Statistische und neuronale Lernverfahren

Martin Stetter

WS 05/06, 2 SWS

Bereich:
prüfbare Vorlesung im Bereich praktischer und theoretischer Informatik:
Künstliche Intelligenz / Maschinelles Lernen

Zeit: Dienstag 8.30 (s.t.) - 10.00

Ort: 00.08.038

Beginn: 25.10.2005

Behandelte Themen

0. „Motivation“: Lernen in Statistik und Biologie

1. Überblick über statistische Datenmodellierungs-Verfahren

2. Das lineare Modell (Regression)

3. Perceptron und Multilagen-Perceptron (Funktionsapproximation)

4. Selbstorganisierende Merkmalskarten (Dichteschätzung)

5. Lernen von Datenmodellen

- Approximation: Bayes‘sches Schließen, MAP, ML
- Maximum Likelihood Schätzung und Fehlerminimierung
- Generalisierung und Regularisierung
- Optimierungsverfahren

6. Bayesianische Netze (Dichteschätzung und Funktionsapproximation)

7. Kern-Trick und Support Vector Machine

Literatur

Statistische Lernverfahren:

- B. Schölkopf, A. Smola: *Learning with Kernels.* MIT Press, Cambridge, MA (2002)
- S. Lauritzen. *Graphical Models*, Oxford Univ. Press (1996)
- Kevin Murphy´s tutorial. *A Brief Introduction to Graphical Models and Bayesian Networks* http://www.ai.mit.edu/~murphyk/Bayes/bayes.html

Statistische und neuronale Verfahren:

- C. M. Bishop, *Neural Networks for Pattern Recognition*, Clarendon Press Oxford (1995)

Neuronale Netze:

- J. Hertz, A. Krogh, R. G. Palmer, *Introduction to the Theory of Neural Computation*, Addison Wesley, Redwood City CA (1991)

Computational Neuroscience und Bioinformatik

- M. Stetter, *Exploration of Cortical Function*, Kluwer Academic Publishers, Dordrecht (2002)
- P. Baldi, G. W. Hatfield, *DNA Microarrays and Gene Expression*, Cambridge University Press Cambridge, MA (2002)

Einführung: Lernen in Statistik und Biologie

Biologisches Lernen	Maschinelles Lernen
Erkennen von Zusammenhängen im Lebensraum	Erkennen von statistischer Struktur in Datensätzen
„Synaptische Plastizität“ im Gehirn	Einstellung der Modellparameter
Erlernen von Fähigkeiten aus Beispielen (prozedurales Lernen)	Überwachtes Lernen
Erlernen von Fähigkeiten durch Probieren	Unüberwachtes Lernen
Finden einfacher Lösungen	Regularisierung
Schlußfolgern	Bayes‘sches Schließen (Inferenz)

Statistisches Lernen: Beispiel

Statistisches / Maschinelles Lernen: „Entdeckung von Struktur in Daten“

Beispiel: 2D-Klassifikation:

