



Seminar zum Thema „Künstliche Intelligenz“

Diskriminanzanalyse

Patrick Wolf (wi5706)

18.05.2005

I. Einführung

1. Problemstellung
2. Zielsetzung der Präsentation

II. Grundlagen der Diskriminanzanalyse

1. Phasen der Diskriminanzanalyse
2. Auswahlkriterien für eine Diskriminanzregel
 - a) Arten von Merkmalen
 - b) Beurteilung der Klassifikationsgüte

III. Diskriminanzanalytische Verfahren

1. Bayes-Klassifikator
2. Neuronale Netze
3. K-Nearest-Neighbor
4. CART

IV. Zusammenfassung

I. Einführung

1. Problemstellung
2. Zielsetzung der Präsentation

II. Grundlagen der Diskriminanzanalyse

1. Phasen der Diskriminanzanalyse
2. Auswahlkriterien für eine Diskriminanzregel
 - a) Arten von Merkmalen
 - b) Beurteilung der Klassifikationsgüte

III. Diskriminanzanalytische Verfahren

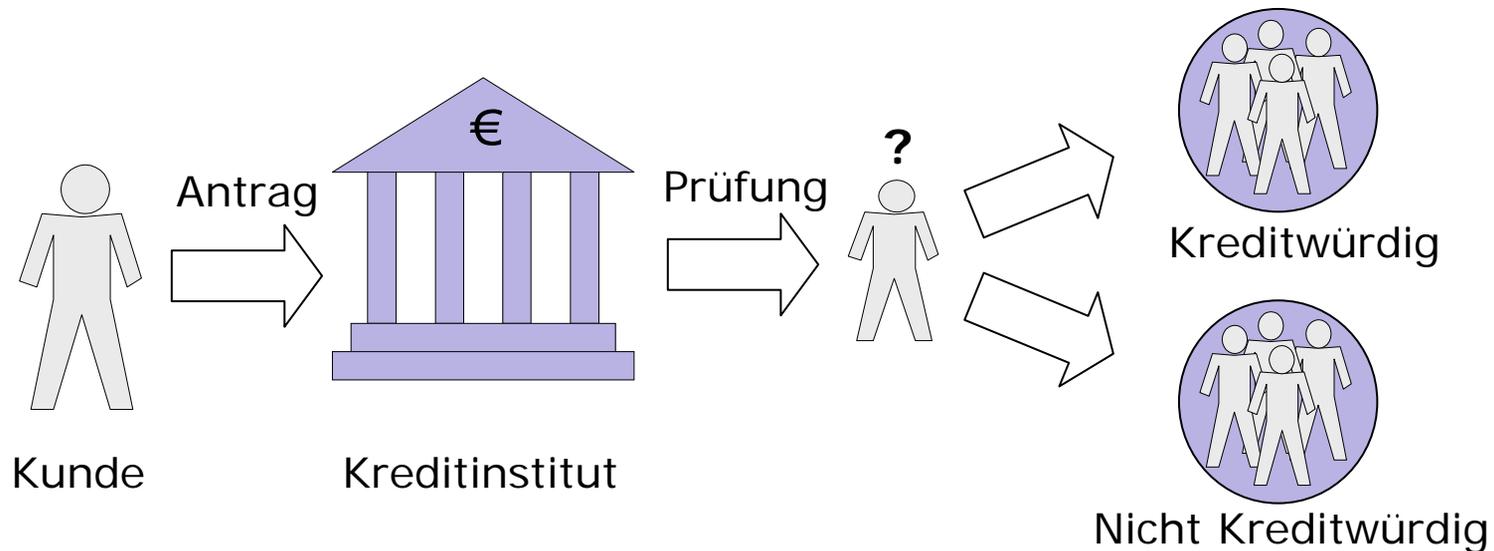
1. Bayes-Klassifikator
2. Neuronale Netze
3. K-Nearest-Neighbor
4. CART

IV. Zusammenfassung

I. Einführung

1. Problemstellung - Beispiel

- Eine Bank möchte möglichst nur denjenigen Kunden einen Kredit gewähren, von denen sie eine problemlose Rückzahlung der Verbindlichkeiten erwarten kann.
- Daher prüft sie im Vorfeld der Kreditvergabe anhand bestimmter Merkmale (z.B. Alter, Gehalt) die Kreditwürdigkeit des Kunden.



I. Einführung

1. Problemstellung

- Welche Merkmale führen zu einer signifikanten Trennung zwischen den Gruppen „Kreditwürdig“ und „Nicht Kreditwürdig“?
- In welche Gruppe ist ein neuer Kunde aufgrund seiner Merkmalsausprägungen einzuordnen?
- Diese Probleme bilden den Ausgangspunkt der Diskriminanzanalyse:

Die Diskriminanzanalyse beschäftigt sich mit der Untersuchung von Gruppenunterschieden und der Vorhersage der Gruppenzugehörigkeit von Objekten

- Im Gegensatz zur Clusteranalyse werden hierbei keine neuen Gruppen gebildet, sondern nur bereits bestehende betrachtet.

I. Einführung

2. Zielsetzung der Präsentation

- *Die Präsentation soll ...*
 - die Grundlagen der Diskriminanzanalyse vermitteln und ausgewählte Verfahren zur Vorhersage der Gruppenzugehörigkeit von Objekten vorstellen.
 - sich auf den Fall zweier Gruppen beschränken, da dieser in der Praxis häufig anzutreffen ist.
 - auf eine Erläuterung des Vorgehens bei der Untersuchung von Gruppenunterschieden verzichten und stattdessen von bereits selektierten Merkmalen ausgehen.

I. Einführung

1. Problemstellung
2. Zielsetzung der Präsentation

II. Grundlagen der Diskriminanzanalyse

1. Phasen der Diskriminanzanalyse
2. Auswahlkriterien für eine Diskriminanzregel
 - a) Arten von Merkmalen
 - b) Beurteilung der Klassifikationsgüte

