

Maschinelles Lernen WS 2008/09

Benno Stein

Raum: 110b, B11
Tel: 3795
URL: webis.de
E-Mail: benno.stein@medien.uni-weimar.de
Sprechstunde: nach Vereinbarung

Inhalt

- I. Einführung
- II. Grundlagen des Maschinellen Lernens
- III. Entscheidungsbäume
- IV. Neuronale Netze
- V. Statistische Lernverfahren
- VI. Fortgeschrittene Lernverfahren
- VII. Lerntheorie
- VIII. Naturanaloge Verfahren
- IX. Reinforcement-Lernen
- X. Cluster-Analyse
- XI. Assoziationsanalyse

Übungen

Organisation:

- Übungsaufgaben werden im Netz zur Verfügung gestellt
- bei praktischen Aufgaben: Teams (2 bis 3 Studenten/innen) möglich

Aufgaben:

- theoretisch
- praktisch

Beginn der Übungen: Ende Oktober

Ziele

- ❑ Grundbegriffe des Maschinellen Lernens verstehen
- ❑ formale Beschreibungen von Verfahren verstehen und in Algorithmen umsetzen können
- ❑ einschlägige Algorithmen sinnvoll anwenden
- ❑ sich selbst weiterbilden können

Angrenzende Gebiete

1. Statistik

[Paradigmen, Modelle]

2. Algebra

3. Information Retrieval

[Techniken, Algorithmen]

4. Wissensverarbeitung

5. Heuristische Suche

6. Entscheidungsunterstützungssysteme

[Anwendungen]

7. Business Intelligence

8. Web-Technologie

Literatur

Maschinelles Lernen:

- ❑ Tom Mitchell.
Machine Learning
McGraw Hill 1997, ISBN 0070428077. www.cs.cmu.edu/~tom/mlbook.html
- ❑ Leo Breiman, Jerome H. Friedman, Richard A. Olshen, Charles J. Stone.
Classification and Regression Trees
- ❑ Trevor Hastie, Robert Tibshirani, Jerome Friedman.
The Elements of Statistical Learning
Springer 2001, ISBN 0-387-95284-5.
- ❑ Vladimir Vapnik.
The Nature of Statistical Learning Theory
Springer 2000, ISBN 387-98780-0.

Literatur

Data Mining:

- ❑ David Hand, Heikki Mannila, Padhraic Smyth.
Principles of Data Mining
Massachusetts Institute of Technology, 2001, ISBN 0-262-08290-X.
- ❑ Tan, Steinbach, Kumar.
Introduction to Data Mining
Addison Wesley 2005, ISBN 0-321-32136-7.
- ❑ Ian H. Witten, Eibe Frank.
Data Mining: Practical Machine Learning Tools and Techniques
Second Edition, Morgan Kaufmann 2005, ISBN 0-12-088407-0

Software

Entwicklungsumgebung:

- Borland, IBM, MERANT, QNX, Rational Software, Red Hat, SuSE, TogetherSoft, Webgain, Ericsson, HP, Intel, MontaVista Software, SAP, Serena Software, Actuate, et al.

Eclipse SDK

Version 3.2.2 <http://www.eclipse.org/downloads/>

Machine Learning Library:

- Eibe Frank, Mark Hall, Geoff Holmes, Mike Mayo, Bernhard Pfahringer, Tony Smith, Ian Witten.

Weka Machine Learning Project

Version 3.5.6 <http://www.cs.waikato.ac.nz/ml/weka/>

Kapitel ML: I

I. Einführung

- Beispiele für Lernaufgaben
- Spezifikation von Lernproblemen

Beispiele für Lernaufgaben

Autoeinkaufsberater



Welche Kriterien liegen einer Entscheidung zugrunde?

Beispiele für Lernaufgaben

Risikoanalyse bei Kreditvergabe

Kunde 1	
Hausbesitzer	ja
Einkommen (p.a.)	51.000 EUR
Tilgung (p.m.)	1.000 EUR
Kreditlaufzeit	7 Jahre
SCHUFA-Eintrag	nein
Alter	37
verheiratet	ja
...	

...

Kunde n	
Hausbesitzer	nein
Einkommen (p.a.)	55.000 EUR
Tilgung (p.m.)	1.200 EUR
Kreditlaufzeit	8 Jahre
SCHUFA-Eintrag	nein
Alter	?
verheiratet	nein
...	

