

Kapitel 8

Neuronale Netze

Ansätze zum Entwurf eines Klassifikators

Es gibt zwei prinzipiell unterschiedliche Ansätze zum Entwurf eines Klassifikators:

1. Statistische parametrische Modellierung der Klassenverteilungen, dann MAP
2. Lösung eines Abbildungsproblems durch Funktionsapproximation (nichtlineare Regression)

$$\min E \left\{ \|\mathbf{f}(\mathbf{x}) - \mathbf{y}\|^2 \right\}$$

\mathcal{X} – Merkmalsraum

\mathcal{Y} – Entscheidungsraum

Zu 1.) Die bisher beschriebene Vorgehensweise beim Entwurf eines Klassifikators beruht darauf, dass man die *klassenspezifischen Verteilungsdichten*

$$p(\mathbf{x}|\omega)$$

parametrisch durch statistische Modelle annähert (durch Schätzung der Parameter, z.Bsp. einer Gauß-Verteilung) und durch eine Maximumselektion zu einer Entscheidung kommt. Lernen bedeutet hierbei: Verbesserung des Parameter-Fitting.

Zu 2.) Es gibt nun eine zweite Möglichkeit der Herangehensweise, welche auf die Auswertung der a-posteriori-Wahrscheinlichkeitsdichte

$$p(\omega|\mathbf{x})$$

aufbaut und durch ein Problem der *Funktionsapproximation* beschrieben werden kann.

Diese Funktionsapproximation kann z.Bsp. durch eine *nichtlineare Regression mit Polynomen* oder auch mit Hilfe eines *künstlichen Neuronalen Netzwerkes* (NN) durchgeführt werden. Im Rahmen dieser Veranstaltung sollen die Grundlagen für beide Vorgehensweisen behandelt werden.

Grundsätzlich ist die Suche nach einer besten Näherungsfunktion ein *Variationsproblem*, welches durch die Wahl von Basisfunktionen auf ein *parametrisches Optimierungsproblem* zurückgeführt werden kann. Lernen bedeutet dann auch hier: Parameterfitting.

Die Gleichwertigkeit der beiden genannten Vorgehensweisen ergibt sich aus dem Bayes-Theorem:

$$p(\omega | \mathbf{x}) = \frac{p(\mathbf{x} | \omega) p(\omega)}{p(\mathbf{x})}$$

$\xrightarrow{1/K}$ (über $p(\omega)$)
 $\xrightarrow{\text{unabhängig von } \omega}$ (unter $p(\mathbf{x})$)

Die Gleichwertigkeit ergibt sich daraus, dass der Nenner unabhängig von ω ist und die a-priori-W. i.a. $1/K$ gesetzt werden kann.

Im folgenden soll nun die Überführung in ein Funktionsapproximationsproblem begründet werden.

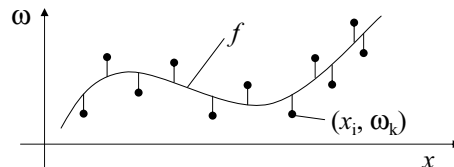
Bei bekannter a-posteriori-W. $p(\omega|\mathbf{x})$ würde für jedes kontinuierliche \mathbf{x} bestmöglich ein ω zugeordnet werde (funktionale Zuordnung $f: \mathbf{x} \rightarrow \omega$). Gegeben sind jedoch nur *Stichproben* und man sucht nach einer Funktion f , welche bestmöglich die Einzelexperimente fittet und damit eine Abbildung realisiert:

$$f: \underset{\mathbf{x}}{\text{Merkmalsraum}} \rightarrow \underset{\omega}{\text{Bedeutungsraum}}$$

Diese Aufgabenstellung kann mit Hilfe der Variationsrechnung gelöst werden.