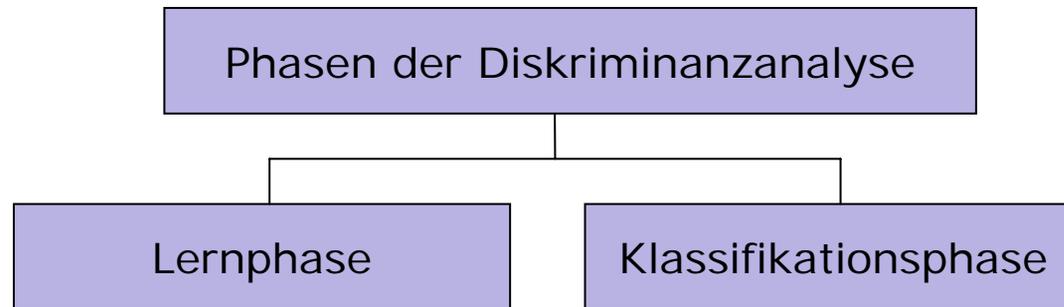
III. Diskriminanzanalytische Verfahren

1. Bayes-Klassifikator
2. Neuronale Netze
3. K-Nearest-Neighbor
4. CART

IV. Zusammenfassung

II. Grundlagen der Diskriminanzanalyse

1. Phasen der Diskriminanzanalyse

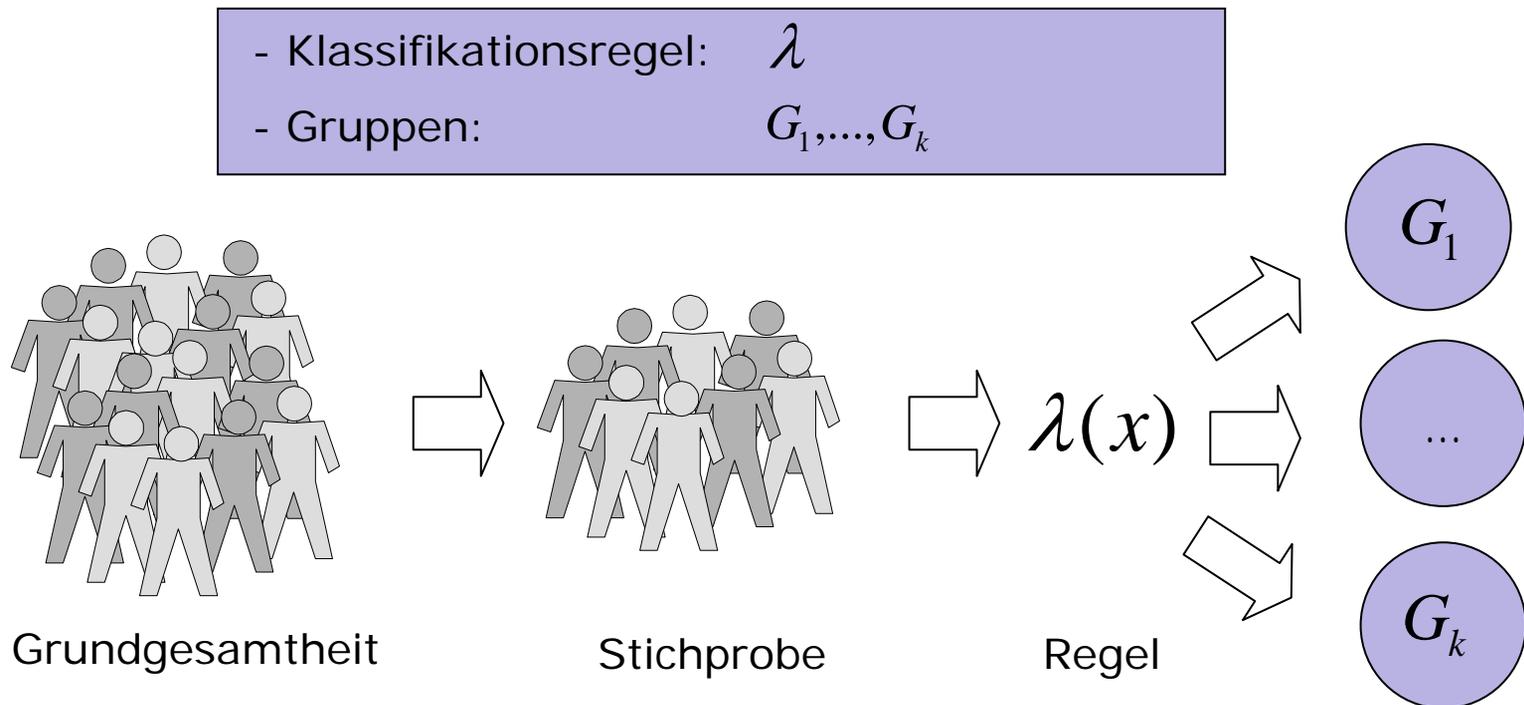


- Lernphase:
Aus der Grundgesamtheit wird eine Trainingsstichprobe entnommen und für jedes darin befindliche Objekt die Gruppenzugehörigkeit sowie die p Ausprägungen x_1, \dots, x_p in den Merkmalen X_1, \dots, X_p bestimmt.

- Grundgesamtheit: Ω
- Merkmalsvektor: $x = (x_1, \dots, x_p)$
- Merkmale X_1, \dots, X_p

II. Grundlagen der Diskriminanzanalyse

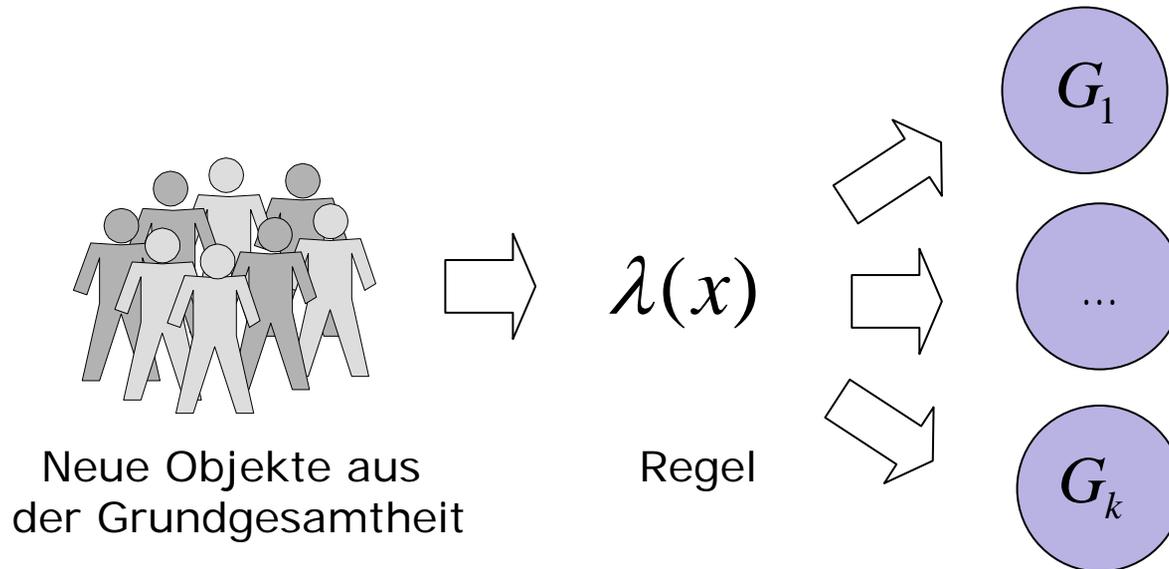
- Lernphase – Fortsetzung:
 Gesucht wird nun eine Klassifikationsregel, die jedes Trainingsobjekt anhand seines spezifischen Merkmalsvektors x genau einer der k vordefinierten Gruppen G_1, \dots, G_k zuordnet.



II. Grundlagen der Diskriminanzanalyse

- Klassifikationsphase:

In der Klassifikationsphase wird die Zuordnungsregel auf neue Objekte aus der Grundgesamtheit angewendet und diese entsprechend klassifiziert.



II. Grundlagen der Diskriminanzanalyse

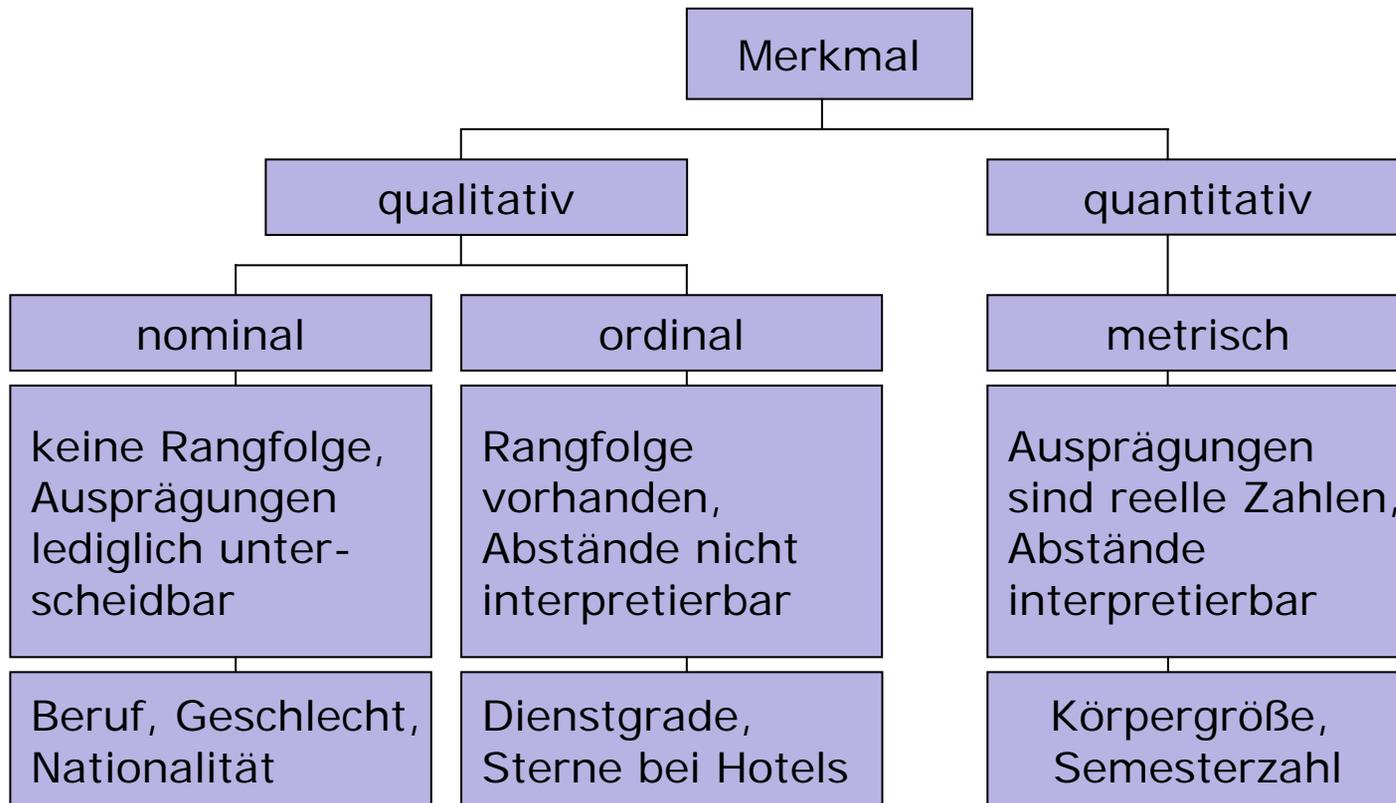
2. Auswahlkriterien für eine Diskriminanzregel

- Für die Auswahl eines Klassifikationsverfahrens spielen eine Vielzahl von Kriterien eine wichtige Rolle, z.B.:
 - Arten der zugrunde liegenden Merkmale
 - Güte der Klassifikation
 - Geschwindigkeit eines Verfahrens
 - Interpretierbarkeit der Ergebnisse
 - etc.
- Im Folgenden sollen die Arten der zugrunde liegenden Merkmale sowie ein Maß zur Beurteilung der Klassifikationsgüte näher betrachtet werden.

II. Grundlagen der Diskriminanzanalyse

2.1 Arten von Merkmalen

- Die bei der Klassifikation relevanten Merkmale X_1, \dots, X_p können in quantitative und qualitative Merkmale unterschieden werden.