M **Fiktive Daten eines Geldinstituts:**

x_1 **= Mittl. abgehobener Geldbetrag**

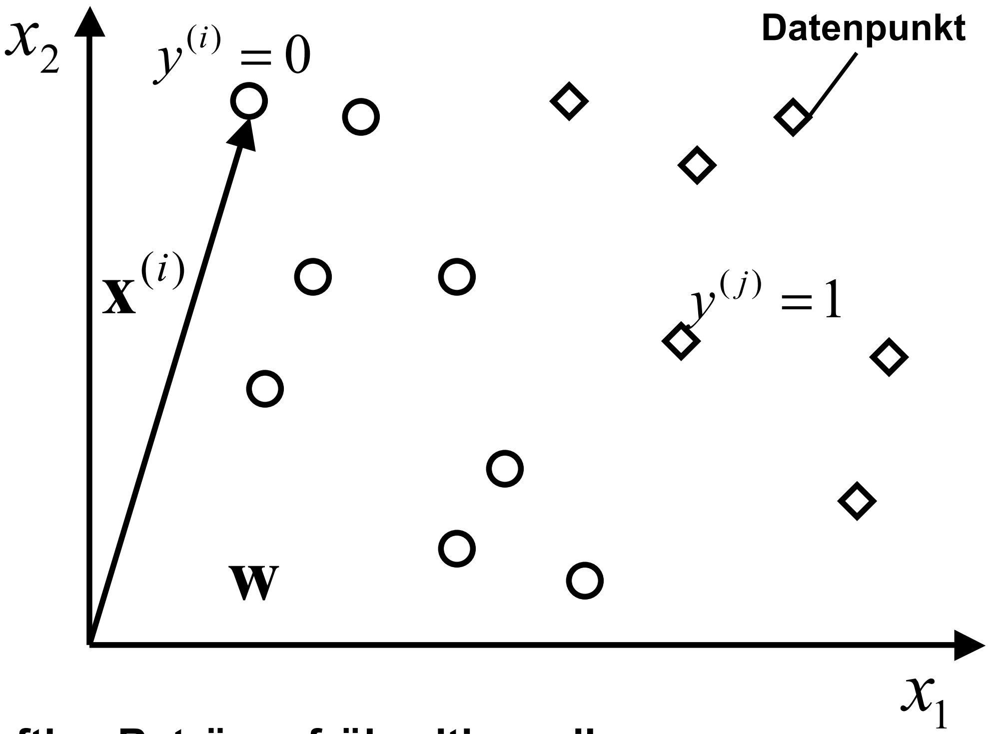
x_2 **= Häufigkeit Abhebevorgänge**

$\mathbf{x} = (x_1, x_2)$ = **„Muster“**

y **= Scheckkartenbetrug (1=ja, 0=nein)**

y **= „Klassenlabel“, Soll, Output**

$(\mathbf{x}^{(1)}, y^{(1)}) \ldots (\mathbf{x}^{(M)}, y^{(M)})$



Ziel: **Lerne Klassifikationsregel, um künftige Betrüger frühzeitig an ihrem Verhalten zu erkennen.**

Lernen eines Klassifikators:

Linearer Klassifikator:

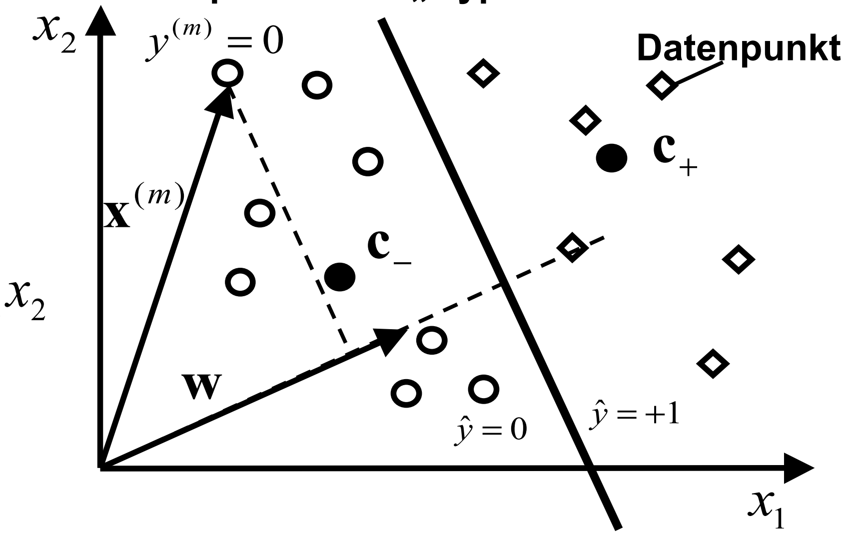
$$\hat{y} = \Theta(\mathbf{w} \cdot \mathbf{x} + b)$$

$$\Theta(x) = 1, x \geq 0$$
$$\Theta(x) = 0, x < 0$$
$$\mathbf{w} \cdot \mathbf{x} := w_1 x_1 + w_2 x_2$$

$(\mathbf{w}, b)$ **= Parametersatz, „Modell“, „Hypothese“**

„Lernen“: Finde besten Parametersatz

Separierende „Hyperebene“ $\mathbf{w} \cdot \mathbf{x} + b = 0$



Offline-Lernen:

$$\mathbf{c}_{+/-} = \frac{1}{M_{+/-}} \sum_{\{m | y^{(m)} = \pm 1\}} \mathbf{x}^{(m)}$$ **Klassenzentren**

$$\hat{\mathbf{w}} = \mathbf{c}_+ - \mathbf{c}_-$$

$$\hat{b} = -\mathbf{w} \cdot (\mathbf{c}_+ + \mathbf{c}_-) / 2$$

Online-Lernen (Beispiel für $b=0$):

Für jeden Datenpunkt ändere $\mathbf{w}$ gemäß

$$\Delta \mathbf{w} = \eta (y^{(m)} - \hat{y}^{(m)}) \mathbf{x}^{(m)}$$

(Perceptron-Lernregel, siehe später)

