \frac{(x - x_i)}{2}}$	Trigonometrisches Polynom
...	...

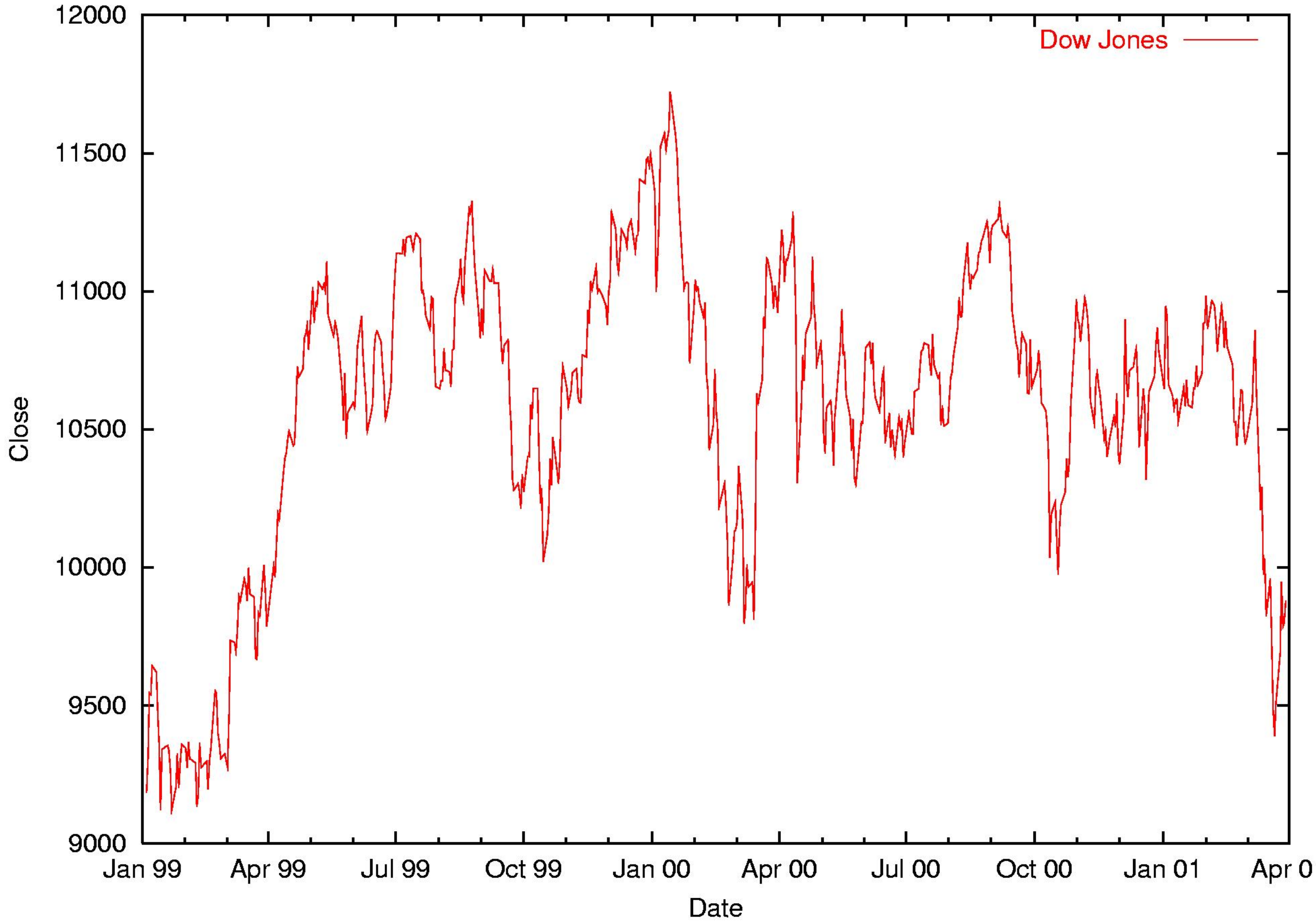
Probleme

1. "Fluch der Dimensionalität" durch exponentielle Abhängigkeit des Speicherbedarfs von der Dimension des Eingaberaumes.

2. *Aliasing* bei der Merkmals-Extrahierung.

3. Nicht vorhandene Soll-Daten (y).

4. Nicht vorhandene Eingangsfaktoren.



Dow Jones
Close
Date
12000
11500
11000
10500
10000
9500
9000
Jan 99
Apr 99
Jul 99
Oct 99
Jan 00
Apr 00
Jul 00
Oct 00
Jan 01
Apr 01