Beispiele für Lernaufgaben

Risikoanalyse bei Kreditvergabe

Kunde 1	
Hausbesitzer	ja
Einkommen (p.a.)	51.000 EUR
Tilgung (p.m.)	1.000 EUR
Kreditlaufzeit	7 Jahre
SCHUFA-Eintrag	nein
Alter	37
verheiratet	ja
...	

...

Kunde n	
Hausbesitzer	nein
Einkommen (p.a.)	55.000 EUR
Tilgung (p.m.)	1.200 EUR
Kreditlaufzeit	8 Jahre
SCHUFA-Eintrag	nein
Alter	?
verheiratet	nein
...	

Gelernte Regeln:

IF (Einkommen>40.000 **AND** Kreditlaufzeit<3)

OR Hausbesitzer=ja

THEN Kreditvergabe=ja

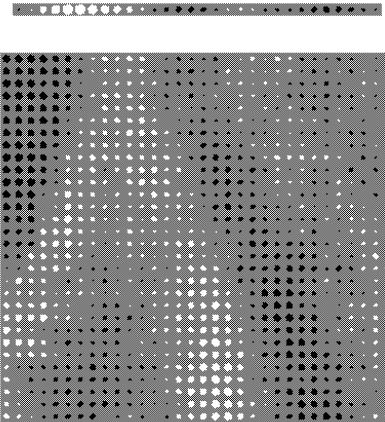
IF SCHUFA-Eintrag=ja **OR**

(Einkommen<20.000) **AND** Tilgung>800)

THEN Kreditvergabe=nein

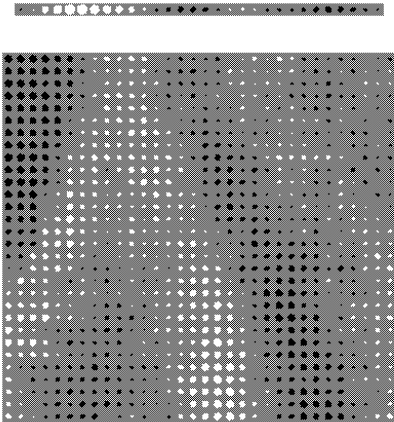
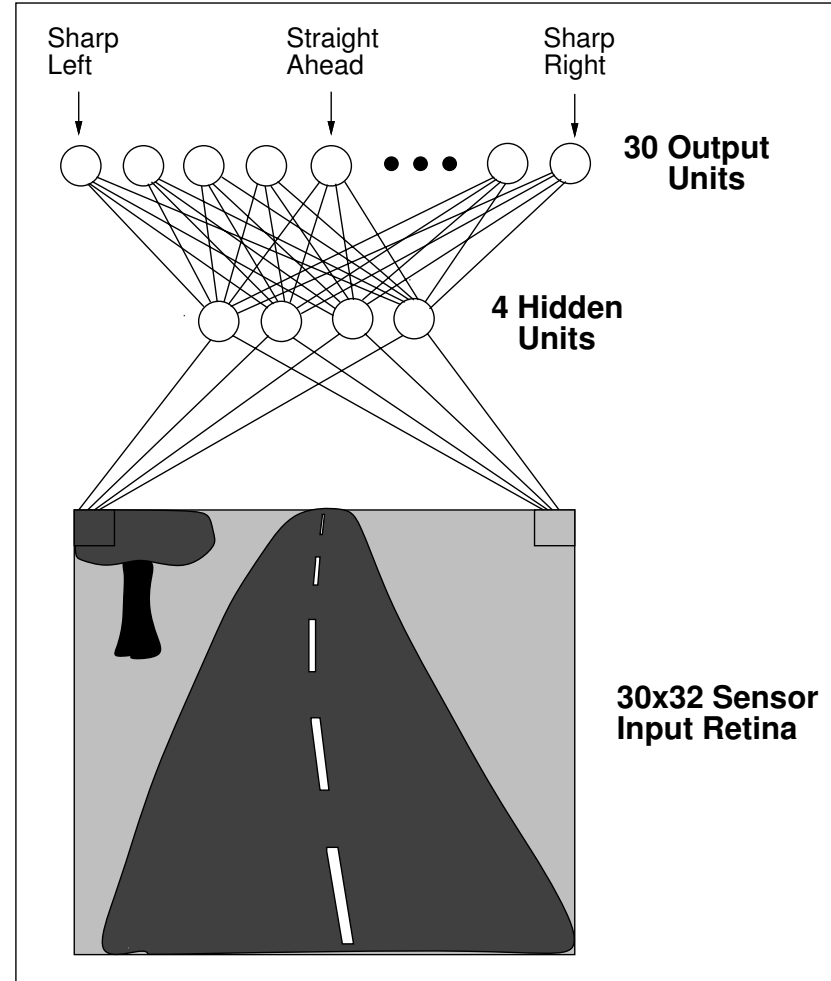
Beispiele für Lernaufgaben