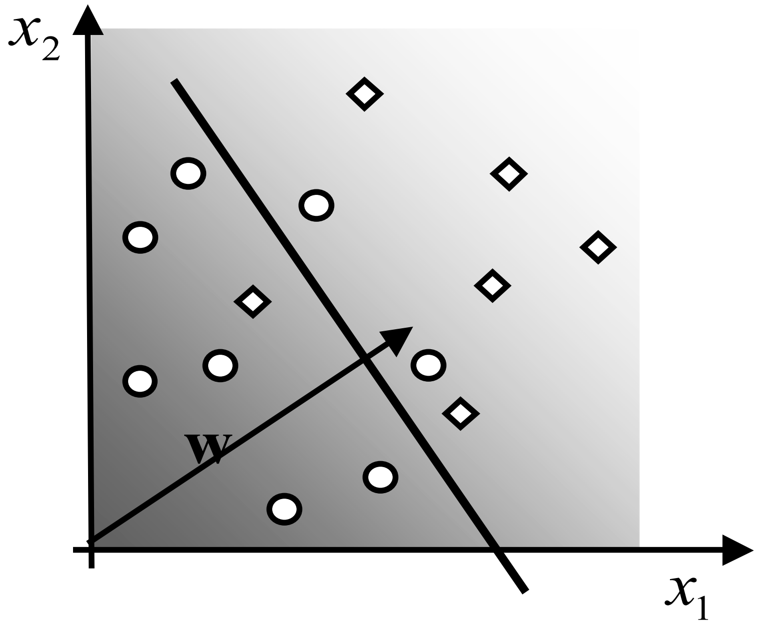
Warum „statistisches“ Lernen?

Daten sind unsicher:

-- Datenpunkte $\mathbf{x}$ könnten versetzt sein
-- Gemessene Klassen können falsch sein

Klassifikationsgesetz ist unsicher

-- Mitglieder unterschiedlicher Klassen könnten dasselbe Muster $\mathbf{x}$ aufweisen



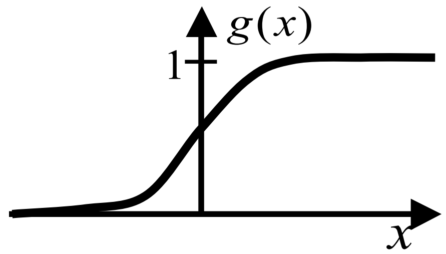
Lösung: Probabilistischer Klassifikator („soft classifier“)

Spezifiziere Wahrscheinlichkeit für Klassenmitgliedschaft

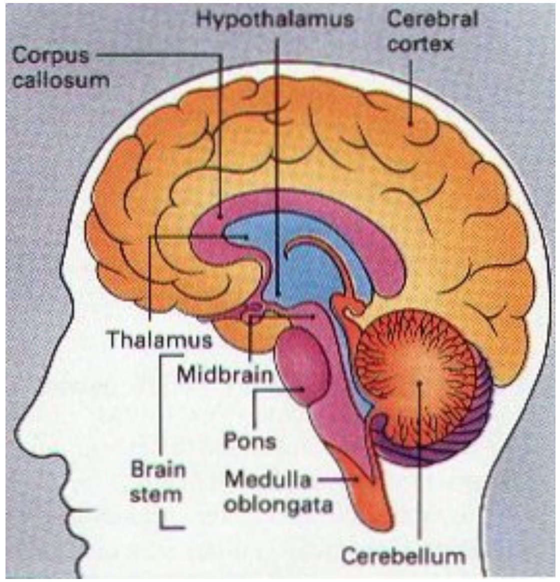
$$\Pr(y(\mathbf{x})=1)=g(\mathbf{w}\cdot\mathbf{x}+b)$$

Beispiel: Logistische Transferfunktion :