2.2 Beurteilung der Klassifikationsgüte

- Ziel einer Zuordnungsregel ist die korrekte Klassifikation der unbekanntem Objekte aus der Grundgesamtheit.
- Korrekte Klassifikation → prognostizierte Gruppenzugehörigkeit entspricht der tatsächlichen Gruppenzugehörigkeit des Objekts.
- Jede Zuordnungsregel ist fehlerbehaftet:
 1. Die Regel kann ein Objekt aus der Gruppe G1 irrtümlich der Gruppe G2 zuordnen.
 2. Die Regel kann ein Objekt aus der Gruppe G2 irrtümlich der Gruppe G1 zuordnen.

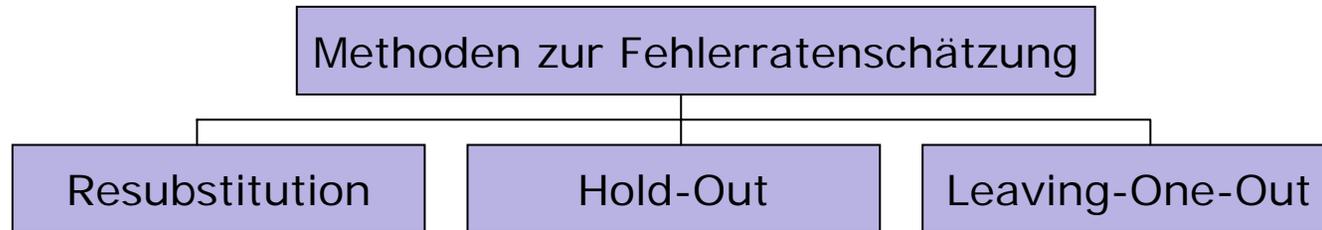
II. Grundlagen der Diskriminanzanalyse

2.2 Beurteilung der Klassifikationsgüte - Fehlerrate

- Die Güte einer Klassifikation bemisst sich anhand der Fehlerrate:

$$E(\lambda) = \frac{\text{Anzahl falsch klassifizierter Objekte}}{\text{Anzahl klassifizierter Objekte}}$$

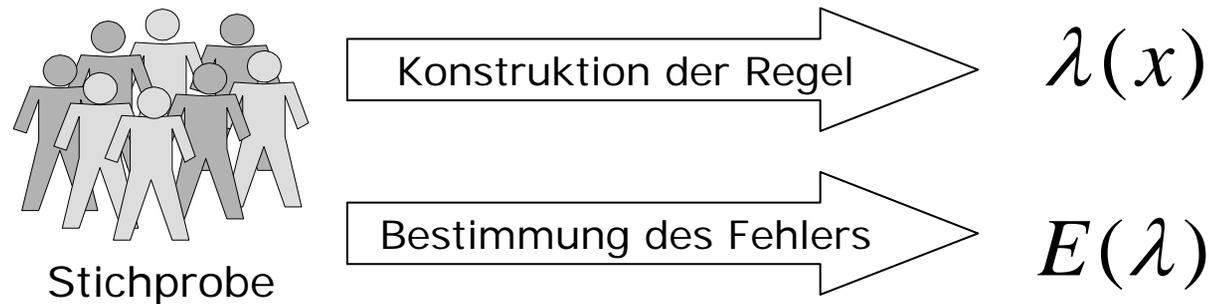
- Verfahren zur Fehlerratenschätzung:



II. Grundlagen der Diskriminanzanalyse

2.2 Beurteilung der Klassifikationsgüte - Resubstitutionsmethode

- Die Trainingsstichprobe wird sowohl zur Konstruktion der Regel als auch zur Ermittlung der Fehlerrate eingesetzt.

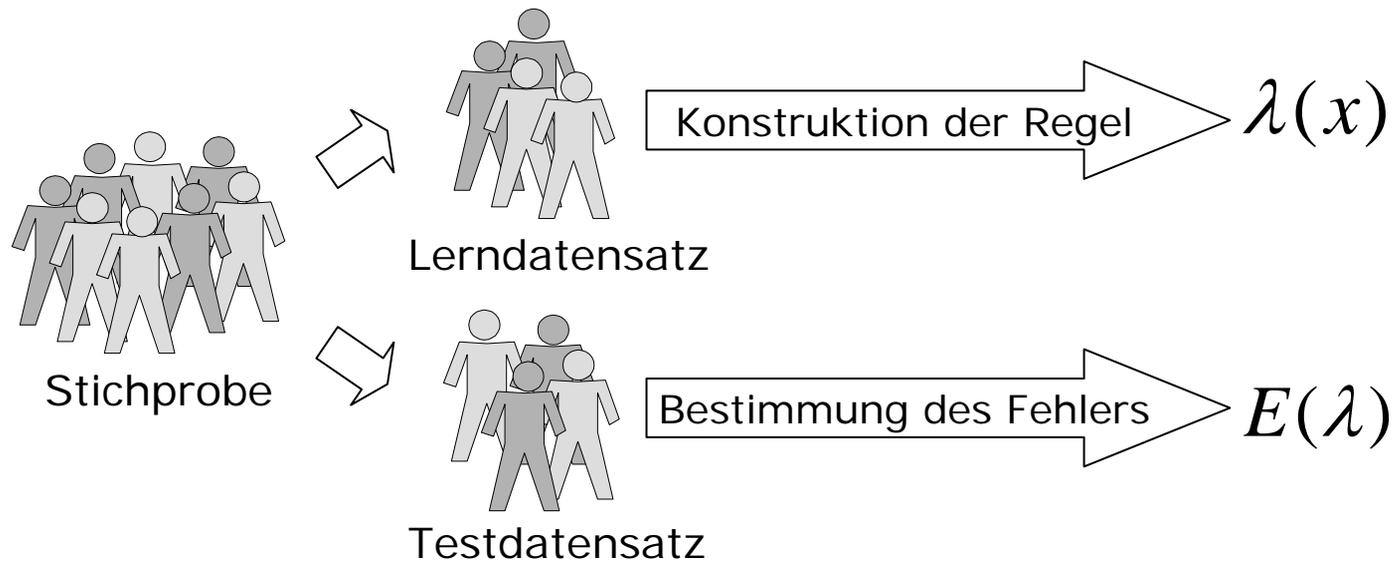


Ergibt häufig eine zu optimistische Fehlerratenschätzung. (1)

II. Grundlagen der Diskriminanzanalyse

2.2 Beurteilung der Klassifikationsgüte – Hold-Out-Methode

- Bei der Hold-Out-Methode wird die Trainingsstichprobe zufällig in einen Lern- und in einen Testdatensatz aufgeteilt.



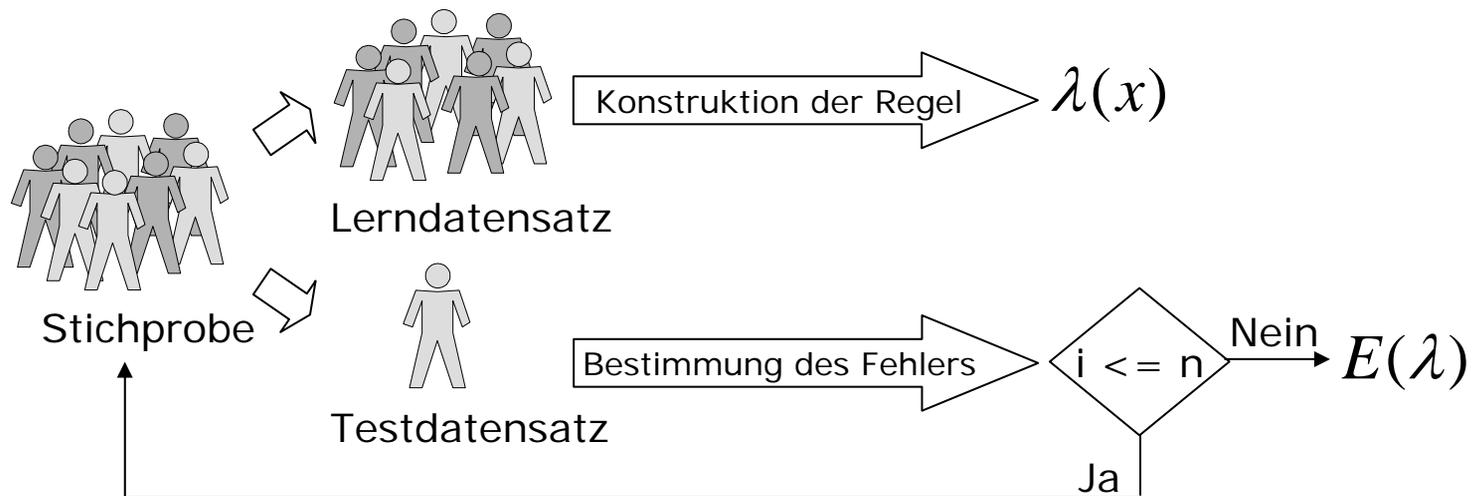
Die Aufteilung vermeidet die Verfälschung der Fehlerrate und führt ab einem Umfang von 1000 Testobjekten zu einer guten Approximation der wahren Fehlerrate.(1)

II. Grundlagen der Diskriminanzanalyse

2.2 Beurteilung der Klassifikationsgüte – Leaving-One-Out

- Das i -te Objekt wird als Testdatensatz aus der Stichprobe entnommen, die verbleibenden $n-1$ Objekte bilden den Lerndatensatz.