Bildanalyse



Beispiele für Lernaufgaben

Bildanalyse



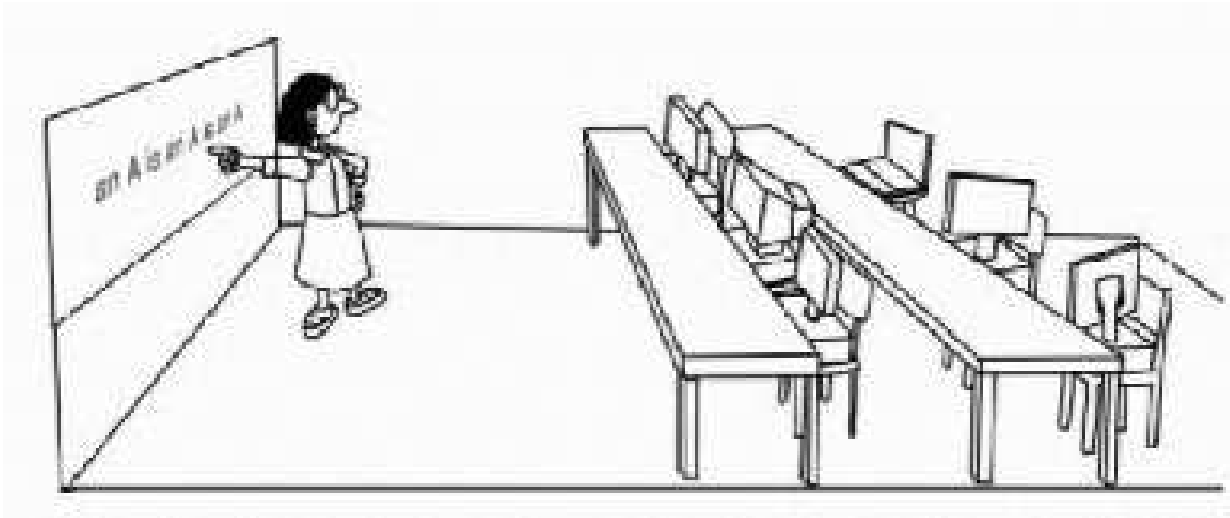
Spezifikation von Lernproblemen

Definition 1 (Maschinelles Lernen, Machine Learning) [Mitchell 1997]

A computer program is said to learn

- from experience (Erfahrung)
- with respect to some class of tasks (Aufgaben) and
- a performance measure (Gütemaß),

if its performance at the tasks improves with the experience.



Bemerkungen:

- Beispiel Schach.
 - Aufgabe = Schachspielen
 - Gütemaß = Anteil gewonnener Spiele bei einer Weltmeisterschaft
 - Erfahrung = Möglichkeit, gegen sich selbst zu spielen

- Beispiel Schrifterkennung.
 - Aufgabe = Isolation und Klassifikation handgeschriebener Worte in Bitmaps
 - Gütemaß = Anteil korrekt klassifizierter Worte
 - Erfahrung = Kollektion mit korrekt klassifizierten, handgeschriebenen Worten

Spezifikation von Lernproblemen

Lernparadigmen

1. Überwachtes Lernen (*Supervised Learning*)
2. Unüberwachtes Lernen (*Unsupervised Learning*)
3. Bekräftigungslernen (*Reinforcement Learning*)

Spezifikation von Lernproblemen

Lernparadigmen

1. Überwachtes Lernen (*Supervised Learning*)

Gelernt werden soll eine Funktion aus gegebenen Paaren von Ein- und Ausgaben. Wichtiges Teilgebiet des überwachten Lernens ist die automatische Klassifikation. Beispiel: Handschrifterkennung

2. Unüberwachtes Lernen (*Unsupervised Learning*)

Ziel ist das Erkennen von Strukturen in Daten. Hierzu gehört die automatische Einteilung in Kategorien (Clustering), Optimierung von Modellparametern (Expectation Maximization), oder die Übersetzung der beobachteten Daten in eine einfachere Repräsentation (Faktoranalyse).