Wählt man als Gütekriterium das minimale Fehlerquadrat, so geht es um die Minimierung von:

$$J = \min_{f(\mathbf{x})} E\{\|\mathbf{f}(\mathbf{x}) - \mathbf{y}\|^2\}$$



Dabei entstehen die Zielvektoren $\{\mathbf{y}_i\}$ im *Entscheidungsraum* \mathcal{Y} durch einfache Abbildung der skalaren Bedeutungswerte $\{\omega_i\}$

$$\Omega := \{\omega_1, \omega_2, \dots, \omega_K\}$$

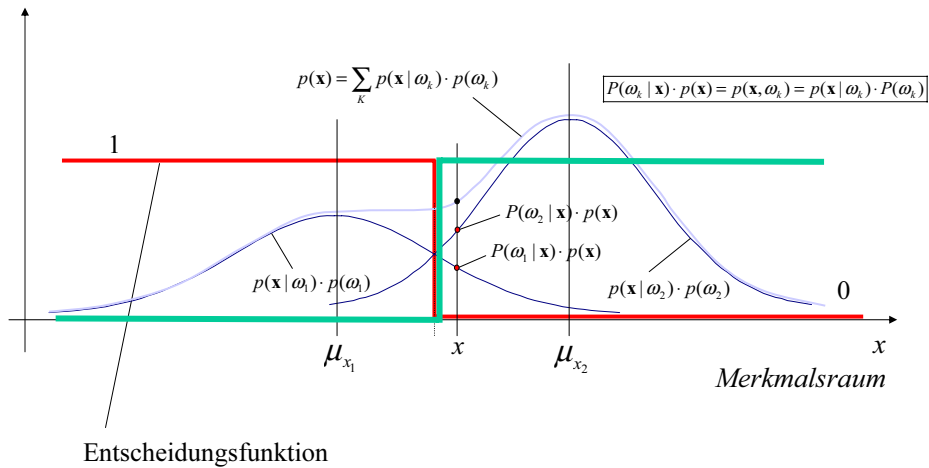


$$\mathcal{Y} := \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_K\}$$

mit:

$$\mathbf{y}_i = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \text{i-ter Einheitsvektor}$$

Zwei-Klassen-Problem mit Gaußverteiligdichte



Regression mit Hilfe *künstlicher* Neuronaler Netze

Verwendung eines mehrschichtigen Perceptrons zur Funktionsapproximation (multilayer perceptron)

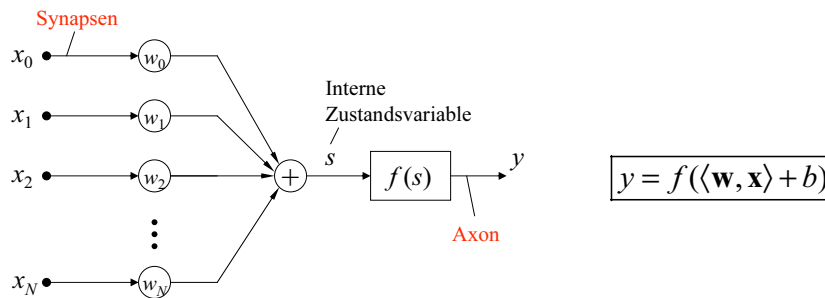
Motivation:

- Anlehnung an menschliche Vorgehensweise (?)
- Parallele Auswertung mit NN-Rechner (Hardware)

Menschliches Gehirn: 10^{11} Neuronen mit bis zu 10^4 Verbindungen pro Neuron

Modell eines Neurons

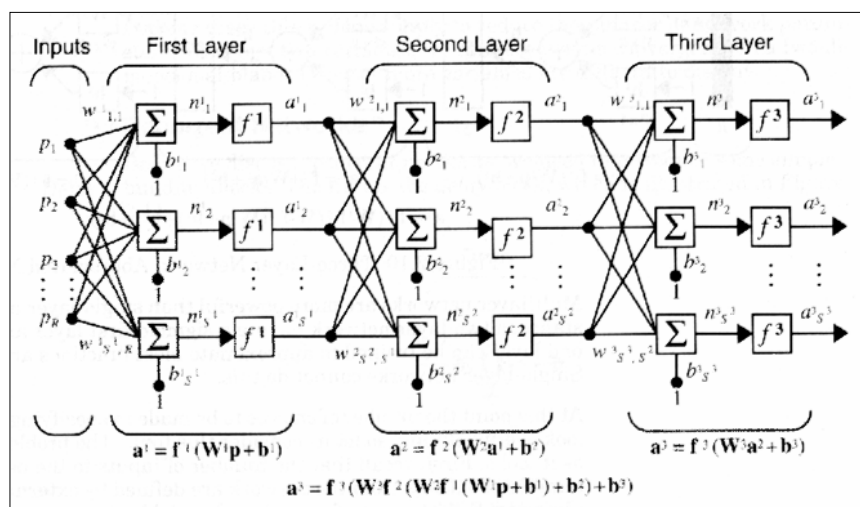
(McCulloch & Pitts, 1943)