Die Regel wird anhand der Lerndaten bestimmt und mit dem Testfall überprüft. Der Prozess wird für $i = 1, \dots, n$ wiederholt.



Liefert im Mittel einen richtigen Schätzwert. Wird als Standardmethode zur Fehlerratschätzung empfohlen. (1)

II. Grundlagen der Diskriminanzanalyse

2.2 Beurteilung der Klassifikationsgüte - Fehlklassifikationskosten

- Bisher unterstellt:
Beide Arten von Fehlentscheidungen haben die gleiche Bedeutung.
- In der Praxis ist diese Annahme meist nicht erfüllt:

„Beispielsweise ist das Risiko, einen gesunden Patienten als krank zu klassifizieren in der Regel geringer zu bewerten als das Risiko, einen kranken Patienten als gesund zu klassifizieren.“

(Bonne, Thorsten / Arminger, Gerhard, Diskriminanzanalyse, 2001, S. 197)

- Einführung von Fehlklassifikationskosten $C(i,j)$.
 - $C(i,j) > 0$, wenn i nicht j .
 - $C(i,j) = 0$, sonst
- Allerdings sind die Fehlklassifikationskosten in der Praxis meist schwer zu bestimmen und sehr subjektiv, daher werden identische Kosten unterstellt.

I. Einführung

1. Problemstellung
2. Zielsetzung der Präsentation

II. Grundlagen der Diskriminanzanalyse

1. Phasen der Diskriminanzanalyse
2. Auswahlkriterien für eine Diskriminanzregel
 - a) Arten von Merkmalen
 - b) Beurteilung der Klassifikationsgüte

III. Diskriminanzanalytische Verfahren

1. Bayes-Klassifikator
2. Neuronale Netze
3. K-Nearest-Neighbor
4. CART

IV. Zusammenfassung

III. Diskriminanzanalytische Verfahren

1. Bayes-Klassifikator

- Basiert auf wahrscheinlichkeitstheoretischen Überlegungen.
- Setzt bestimmte Kenntnisse über die Verteilung der Gruppen sowie der darin befindlichen Objekte in der Grundgesamtheit voraus.

Funktionsweise:

$$\lambda_{\text{Bayes}}(\mathbf{x}) = \begin{cases} G_1 & \text{wenn } P(G_1 | \mathbf{x}) \geq P(G_2 | \mathbf{x}) \\ G_2 & \text{wenn } P(G_1 | \mathbf{x}) < P(G_2 | \mathbf{x}) \end{cases}$$

Wahrscheinlichkeit für das Auftreten des Merkmalsvektors \mathbf{x} in der Gruppe G_i

Theorem von Bayes:

$$P(G_i | \mathbf{x}) = \frac{\pi_i * P(\mathbf{x} | G_i)}{\sum_{j=1}^k \pi_j P(\mathbf{x} | G_j)}$$

Wahrscheinlichkeit mit der ein Objekt aus G_i den Vektor \mathbf{x} besitzt

Wahrscheinlichkeit für die Gruppe G_i in der Grundgesamtheit

III. Diskriminanzanalytische Verfahren

1. Bayes-Klassifikator – Beispiel für die Berechnungen

- Für Schüler soll anhand der Merkmale „Gemessener IQ“ und „Schulabschluss der Eltern“ das zukünftige Abschneiden beim Abitur vorhergesagt werden. (1)

| Gemessener IQ | Abschluss der Eltern | Notenschnitt Abitur < 7 Pkt. (Gruppe G1) | Notenschnitt Abitur >= 7 Pkt. (Gruppe G2) |
|----------------------------------|----------------------|--|---|
| < 104 | Uni Abschluss | 55 | 38 |
| | Kein Uni-Abschluss | 114 | 76 |
| 104-121 | Uni Abschluss | 115 | 227 |
| | Kein Uni-Abschluss | 167 | 143 |
| > 121 | Uni Abschluss | 22 | 114 |
| | Kein Uni-Abschluss | 16 | 27 |
| Summe (Gesamt 1114 Fälle) | | 489 | 625 |

$$\pi_1 = \frac{489}{1114} = 0,44$$

$$\pi_2 = \frac{625}{1114} = 0,56$$

III. Diskriminanzanalytische Verfahren

1. Bayes-Klassifikator – Beispiel für die Berechnungen (2)

- Bedingte Wahrscheinlichkeiten für die Stichprobe:

| Gemessener IQ | Abschluss der Eltern | Notenschnitt Abitur < 7 Pkt. (Gruppe G1) | Notenschnitt Abitur >= 7 Pkt. (Gruppe G2) |
|----------------------------------|------------------------|--|---|
| < 104 (a) | Uni Abschluss (y) | 0,11 | 0,06 |
| | Kein Uni-Abschluss (z) | 0,23 | 0,12 |
| 104-121 (b) | Uni Abschluss (y) | 0,24 | 0,36 |
| | Kein Uni-Abschluss (z) | 0,34 | 0,23 |
| > 121 (c) | Uni Abschluss (y) | 0,04 | 0,18 |
| | Kein Uni-Abschluss (z) | 0,03 | 0,04 |
| Summe (Gesamt 1114 Fälle) | | 0,99 | 0,99 |

Berechnungsbeispiel:
(Werte siehe Folie 22)

$$P((a, y) | G_1) = \frac{55}{489} = 0,11$$

III. Diskriminanzanalytische Verfahren

1. Bayes-Klassifikator – Beispiel für die Berechnungen (3)

- Fertige Entscheidungstabelle:

| Gemessener IQ | Abschluss der Eltern | Notenschnitt Abitur < 7 Pkt. (Gruppe G1) | Notenschnitt Abitur >= 7 Pkt. (Gruppe G2) |
|----------------------------------|------------------------|--|---|
| < 104 (a) | Uni Abschluss (y) | 0,05 | 0,03 |
| | Kein Uni-Abschluss (z) | 0,10 | 0,07 |
| 104-121 (b) | Uni Abschluss (y) | 0,11 | 0,20 |
| | Kein Uni-Abschluss (z) | 0,15 | 0,13 |
| > 121 (c) | Uni Abschluss (y) | 0,02 | 0,10 |
| | Kein Uni-Abschluss (z) | 0,01 | 0,02 |
| Summe (Gesamt 1114 Fälle) | | 0,44 | 0,55 |

Berechnungsbeispiel
(Werte siehe Folie 23):

$$\pi_1 * P((a, y) | G_1) = 0,44 * 0,11 = 0,05$$

III. Diskriminanzanalytische Verfahren

1. Bayes-Klassifikator – Beispiel für die Berechnungen (4)

- Zuordnung der Objekte nach der Bayes-Regel:

| Gemessener IQ | Abschluss der Eltern | Notenschnitt Abitur < 7 Pkt. (Gruppe G1) | Notenschnitt Abitur >= 7 Pkt. (Gruppe G2) | Zuordnung |
|----------------------------------|------------------------|--|---|-----------|
| < 104 (a) | Uni Abschluss (y) | 0,05 | 0,03 | G1 |
| | Kein Uni-Abschluss (z) | 0,10 | 0,07 | G1 |
| 104-121 (b) | Uni Abschluss (y) | 0,11 | 0,20 | G2 |
| | Kein Uni-Abschluss (z) | 0,15 | 0,13 | G1 |
| > 121 (c) | Uni Abschluss (y) | 0,02 | 0,10 | G2 |
| | Kein Uni-Abschluss (z) | 0,01 | 0,02 | G2 |
| Summe (Gesamt 1114 Fälle) | | 0,44 | 0,55 | |

- Berechnung der Fehlerrate:

$$E(\lambda_{Bayes}) = 0,03 + 0,07 + 0,11 + 0,13 + 0,02 + 0,01 = 0,37$$

III. Diskriminanzanalytische Verfahren

1. Bayes-Klassifikator – Problem bei der Konstruktion der Regel

- Für jeden möglichen Merkmalsvektor muss die Wahrscheinlichkeit seines Auftretens bestimmt werden.
- Bei einer Vielzahl von Merkmalen und Ausprägungen ist dies häufig nicht praktikabel, da Aufwand zu groß.
- Lösung des Problems:
Unabhängigkeit der Merkmale wird unterstellt.
Für die Wahrscheinlichkeit des Merkmalsvektors gilt dann:

$$P(x | G_i) = \prod_{j=1}^p P(x_j | G_i)$$

- In der Praxis wird $P(x_j | G_i)$ meist noch geglättet, um zu verhindern, dass bei $P(x_j | G_i) = 0$ auch $P(x | G_i) = 0$ gilt.

III. Diskriminanzanalytische Verfahren

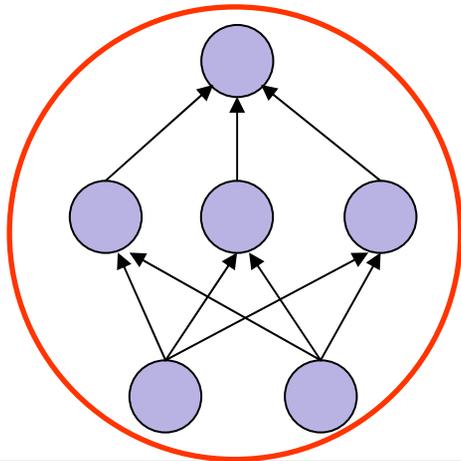
1. Bayes-Klassifikator – Eigenschaften

- + Besitzt theoretisch eine minimale Fehlerrate.
- + Hohe Geschwindigkeit und Präzision, daher besonders für große Datenmengen geeignet.
- Annahme der Merkmalsunabhängigkeit oftmals nicht zutreffend (z.B. ist das Merkmal „Gehalt“ nicht unabhängig vom Merkmal „Beruf“).
- Verlässliche Schätzung der Wahrscheinlichkeiten erfordert einen sehr großen Stichprobenumfang.