$$g(x)=\frac{1}{1+\exp(-x)}$$



Neuronales Lernen: Gehirn und Nervenzelle



<u>Das menschliche Gehirn</u>

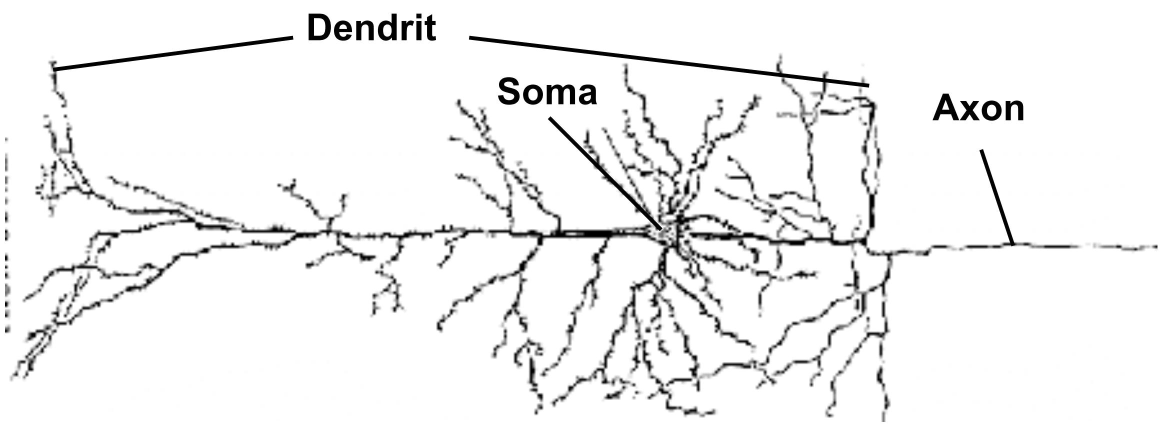
Besteht aus Nervenzellen (Neuronen) und Hilfszellen

Hochstrukturiert (Kerne, Areale der Grosshirnrinde)

Gigantisches Netz aus Neuronen:
-- 100 000 000 000 Nervenzellen
-- Jede Zelle erhält synaptischen Input von ca. 10 000 anderen Nervenzellen (Konvergenz)
-- Jede Zelle sendet ca. 10 000 Outputs (Divergenz)
-- Gesamte Leitungslänge: 750 000 km !!!

<u>Nervenzelle</u>

Besteht aus Dendrit,
Soma (Zellkörper)
Axon



Neuronales Lernen: Reizleitung in Neuronen

<u>Funktionsweise des Neurons</u>

Signal: Aktionspotential, „Spike“

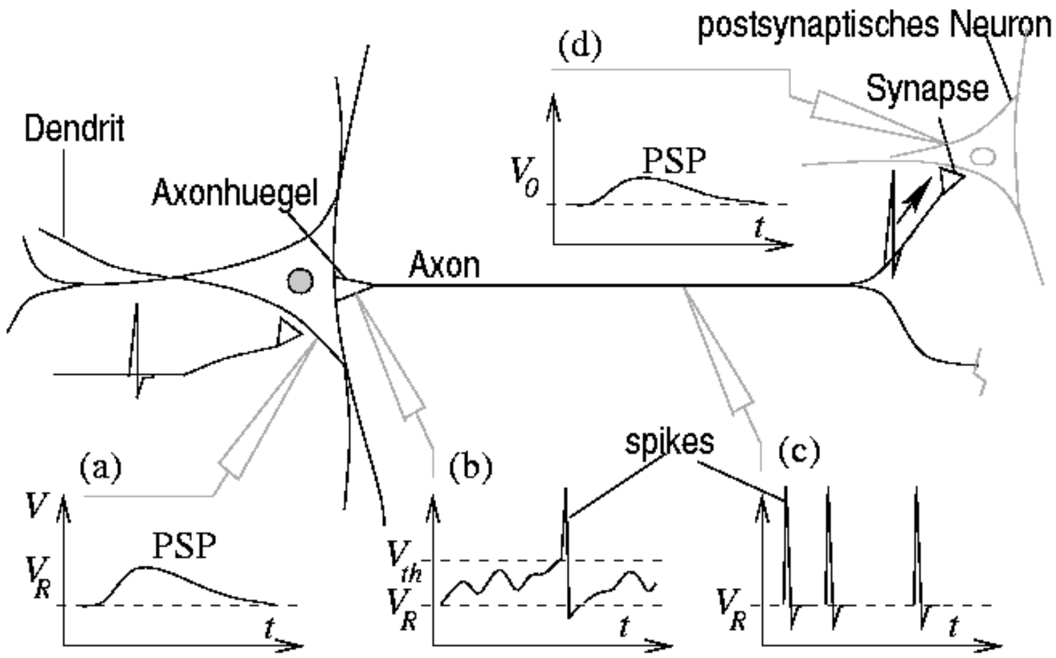
Signalfluss:
Dendrit --> Soma --> Axon-->
Synapse--> Dendrit ...

(a) Spike kommt an;
Synapse injiziert Strom I
Membranspannung steigt (PSP)

(b) Viele PSPs summieren sich
Bei Schwellenspannung: Spike

(c) Spike läuft Axon entlang
verzweigt sich mit dem Axon

(d) Spike kommt an.....



<u>Biologisches Lernen (Hypothese):</u>

Synaptischer Strom I ändert sich in Abhängigkeit von der Zeit und der Hirnaktivität („LTP, LTD“.....)