3. Bekräftigungslernen (*Reinforcement Learning*)

Ziel ist das Erlernen bzw. Anpassen oder Optimieren einer Verhaltensstrategie durch Feedback aus der Umwelt unter dem Prinzip der Nutzenmaximierung. Beispiel: Agenten in einer virtuellen Welt

Spezifikation von Lernproblemen

Beispiel Schach: Art der Erfahrung

1. Feedback

- direkt: Zu einer Brettkonfiguration ist der beste Zug gegeben.
- indirekt: Es gibt nur das Resultat eine ganzen Serie von Zügen, am Ende eines Spiels.

Spezifikation von Lernproblemen

Beispiel Schach: Art der Erfahrung

1. Feedback

- direkt: Zu einer Brettkonfiguration ist der beste Zug gegeben.
- indirekt: Es gibt nur das Resultat eine ganzen Serie von Zügen, am Ende eines Spiels.

2. Reihenfolge der Beispiele

- Ein Lehrer präsentiert dem Lernenden wichtige Beispiele einschließlich deren Lösung.
- Der Lernende wählt die Beispiele selbst aus; z. B. Brettkonfiguration, für die er keinen passenden Zug kennt.

Spezifikation von Lernproblemen

Beispiel Schach: Art der Erfahrung

1. Feedback

- direkt: Zu einer Brettkonfiguration ist der beste Zug gegeben.
- indirekt: Es gibt nur das Resultat eine ganzen Serie von Zügen, am Ende eines Spiels.

2. Reihenfolge der Beispiele

- Ein Lehrer präsentiert dem Lernenden wichtige Beispiele einschließlich deren Lösung.
- Der Lernende wählt die Beispiele selbst aus; z. B. Brettkonfiguration, für die er keinen passenden Zug kennt.

3. Relevanz hinsichtlich des Gütemaßes

- Wieviel lässt sich aufgrund der Erfahrung lernen, um in der realen Situation zu bestehen? Allein durch Spielen gegen sich selbst kann man bei Schach nicht auf Weltmeisterniveau kommen.

Spezifikation von Lernproblemen

Beispiel Schach: eine ideale Zielfunktion γ

(a) *chooseMove* : *Boards* \rightarrow *Moves*

(b) γ : *Boards* \rightarrow \mathbb{R}

Rekursive Definition von γ im Stil einer *Means-End-Analyse*:

Sei $o \in$ *Boards*.

1. $\gamma(o) = 100$, falls o einen gewonnenen Endzustand repräsentiert.
2. $\gamma(o) = -100$, falls o einen verlorenen Endzustand repräsentiert.
3. $\gamma(o) = 0$, falls o einen unentschiedenen Endzustand repräsentiert.
4. $\gamma(o) = \gamma(o^*)$ sonst.

Dabei bezeichne o^* den besten Endzustand, der erreicht werden kann, wenn beide Seiten optimal spielen. Siehe Vorlesung [Suchalgorithmen](#) [Stein 1998-2008]: Spielbaumsuche, MiniMax-Strategie, α - β -Pruning

Spezifikation von Lernproblemen

Beispiel Schach: eine Approximation y der idealen Zielfunktion γ

$$\gamma(o) \sim y(\alpha(o)) \equiv y(\mathbf{x}) := w_0 + w_1 \cdot x_1 + w_2 \cdot x_2 + w_3 \cdot x_3 + w_4 \cdot x_4 + w_5 \cdot x_5 + w_6 \cdot x_6$$

mit

x_1 = Anzahl schwarzer Figuren in Konfiguration o

x_2 = Anzahl weißer Figuren in Konfiguration o

x_3 = Anzahl schwarzer Könige in Konfiguration o

x_4 = Anzahl weißer Könige in Konfiguration o

x_5 = Anzahl bedrohter weißer Figuren in Konfiguration o

x_6 = Anzahl bedrohter schwarzer Figuren in Konfiguration o

Weitere Möglichkeiten:

- Regelmenge
- neuronales Netz
- Polynom über Merkmale einer Brettkonfiguration
- ...

Bemerkungen:

- Ideale Zielfunktionen sind meist unbekannt bzw. können nicht gelernt werden. Die Funktion y kann als eine Approximation einer idealen Zielfunktion γ aufgefasst werden.
- Ein wesentlicher Unterschied zwischen einer idealen Zielfunktion γ und einer Approximationsfunktion y besteht in der Größe und Darstellung des Eingaberaums: die ideale Zielfunktion γ schließt von der realen Situation bzw. der vollständigen Repräsentation eines Objektes auf einen Zielfunktionswert; y beschränkt sich auf ausgewählte, typischerweise leicht zu beobachtende Merkmale x . Beispiele:
 - Ein Schachgroßmeister bewertet eine Stellung in seiner Gesamtheit (intuitiv und analytisch); ein Schachprogramm beschränkt sich auf bestimmte Merkmale.
 - Ein menschlicher Pilzsammler bewertet einen Pilz als Ganzes (intuitiv, analytisch, sensitiv); ein Klassifikationsprogramm beschränkt sich auf einige leicht messbare Merkmale.
- Im Schach ist man an einer reellwertigen Bewertungsfunktion interessiert; solche Probleme stellen ein Regressionsproblem dar. Ist man nur an festen Bewertungsstufen interessiert, z.B. Bewertung mit Zahlenwert von 1 bis 10 oder Schulnoten, so liegt ein Klassifikationsproblem vor. Durch eine Diskretisierung des Bildraumes kann jedes Regressionsproblem in ein Klassifikationsproblem umgewandelt werden.
- Regressionsprobleme und Klassifikationsprobleme unterscheiden sich häufig durch die verwendeten Funktionen zur Bestimmung von Fehlern. Bei Regressionsproblemen wird die Güte z.B. durch die Summe der quadratischen Abstände für eine Menge von Beispielen geschätzt; bei Klassifikationsproblemen dient die Anzahl der fehlklassifizierten Beispiele.

Spezifikation von Lernproblemen

Spezifikation von Klassifikationsproblemen

Situation in der realen Welt:

- O sei ein Menge von Objekten.
- C sei eine Menge von Klassen.
- $\gamma : O \rightarrow C$ sei der ideale Klassifikator für O .

Spezifikation von Lernproblemen

Spezifikation von Klassifikationsproblemen

Situation in der realen Welt:

- O sei eine Menge von Objekten.
- C sei eine Menge von Klassen.
- $\gamma : O \rightarrow C$ sei der ideale Klassifikator für O .

Klassifikation:

- Feststellung der Klasse $\gamma(o) \in C$ für gegebene $o \in O$.

Ansatz zur Automatisierung:

1. Abstraktion der Objekte $o \in O$ zu Merkmalsvektoren $\mathbf{x} = \alpha(o)$.
2. Erstellung einer Menge von Beispielen $(\alpha(o), \gamma(o))$ bzw. $(\mathbf{x}, c(\mathbf{x}))$. Das heißt, für $\mathbf{x} = \alpha(o)$ ist $c(\mathbf{x})$ definiert als $\gamma(o)$.
3. Approximation von c .

Spezifikation von Lernproblemen

Spezifikation von Klassifikationsproblemen

Situation im Modell:

- X sei ein Instanzenraum (Merkmalsraum) über endlich vielen Merkmalen.
- C sei eine Menge von Klassen.
- $c : X \rightarrow C$ sei der zu lernende Klassifikator für X .
- $D = \{(\mathbf{x}_1, c(\mathbf{x}_1)), \dots, (\mathbf{x}_n, c(\mathbf{x}_n))\} \subseteq X \times C$ sei eine Menge von Lernbeispielen.

Spezifikation von Lernproblemen

Spezifikation von Klassifikationsproblemen

Situation im Modell:

- X sei ein Instanzenraum (Merkmalsraum) über endlich vielen Merkmalen.
- C sei eine Menge von Klassen.
- $c : X \rightarrow C$ sei der zu lernende Klassifikator für X .
- $D = \{(\mathbf{x}_1, c(\mathbf{x}_1)), \dots, (\mathbf{x}_n, c(\mathbf{x}_n))\} \subseteq X \times C$ sei eine Menge von Lernbeispielen.

Maschinelles Lernproblem:

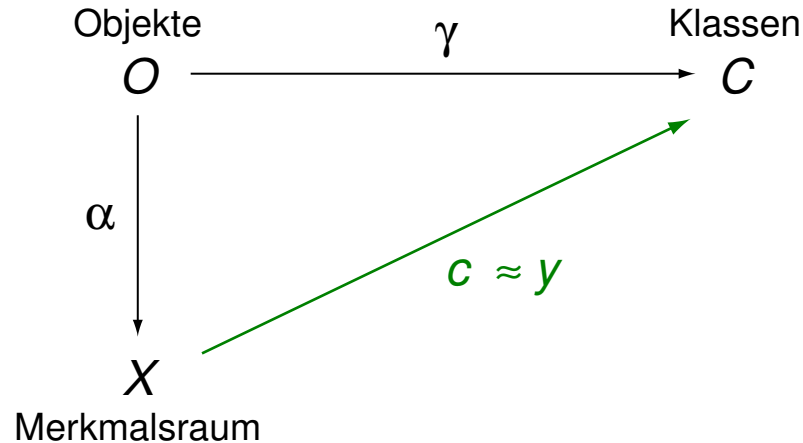
- Bestimmung eines Klassifikators $y : X \rightarrow C$ auf Basis von D .