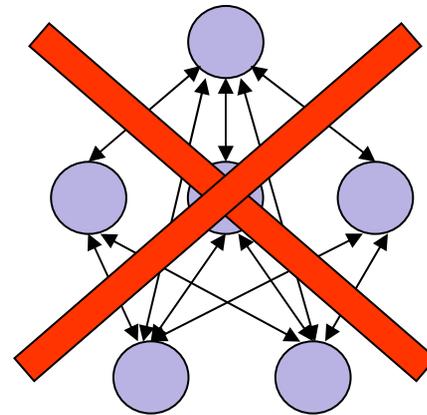
III. Diskriminanzanalytische Verfahren

2. Neuronale Netze (NN)

- Neuronale Netze simulieren die Funktionsweise des Gehirns.
- Sie bestehen aus einer Vielzahl künstlicher Neuronen, die über gerichtete und gewichtete Verbindungen Informationen austauschen.
- Hinsichtlich der zulässigen Flussrichtung der Informationen können zwei Arten von Netzen unterschieden werden:



Netze ohne Rückkoppelung
(Feed-Forward)

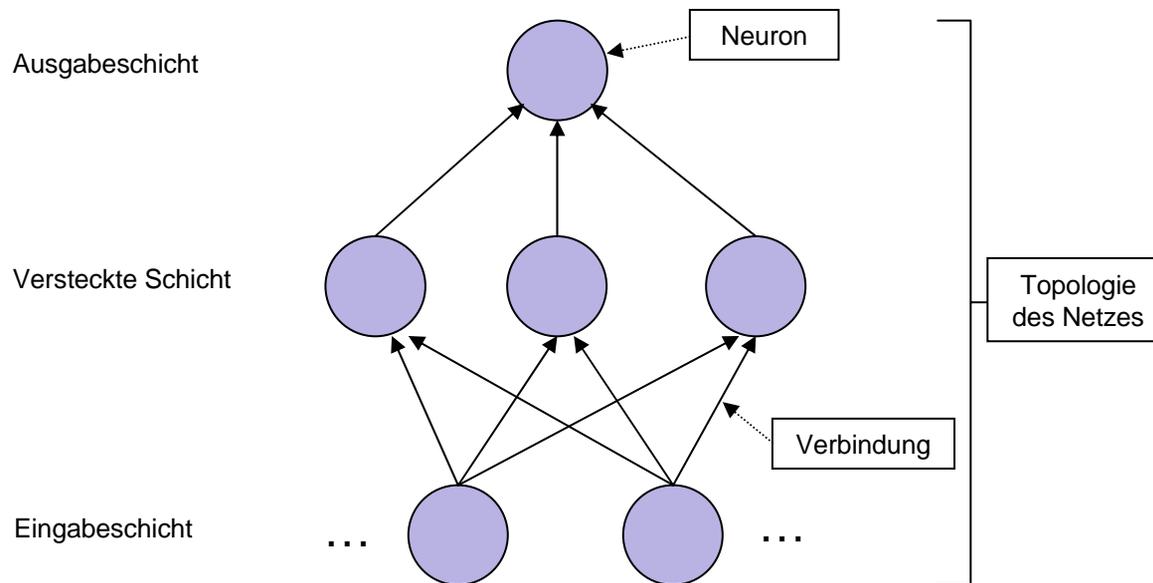


Netze mit Rückkoppelung
(Feed-Backward)

III. Diskriminanzanalytische Verfahren

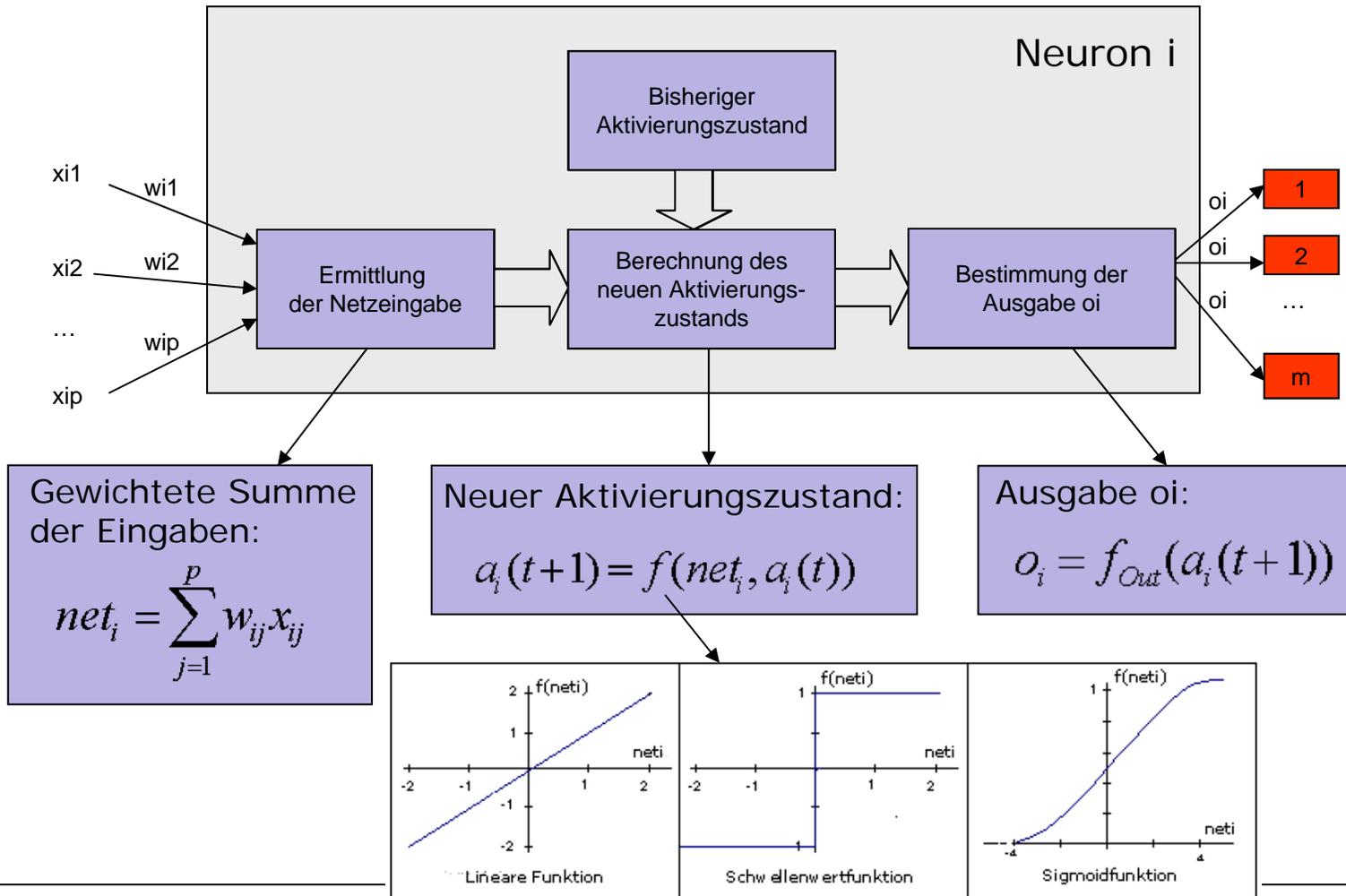
2. Neuronale Netze – Feed-Forward-Netze (FFN)

- Die Neuronen sind in Schichten angeordnet:
 - Eingabeschicht
 - Eine (mehrere) versteckte Schicht(en)
 - Ausgabeschicht
- Jedes Neuron einer vorgelagerten ist mit jedem Neuron einer nachgelagerten Schicht verbunden.