Ratenmodell des Neurons

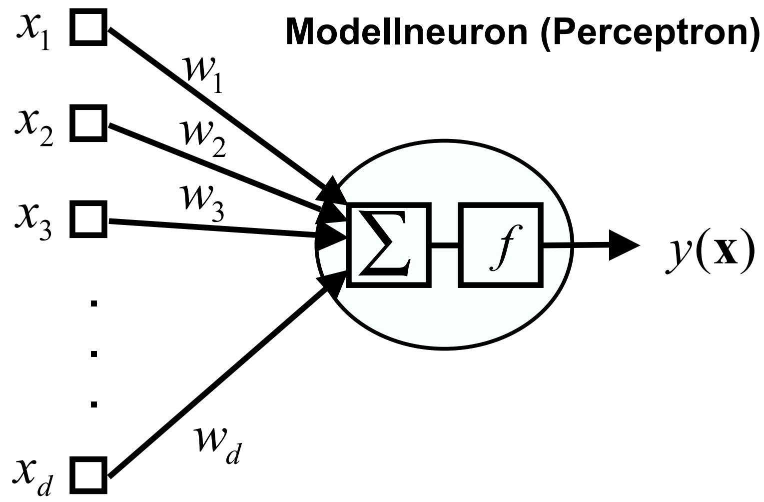
- **Neuron erhält Signale von d Synapsen**
- **An Synapse i kommen Spikes mit der Rate x_i an**
- **Synapse i induziert Spannung $U_i = w_i x_i$**
  w_i heißt synaptisches Gewicht
- **Das Soma summiert die Spannungs-änderungen:**

$$U = \sum_{i=1}^{d} w_i x_i = \mathbf{w} \cdot \mathbf{x}$$

- **Die Spikerate y am Axon ist eine sigmoide Funktion der Summenspannung**

$$y(\mathbf{x}) = g(\mathbf{w} \cdot \mathbf{x} - \theta)$$

Modellneuron (Perceptron)



Lernen im Modellneuron:

Synaptische Gewichte $\mathbf{w}$ ändern sich abhängig von der Aktivität

Biologisch motivierte Lernregeln... (zB. „Hebb-Regel“)

Statistische und Neuronale Lernverfahren: Beispiele

Künstliche Neuronale Netze

Modellneuronen können zu künstlichen neuronalen Netzen zusammengeschaltet werden, zB.

Statistische Datenmodellierung

Modellneuronen können auch als statistische Datenmodelle interpretiert werden, z.B.

Perceptron

Multilagen-Perceptron

Radiale Basisfunktionen-Netzwerk

Hopfield-Netzwerk

Adaptive Resonance Theory Netzwerk

Selbstorganisierende Merkmalskarten

Probabilistischer Klassifikator

Support-Vector Machine

Bayes-Belief-Netzwerk

Helmholtz-Machine

Vektor-Quantisierer

ICA

Zusammenfassung

- **Ein biologisch motiviertes Modellneuron läßt sich mit einem linearen probabilistischen Klassifikator identifizieren**
- **Viele künstliche Neuronale Netze lassen sich mit statistischen Lernverfahren identifizieren**
- **Viele Verfahren, viele Aufgaben**

„Road Map“

- **Bayes‘sche Inferenz als genereller Rahmen für statistische Lernverfahren**
- **Statistische Datenmodellierung durch Optimierung und Regularisierung**
- **Spezielle maschinelle Lernverfahren und Neuronale Netzwerk-Typen**

Parallelveranstaltung Unüberwachte Lernverfahren

PD Dr. Thomas Runkler: Data Mining und Knowledge Discovery

Bereich: prüfbare Vorlesung im Bereich 1.4 Künstliche Intelligenz / Maschinelles Lernen

Zeit: Montag 8.30 - 10.00
Ort: Raum 00.13.009A
Beginn: 24.10.05

Inhalt:
Einführung: Ziel, Definitionen, Schritte der Knowledge Discovery (KDD)
Datenquellen, -charakteristika und Fehlerquellen,
Datenvorverarbeitung und -filterung
Datenvisualisierung; Projektionen, Hauptachsentrafo., mehrdim. Skalierung, Sammon-Methode, selbstorg. Karten
Datentransformationen und Merkmalsgenerierung
Datenanalyse; Korrelationsanalyse und Scheinkorrelationen, Regression, Klassifikation (Entscheidungsbäume, ID3)
Clustering (SAHN, HCM, FCM, GK, FCL, FCE, ACE, RBF, LVQ, FGLVQ), Regeln (scharf, unscharf, lokale Modelle)
Anwendungsbeispiele