Lösung:

- Statistik, Theorie und Algorithmen des Maschinellen Lernens.

Spezifikation von Lernproblemen

Spezifikation von Klassifikationsproblemen



Abbildungen:

- γ . idealer Klassifikator auf den realen Objekten
- α . Modellbildungsfunktion
- c . zu lernender Klassifikator auf dem Merkmalsraum
- y . Approximationsfunktion für c

Bemerkungen:

- ❑ Der Merkmalsraum X enthält die gemäß unserer Vorstellung (= unseres Modells) berechneten Darstellungen $\mathbf{x}_1, \mathbf{x}_2, \dots$ der realen Objekte o_1, o_2, \dots
- ❑ Die Modellbildungsfunktion α bestimmt die Abstraktion / Darstellungstreue / Exaktheit / Qualität / Vereinfachung bei der Berechnung eines \mathbf{x} als $\mathbf{x} = \alpha(o)$.
- ❑ Entscheidungsprobleme sind Klassifikationsprobleme mit genau zwei Klassen.
- ❑ Das Halteproblem für Turingmaschinen ist ein unentscheidbares Klassifikationsproblem.

Spezifikation von Lernproblemen

LMS-Algorithmus zur Bestimmung von y

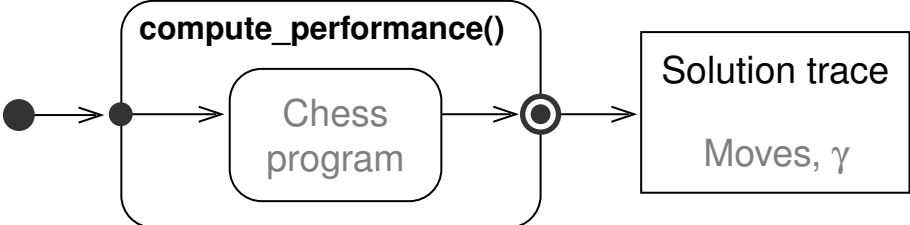
Algorithm:	LMS	Least Mean Square
Input:	D	Lernbeispiele der Form $(\mathbf{x}, c(\mathbf{x}))$ mit $c(\mathbf{x})$ als Zielfunktionswert für \mathbf{x} .
	η	Lernrate; eine positive kleine Konstante.
Internal:	$y(D)$	Menge der $y(\mathbf{x})$ für die Vektoren \mathbf{x} aus D .
Output:	\mathbf{w}	Gewichtsvektor.

$LMS(D, \eta)$

1. $initialize_random_weights((w_0, w_1, \dots, w_p))$
2. **REPEAT**
3. $\langle \mathbf{x}, c(\mathbf{x}) \rangle = random_select(D)$
4. $y(\mathbf{x}) = w_0 + w_1 \cdot x_1 + \dots + w_p \cdot x_p$
5. $error = c(\mathbf{x}) - y(\mathbf{x})$
6. **FOR** $i = 1$ **TO** p **DO** $w_i = w_i + \eta \cdot error \cdot x_i$
7. **UNTIL** $(convergence(D, y(D)))$
8. $return((w_0, w_1, \dots, w_p))$