III. Diskriminanzanalytische Verfahren

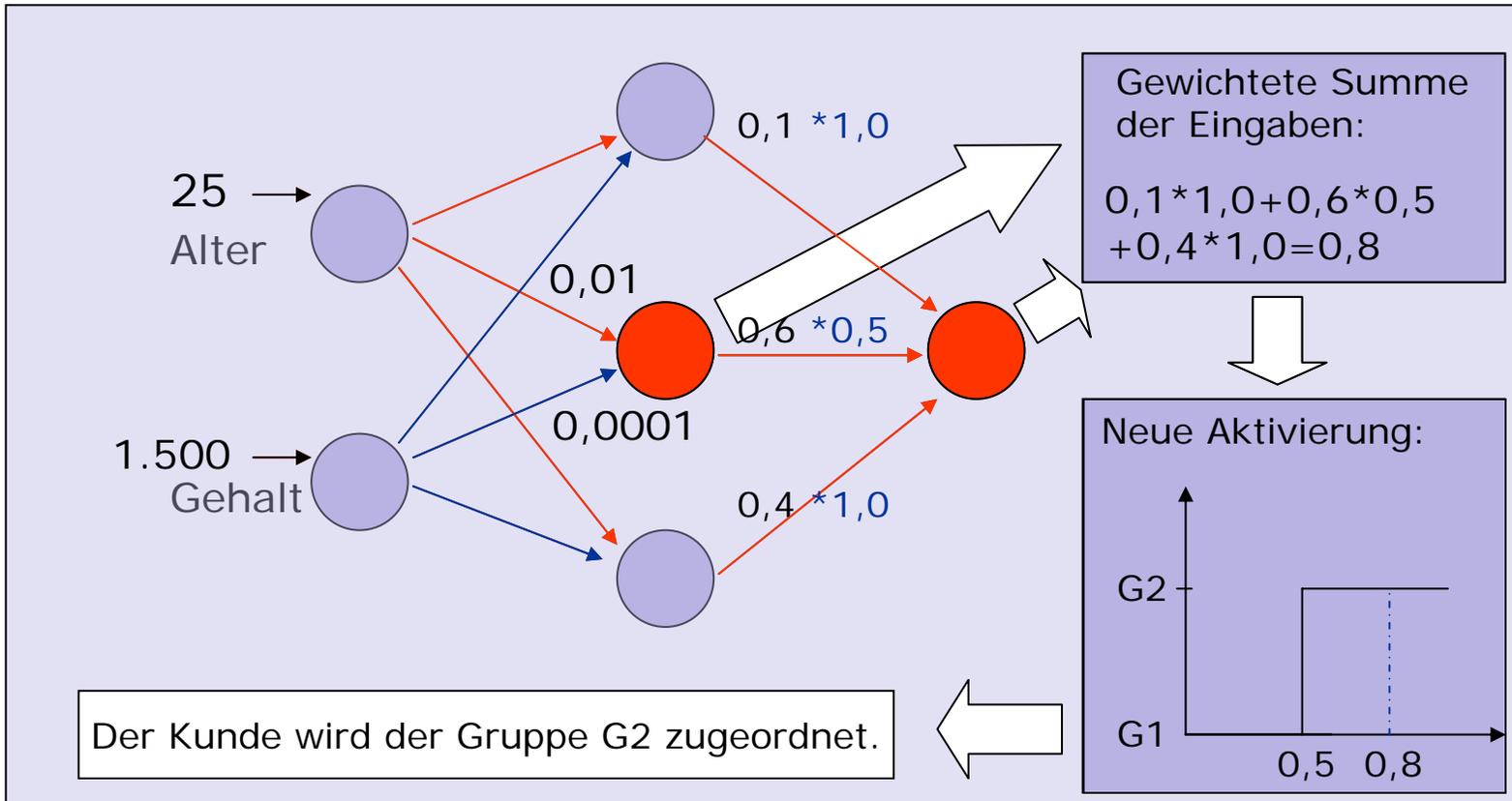
2. Neuronale Netze – Informationsverarbeitung in einem Neuron



III. Diskriminanzanalytische Verfahren

2. Neuronale Netze – Beispiel Kreditwürdigkeitsprüfung

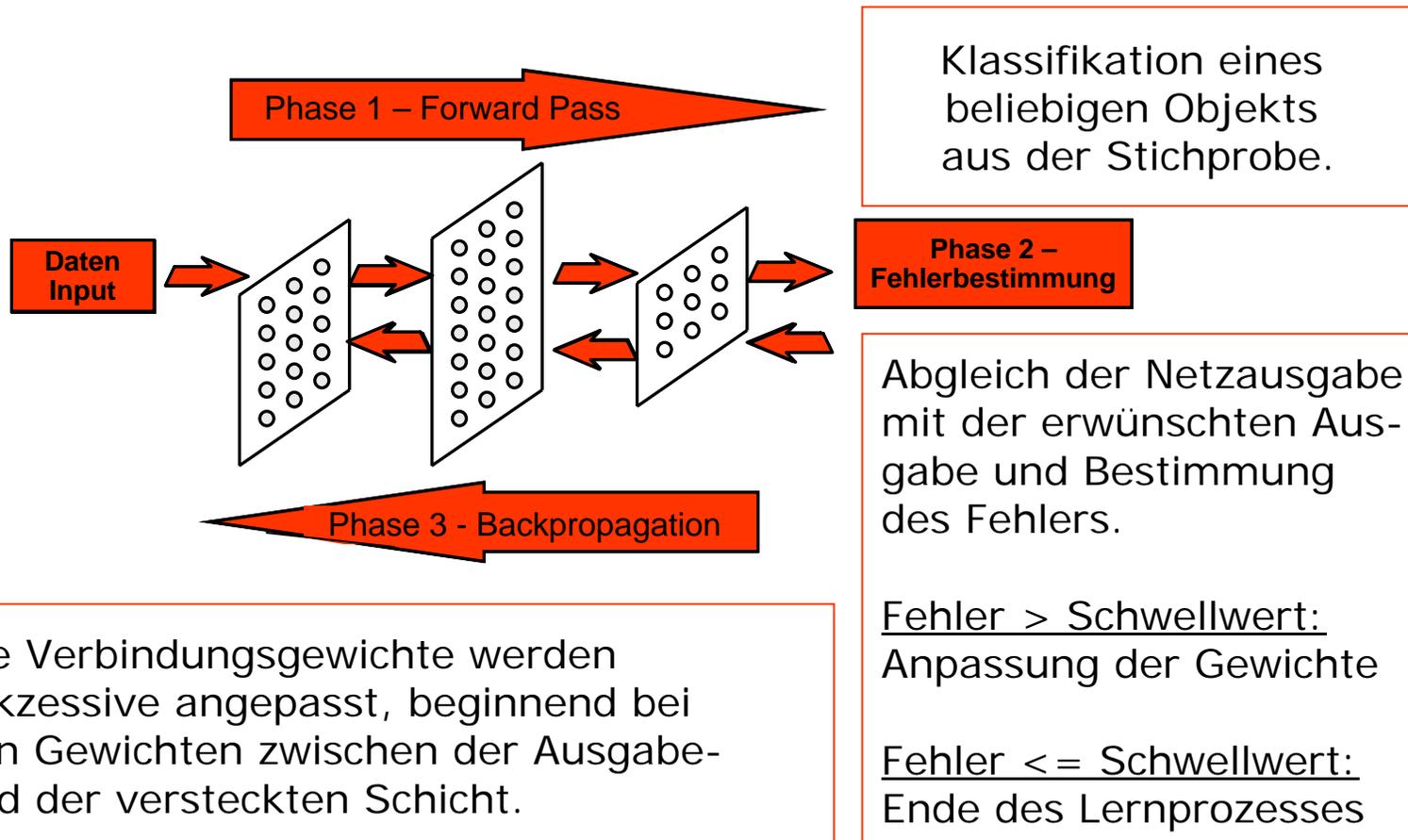
- 25-jähriger Kunde, 1.500€ regelmäßiges Gehalt



III. Diskriminanzanalytische Verfahren

2. Neuronale Netze – Lernprozess

- Lernen im Neuronalen Netz := Modifikation der Gewichte.



III. Diskriminanzanalytische Verfahren

2. Neuronale Netze – Testphase

- Überprüfung der Generalisierungsfähigkeit des Netzes nach Abschluss der Lernphase anhand einer Teststichprobe.
- Die Teststichprobe enthält neue Objekte, die mit dem Netz klassifiziert werden.
- Netzwerkfehler > Schwellwert:
Netz muss nachtrainiert werden.
- Netzwerkfehler \leq Schwellwert:
Netz besitzt genügend hohe Vorhersagegenauigkeit für zukünftige, neu zu klassifizierende Objekte.

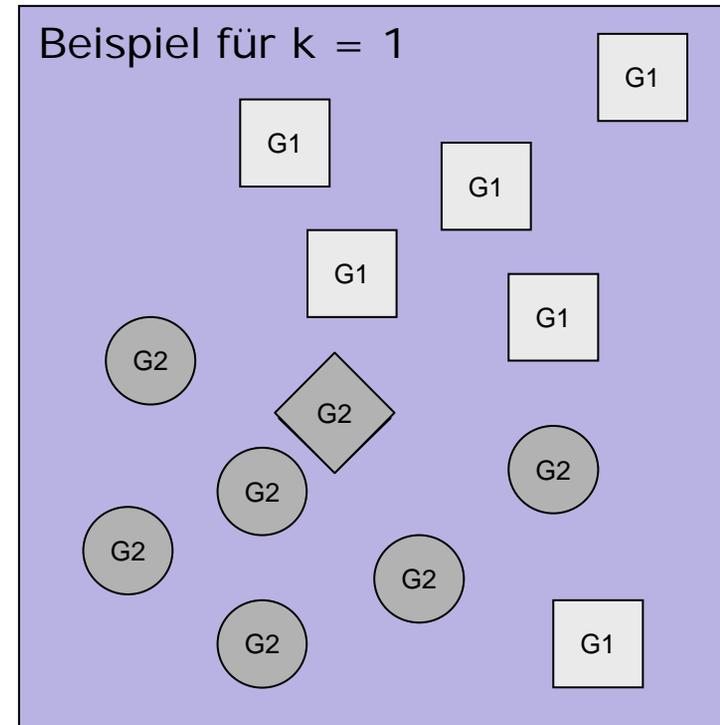
2. Neuronale Netze – Eigenschaften

- + Hohe Lernfähigkeit
- + Hohe Parallelität bei der Informationsverarbeitung
- + Hohe Fehlertoleranz
- + Flexible Struktur, die durch Hinzufügen von versteckten Schichten erweiterbar ist.
- Ergebnisse nur wenig transparent.
- Setup des Netzes erfordert eine Vielzahl von Entscheidungen (z.B. Anzahl der Schichten, Anzahl der Neuronen pro Schicht).