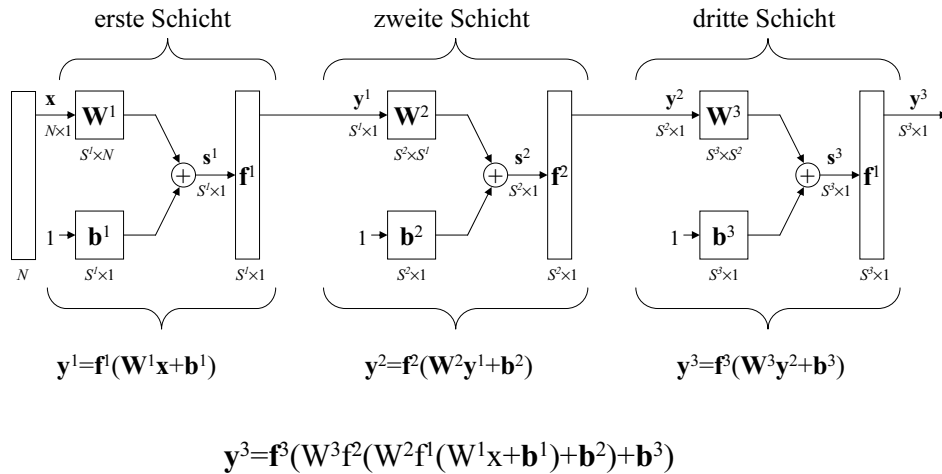
Die neuronale Aktivität wird durch ein Skalarprodukt zwischen dem Gewichtsvektor \mathbf{w} und den Eingangskanälen x_i mit einer sich anschließenden nichtlinearen Aktivierungsfunktion $f(s)$ beschrieben. Dabei können die Erregungen x_i verstärkend oder hemmend wirken, was zu positiven oder negativen Gewichtswerten w_i korrespondiert.

Das Multilagen-Perceptron mit 3 Schichten

(Rosenblatt 1958)



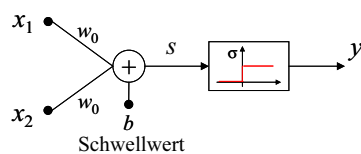
Symbolische Darstellung eines Perceptrons mit 3 Schichten



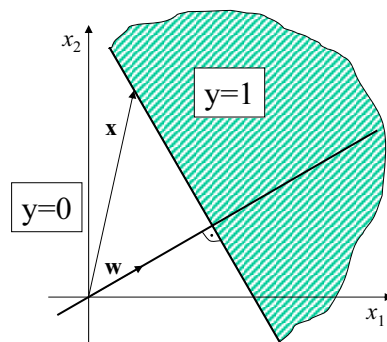
Das Perceptron mit einem Neuron und zwei Eingängen (Rosenblatt, 1958)

In den Anfangsjahren der NN-Forschung benutzte man vorzugsweise Schwellwertlogiken. D.h. der Eingangsvektor war *zweiwertig* und als Aktivierungsfunktion wurde die *Sprungfunktion* ($f(s) = \sigma(s) = 1$ für $s \geq 0$ und $\sigma(s) = 0$ für $s < 0$).

Da jede Boole'sche Funktion als disjunktive oder konjunktive Normalform geschrieben werden kann, ist es möglich, mit einem *zweischichtigen* Perceptron *jede logische Funktion* zu realisieren.



$$y = \sigma(\langle \mathbf{w}, \mathbf{x} \rangle + b) = \sigma(g(\mathbf{x}))$$



Funktionsweise bei reellwertigen Eingängen

Zur Funktionsweise des Perceptrons

Das Neuron unterteilt den Eingansraum von \mathbf{x} mit einer *Hyperebene* (hier: *Gerade*) in zwei Hälften. Die Hyperebene ist der *geometrische Ort*, auf dem alle Vektoren liegen, deren Projektion auf \mathbf{w} konstant ist:

$$g(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle + b = 0$$

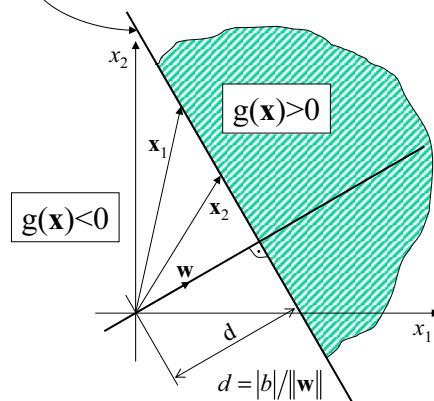
$$\langle \mathbf{w}, \mathbf{x} \rangle = \|\mathbf{w}\| \|\mathbf{x}\| \cos \varphi$$