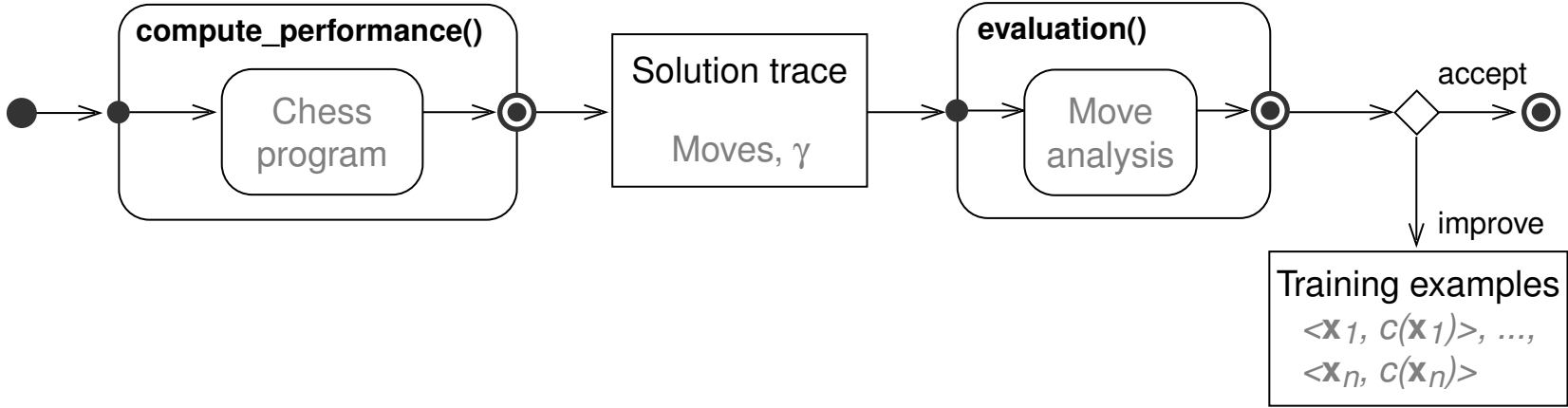
Spezifikation von Lernproblemen

Aufbau von lernenden Systemen [cf. p.12, Mitchell 1997]



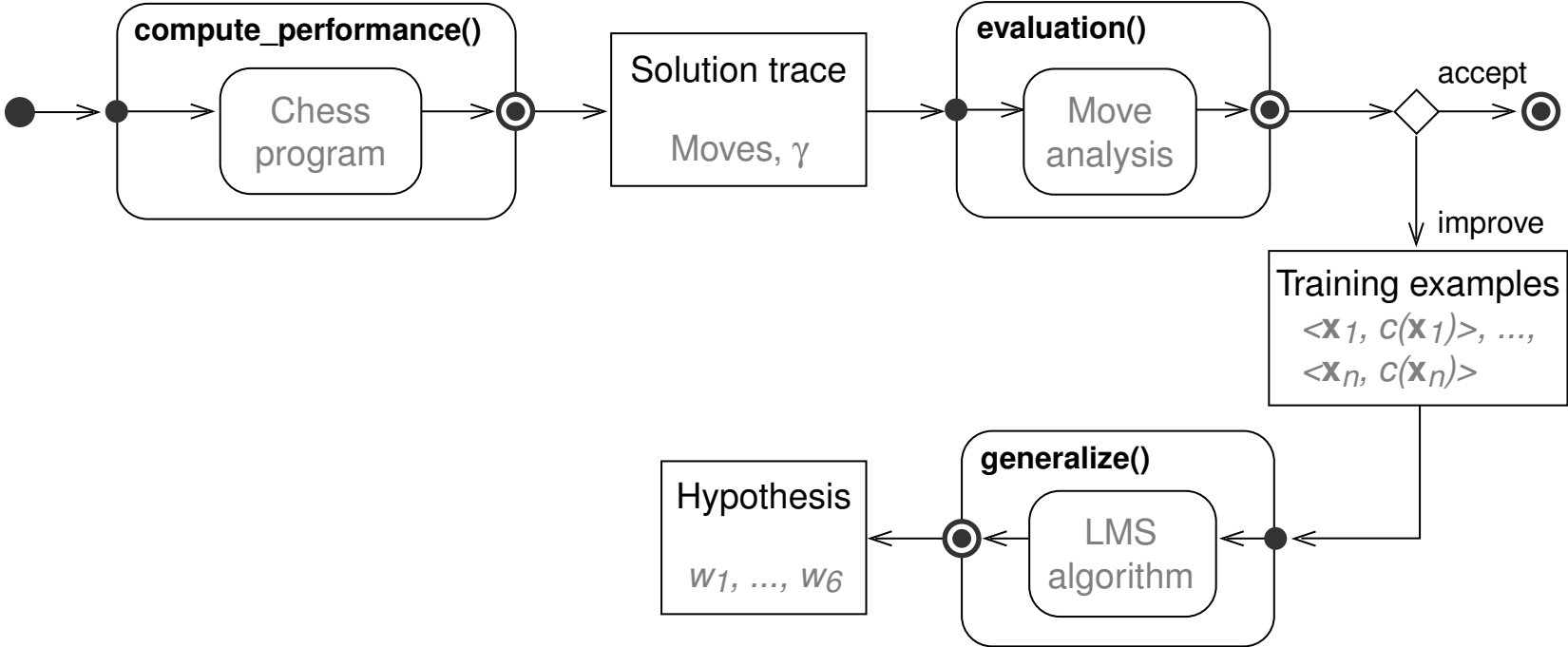
Spezifikation von Lernproblemen

Aufbau von lernenden Systemen [cf. p.12, Mitchell 1997]