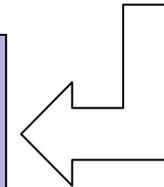
III. Diskriminanzanalytische Verfahren

3. k-Nearest-Neighbor

- Überlegung:
Benachbarte Objekte besitzen häufig ähnliche Merkmalsausprägungen und gehören derselben Gruppe an.
- Regel:
Ordne ein Objekt derjenigen Gruppe zu, aus der die Mehrzahl seiner k Nachbarn stammt.
- Die Nachbarn werden auf Basis der Trainingsstichprobe ermittelt.



Aufgrund seiner Nähe zu einem Objekt der Gruppe G2 wird das neue Objekt (Raute) nach dem kNN-Verfahren der Gruppe G2 zugeordnet.



III. Diskriminanzanalytische Verfahren

3. k-Nearest-Neighbor – Ablauf des Verfahrens

- Abstandsberechnung:
Für jedes Objekt y_i aus der Stichprobe wird der Abstand $D(x, y_i)$ zum Merkmalsvektor x des zu klassifizierenden Objekts berechnet.

Beispiel für ein Distanzmaß:

$$D(x, y_i) = \sqrt{\sum_{j=1}^p (x_j - y_{ij})^2}$$

- Auswahl der k Nachbarn:
Es werden diejenigen k Objekte aus der Stichprobe ausgewählt, die zu x den kleinsten Abstand besitzen.
- Klassifikation des Objekts:
Das Objekt wird derjenigen Gruppe zugeordnet, aus der die Mehrzahl der k Trainingsobjekte stammen.

III. Diskriminanzanalytische Verfahren

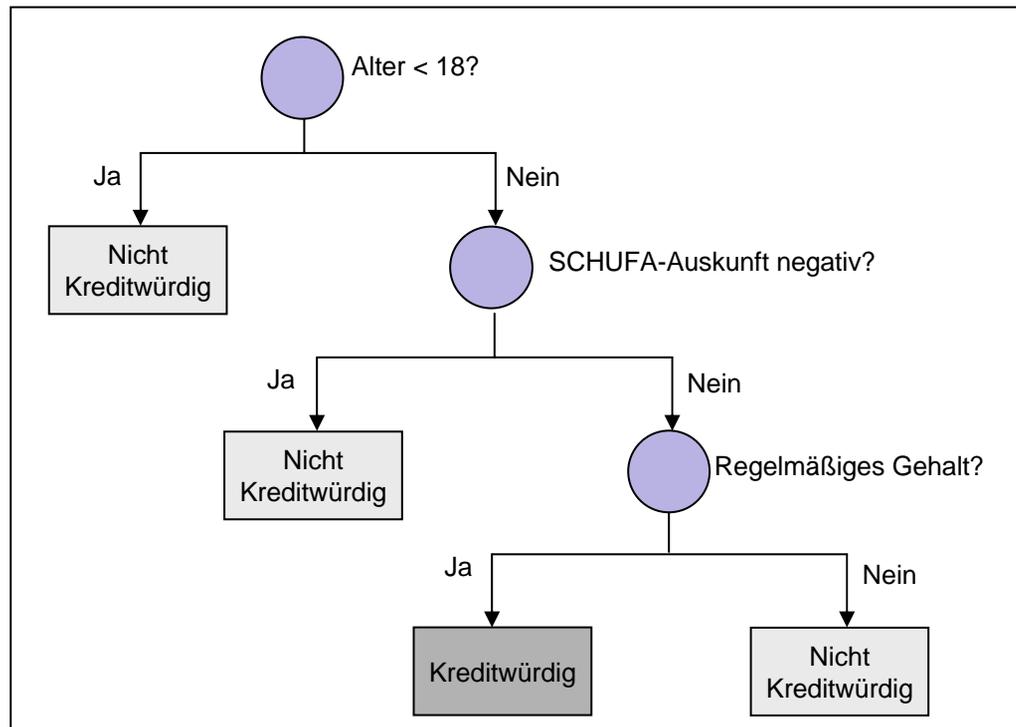
3. k-Nearest-Neighbor – Eigenschaften

- + Lernphase entfällt, da die Auswertung der Stichprobe erst zur Klassifikationszeit erfolgt (instanzbasiertes Verfahren).
- + Das Verfahren erweitert mit jedem neu klassifizierten Objekt seine Datenbasis, es lernt kontinuierlich.
- Alle Trainingsobjekte müssen im Speicher verfügbar gehalten werden.
- Die Auswertung zur Klassifikationszeit erfordert einen hohen Berechnungsaufwand.
- Ergebnisse werden nicht gespeichert, daher muss stets die gesamte Stichprobe ausgewertet werden.

III. Diskriminanzanalytische Verfahren

4. CART - Classification and Regression Tree

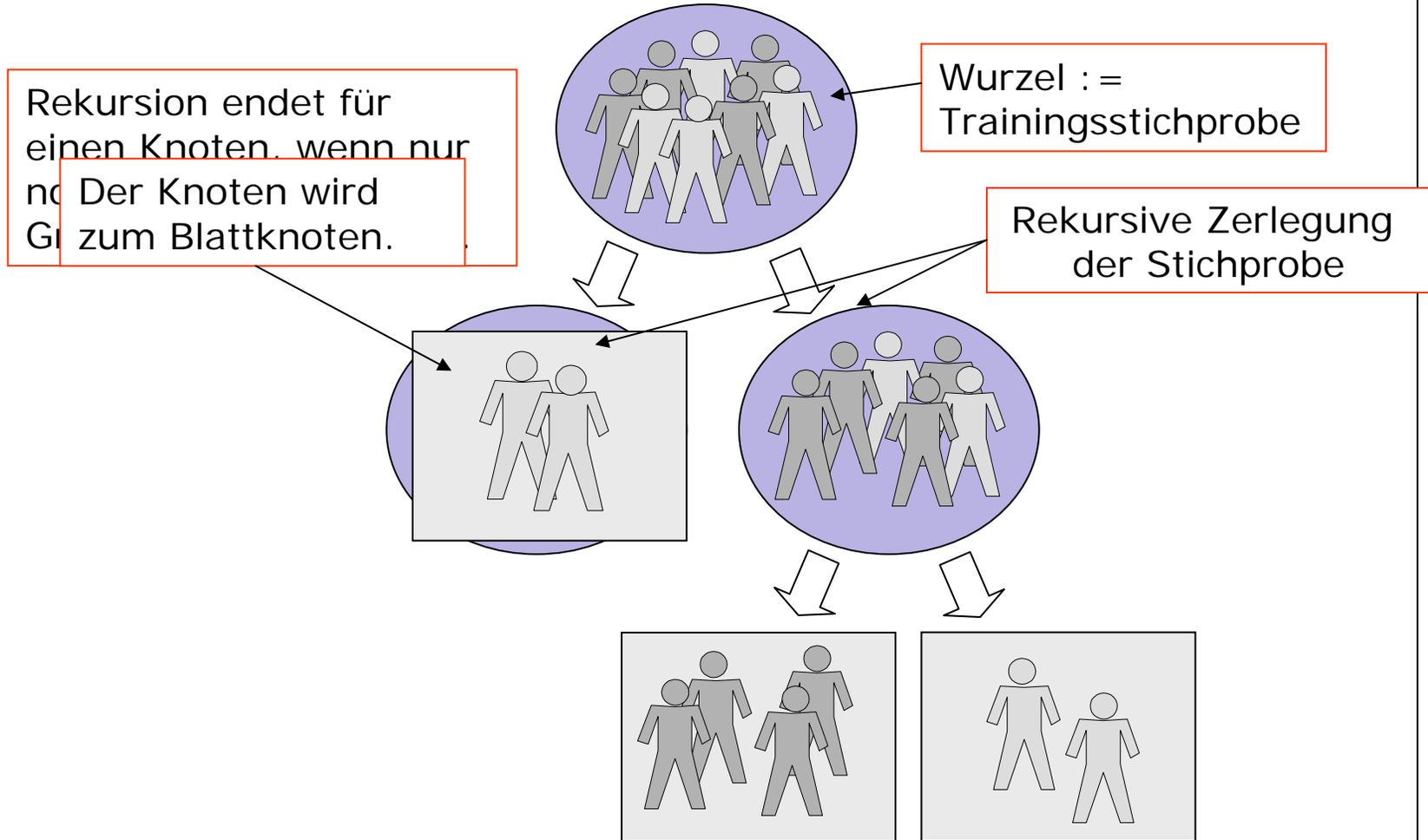
- Auf Basis der Trainingsstichprobe wird zunächst ein Entscheidungsbaum konstruiert und dieser anschließend zur Klassifikation neuer Fälle eingesetzt.