Für zwei Punkte auf der Trennfläche gilt:

$$0 = \langle \mathbf{w}, \mathbf{x}_1 \rangle + b = \langle \mathbf{w}, \mathbf{x}_2 \rangle + b$$

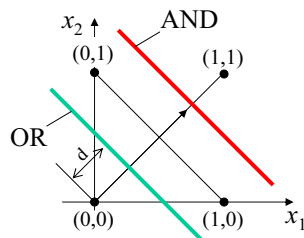
$$\Rightarrow \langle \mathbf{w}, (\mathbf{x}_1 - \mathbf{x}_2) \rangle = 0$$

$$\Rightarrow \mathbf{w} \perp (\mathbf{x}_1 - \mathbf{x}_2)$$



Nichtlinearer Klassifikatorentwurf - das XOR-Problem

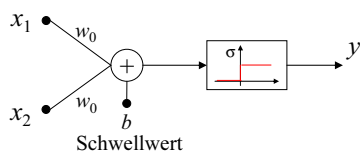
Es ist offensichtlich, dass sich die Booleschen Funktionen AND und OR durch einen linearen Klassifikator und damit mit einem einschichtigen Perceptron lösen lassen:



x_1	x_2	AND	Klasse	OR	Klasse
0	0	0	ω_1	0	ω_1
0	1	0	ω_1	1	ω_2
1	0	0	ω_1	1	ω_2
1	1	1	ω_2	1	ω_2

$$\mathbf{w} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

AND	$b = \frac{3}{4}\sqrt{2}$
OR	$b = \frac{1}{4}\sqrt{2}$



oder auch vereinfacht:

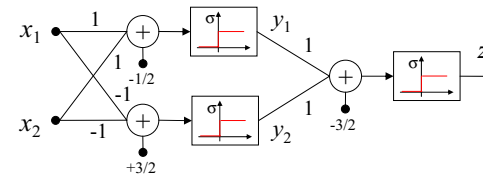
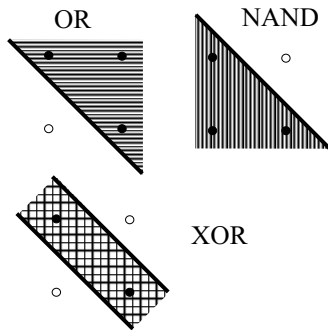
$$\text{AND: } g(\mathbf{x}) = x_1 + x_2 - \frac{3}{2} = 0$$

$$\text{OR: } g(\mathbf{x}) = x_1 + x_2 - \frac{1}{2} = 0$$

Lösung des XOR-Problems mit einem Zwei-Lagen-Perceptron

Das Exklusiv-Oder-Problem lässt sich hingegen im Gegensatz zu AND und OR *nicht* durch einen linearen Klassifikator lösen.

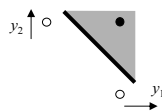
Eine Lösung ergibt sich jedoch durch Kombination zweier Trennlinien. Dies führt auf ein zweilagiges Perceptron!



		erste Schicht		zweite Schicht	
x_1	x_2	y_1	y_2	XOR	Klasse
0	0	0	1	0	ω_1
0	1	1	1	1	ω_2
1	0	1	1	1	ω_2
1	1	1	0	0	ω_1

$$z = (x_1 \vee x_2) \wedge \overline{(x_1 \wedge x_2)} = y_1 \wedge y_2 \quad \text{XOR}$$

Nach der Abbildung der ersten Schicht ergibt sich ein linear trennbares Problem!!



$$y_1 = \sigma(x_1 + x_2 - \frac{1}{2}) \quad \text{OR}$$

$$y_2 = \sigma(-x_1 - x_2 + \frac{3}{2}) \quad \text{NAND}$$

$$z = \sigma(y_1 + y_2 - \frac{3}{2}) \quad \text{AND}$$