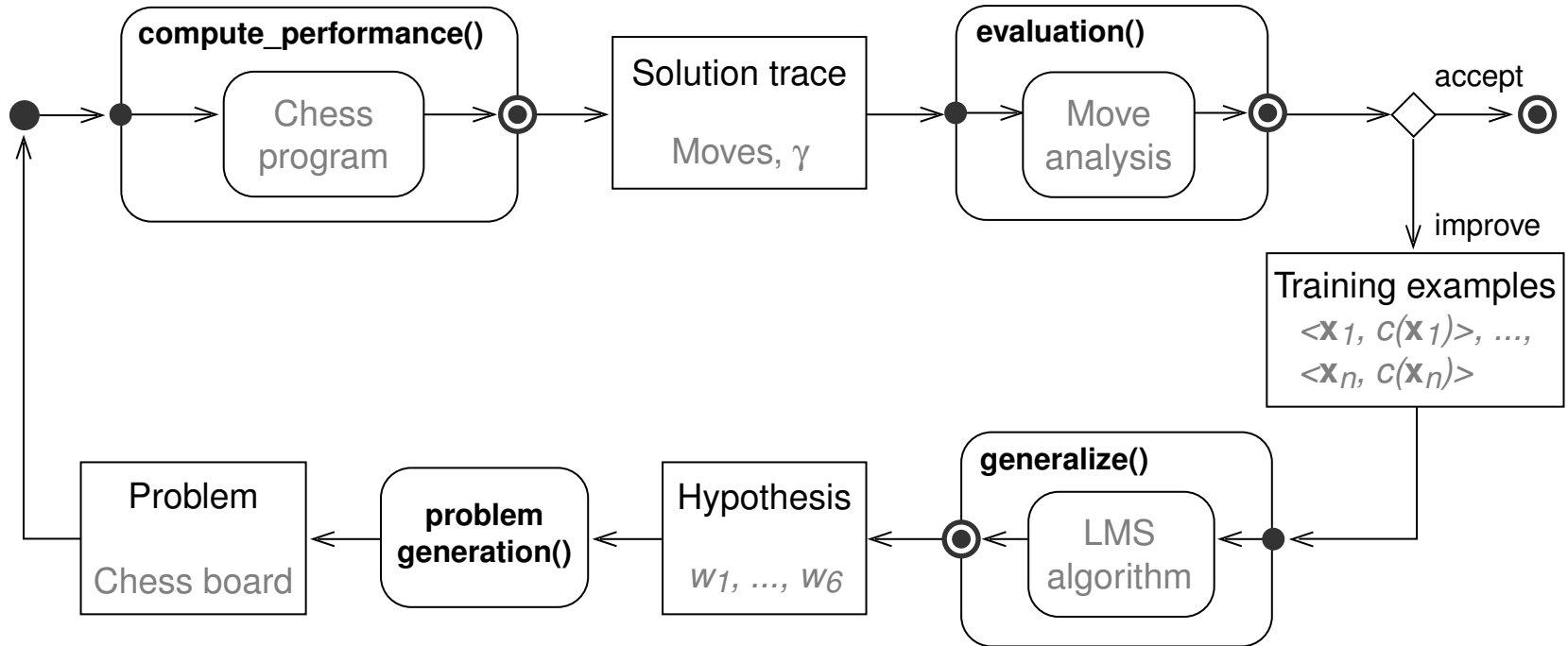
Spezifikation von Lernproblemen

Aufbau von lernenden Systemen [cf. p.12, Mitchell 1997]



Spezifikation von Lernproblemen

Aufbau von lernenden Systemen [cf. p.12, Mitchell 1997]



Design-Entscheidungen:

1. Art der Trainingserfahrung
2. Definition einer idealen Zielfunktion γ
3. Repräsentation und Definition einer Approximation y von γ
4. Lernalgorithmus zur Bestimmung einer geschlossenen Form von y

Spezifikation von Lernproblemen

Fragestellungen

- ❑ Welche Algorithmen sind zur Funktionsapproximation geeignet?
- ❑ Wie sehen Gütemaße zur Beurteilung der Genauigkeit aus?
- ❑ Wie beeinflusst die Anzahl der Trainingsbeispiele die Genauigkeit?
- ❑ Wie beeinflussen verrauschte Daten die Genauigkeit?
- ❑ Wo sind die theoretischen Grenzen der Lernbarkeit?
- ❑ Wie lässt sich Vorwissen in Lernverfahren integrieren?
- ❑ Was kann man bei biologischen Systemen abschauen?