Beispiel eines CART für eine Kreditwürdigkeitsprüfung

III. Diskriminanzanalytische Verfahren

4. CART – Konstruktion



III. Diskriminanzanalytische Verfahren

4. CART – Aufteilung eines Knotens

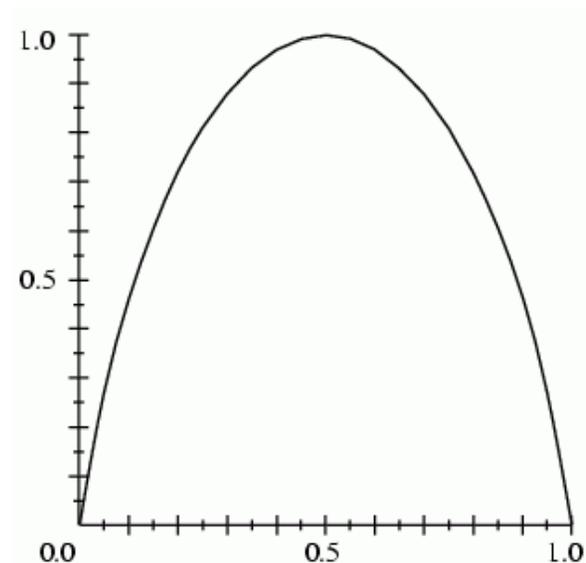
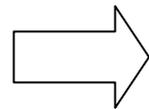
- Erfolgt durch eine Zerlegungsvorschrift („Split“).
- Ein Split ordnet jedes Objekt eines Knotens t anhand der Merkmalsausprägung x_j entweder einem linken Kindknoten t_l oder einem rechten Kindknoten t_r zu.
- Die Form eines Splits ist von der Skalierung des zugrunde liegenden Merkmals abhängig.
- Beispiele für Splits:
 - „Alter < 18 ?“
 - „Schufa-Auskunft negativ?“
 - „Geschlecht männlich?“
 - etc.

III. Diskriminanzanalytische Verfahren

4. CART – Ziel einer Aufteilung

- Das Ziel einer Zerlegung des Knotens t besteht darin, möglichst homogene Kindknoten zu erzeugen.
- Ein Knoten ist homogen, wenn er nur Objekte mit derselben Gruppenzugehörigkeit beinhaltet.
- Als Maß für die Homogenität eines Knotens wird z.B. der Gini-Index verwendet.

$$gini(t) = 1 - \sum_{i=1}^k p_i^2$$



III. Diskriminanzanalytische Verfahren

4. CART – Auswahl einer geeigneten Aufteilungsvorschrift

- Für jeden Knoten existiert eine Vielzahl möglicher Splits, die in der Menge S zusammengefasst werden.
- Die Aufgabe des Algorithmus besteht darin, aus S denjenigen Split auszuwählen, der die größtmögliche Veränderung der Homogenität im Baum bewirkt.
- Diese Veränderung lässt sich für einen Split s wie folgt berechnen:

$$\Delta gini(s, t) = gini(t) - \frac{|t_l|}{|t|} gini(t_l) - \frac{|t_r|}{|t|} gini(t_r)$$

- Für die Wahl eines Splits gilt demnach:

$$s^* = \underset{s \in S}{MAX} (\Delta gini(s, t))$$

III. Diskriminanzanalytische Verfahren

4. CART – Maximalbaum und Zuordnung der Blattknoten

- Der auf Basis des Algorithmus erzeugte Baum wird als Maximalbaum bezeichnet.
- Jeder Blattknoten repräsentiert darin eine der beiden Gruppen, wobei die Zuordnung über die Bayes Regel erfolgt:

Ordne den Knoten t derjenigen Gruppe G_i (mit $i = 1,2$) zu, für die gilt:

$$\lambda_{\text{Bayes}}(t) = \begin{cases} G_1 & \text{wenn } P(G_1 | t) \geq P(G_2 | t) \\ G_2 & \text{wenn } P(G_1 | t) < P(G_2 | t) \end{cases}$$

III. Diskriminanzanalytische Verfahren

4. CART – Stutzen des Maximalbaums („Pruning“)

- Der Maximalbaum ist im Allgemeinen zu stark an die zufällige Struktur der Trainingsstichprobe angepasst („Overfitting“).
- Um die Abhängigkeit von der Stichprobe zu reduzieren, wird der Maximalbaum in seiner Größe verringert („Pruning“).
- Dazu werden schrittweise schwach besetzte Knoten aus dem Baum entfernt und eine Folge von Teilbäumen mit abnehmender Komplexität erzeugt.
- Aus dieser Folge wird derjenige Baum ausgewählt, der im Hinblick auf eine Teststichprobe die kleinste Fehlerrate besitzt bzw. die Fehlklassifikationskosten minimiert.
- Eine detaillierte Beschreibung des Pruning-Algorithmus findet sich bei:

Schillinger, Marco Patrick, 2002: Flächenhafte Schätzung mit Classification and Regression Trees und robuste Gütebestimmung ökologischer Parameter in einem kleinen Einzugsgebiet, Internet <http://opus.ub.uni-bayreuth.de/volltexte/2002/7/>

III. Diskriminanzanalytische Verfahren

4. CART – Eigenschaften

- + Einfache Anwendbarkeit: Nachdem der Baum konstruiert ist, lassen sich neue Objekte schnell und effizient klassifizieren.
- + Die Zuordnung ist transparent und somit leicht interpretierbar.
- Relativ hoher Aufwand bei der Konstruktion eines CART.
- Baumstruktur ist stark von der zugrunde liegenden Stichprobe abhängig.

I. Einführung

1. Problemstellung
2. Zielsetzung der Präsentation

II. Grundlagen der Diskriminanzanalyse

1. Phasen der Diskriminanzanalyse
2. Auswahlkriterien für eine Diskriminanzregel
 - a) Arten von Merkmalen
 - b) Beurteilung der Klassifikationsgüte

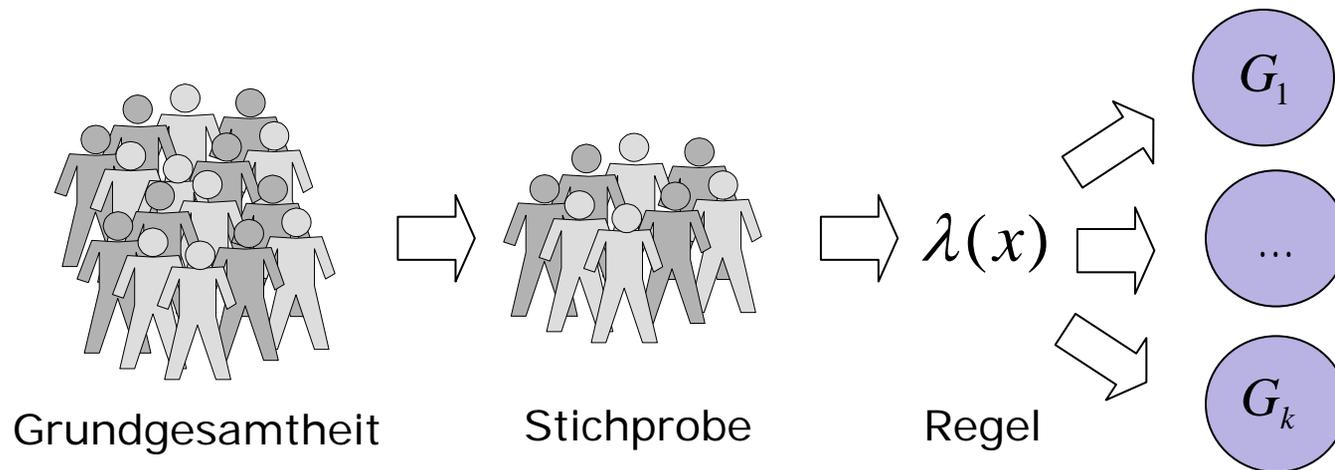
III. Diskriminanzanalytische Verfahren

1. Bayes-Klassifikator
2. Neuronale Netze
3. K-Nearest-Neighbor
4. CART

IV. Zusammenfassung

IV. Zusammenfassung

- In dieser Präsentation wurden die Grundlagen der Diskriminanzanalyse erläutert und einige ausgewählte Klassifikationsverfahren vorgestellt.
- Grundsätzliches Vorgehen:



- Problem:
 Die Regel muss auf Basis einer Stichprobe so konstruiert werden, dass sie auch neue Objekte mit unbekannter Gruppenzugehörigkeit möglichst korrekt klassifiziert.

IV. Zusammenfassung

- Drei der vorgestellten Verfahren werten die Trainingsstichprobe in einer Lernphase zunächst aus und generieren Wissen in Form von:
 - Wahrscheinlichkeitsverteilungen (Bayes-Klassifikator)
 - Gewichtete Verbindungen (Neuronale Netze)
 - Entscheidungsbaum (CART)
- Bei dem k-Nearest-Neighbor-Verfahren entfällt die Lernphase. Die Auswertung der Trainingsstichprobe erfolgt zur Klassifikationszeit.
- Die Auswahl eines Verfahrens ist von zahlreichen Kriterien abhängig:
 - Arten der betrachteten Merkmale
 - Fehlerrate/Fehlklassifikationskosten
 - Geschwindigkeit der Klassifikation
 - ...
- Welche Kriterien letztlich allerdings für die Wahl den Ausschlag geben, ist von Fall zu Fall verschieden.