

FACHHOCHSCHULE WEDEL

Seminararbeit Informatik

in der Fachrichtung

Wirtschaftsinformatik

Themenbereich „Künstliche Intelligenz“

Thema Nr. 3

Diskriminanzanalyse

Eingereicht von: Patrick Wolf
Wedeler Landstr. 65
22559 Hamburg
Tel. (040) 600 877 33

Erarbeitet im: 6. Semester

Abgegeben am: 19.05.2005

Referent (FH Wedel): Prof. Dr. Sebastian Iwanowski
Fachhochschule Wedel
Feldstraße 143
22880 Wedel

Inhaltsverzeichnis

Inhaltsverzeichnis	I
Abbildungsverzeichnis	II
1. Einführung	1
1.1. Problemstellung.....	1
1.2. Gliederung der Arbeit	2
2. Grundlagen der Diskriminanzanalyse.....	3
2.1. Arten von Merkmalen	3
2.1.1. Quantitative Merkmale	4
2.1.2. Qualitative Merkmale	4
2.2. Beurteilung der Klassifikationsgüte	4
2.2.1. Fehlerrate	5
2.2.2. Fehlklassifikationskosten.....	6
3. Diskriminanzanalytische Verfahren	8
3.1. Bayes-Klassifikator.....	8
3.1.1. Funktionsweise	8
3.1.2. Eigenschaften	10
3.2. Neuronale Netze.....	11
3.2.1. Funktionsweise	11
3.2.2. Eigenschaften	16
3.3. k-Nearest-Neighbor.....	16
3.3.1. Funktionsweise	16
3.3.2. Eigenschaften	18
3.4. CART.....	18
3.4.1. Funktionsweise	19
3.4.2. Eigenschaften	22
4. Zusammenfassung	24
Literaturverzeichnis	25

Abbildungsverzeichnis

Abbildung 1: Beispielstruktur eines Neuronalen Netzes	12
Abbildung 2: Informationsverarbeitungsprozess innerhalb eines Neurons.....	12
Abbildung 3: Kurvenverläufe ausgewählter Aktivierungsfunktionen.....	14
Abbildung 4: Phasen des Lernprozesses in einem NN	15
Abbildung 5: k-Nearest-Neighbor-Verfahren (k=1).....	18
Abbildung 6: Beispiel eines CART für eine Kreditwürdigkeitsprüfung	19
Abbildung 7: Verlauf des Gini-Index (k=2)	21

1. Einführung

1.1. Problemstellung

Eine Bank ist bei der Kreditvergabe daran interessiert, möglichst nur denjenigen Kunden ein Darlehen zu gewähren, von denen sie eine problemlose Rückzahlung der Verbindlichkeiten erwarten kann. Aus diesem Grund wird sie für jeden Antragsteller zunächst eine Kreditwürdigkeitsprüfung durchführen, innerhalb derer anhand bestimmter Merkmale (z.B. Höhe des Gehalts, bisheriges Zahlungsverhalten) entschieden wird, ob der jeweilige Antragsteller der Gruppe der kreditwürdigen oder der Gruppe der nicht kreditwürdigen Kunden zuzuordnen ist. Auf Basis dieser Zuordnung kann anschließend eine Entscheidung über die Gewährung oder die Ablehnung des Kredits erfolgen.

Das Problem bei der Klassifikation von Antragstellern, ebenso wie bei der Klassifikation beliebiger anderer Objekte, besteht allerdings darin, diejenigen Merkmale auszuwählen, die eine signifikante Trennung zwischen den einzelnen Gruppen ermöglichen. Eine weitere Schwierigkeit stellt die Bestimmung der Klassifikationsregel dar, mit der die Zuordnung der Objekte zu den einzelnen Gruppen erfolgt.

Die aufgeführten Probleme bilden den Ausgangspunkt der Diskriminanzanalyse, die auf eine Arbeit von R.A. Fisher aus dem Jahre 1936 zurückgeht.¹ Dabei handelt es sich um ein multivariates Verfahren der Statistik², das sich mit der Untersuchung von Gruppenunterschieden sowie der Vorhersage der Gruppenzugehörigkeit von Objekten beschäftigt. Im Gegensatz zur Clusteranalyse werden bei der Diskriminanzanalyse keine neuen Gruppen gebildet, sondern nur bereits bestehende betrachtet.³

Ziel dieser Arbeit ist es, die Grundlagen der Diskriminanzanalyse zu erläutern und die Verfahren zur Vorhersage der Gruppenzugehörigkeit vorzustellen. Zum besseren Verständnis beschränkt sich die Darstellung dabei auf den Fall zweier Gruppen, der auch in der Praxis häufig anzutreffen ist. Auf eine Erläuterung des

¹ Vgl. Fisher, R.A., The use of multiple measurements, 1936.

² „Die multivariate Statistik untersucht die Beziehungsstrukturen zwischen vielen Variablen.“, Diaz-Bone, Rainer, Multiple Lineare Regression, 2005.

³ Vgl. Stahel, Werner, Clusteranalyse, o.J.

Vorgehens für die Untersuchung von Gruppenunterschieden (Merkmalsauswahl) wird im Rahmen der Ausarbeitung verzichtet und stattdessen auf die Arbeiten von *Trampisch*⁴ oder *Hupfeld*⁵ verwiesen.

1.2. Gliederung der Arbeit

Das zweite Kapitel behandelt die Grundlagen der Diskriminanzanalyse. Nach einer kurzen Beschreibung des prinzipiellen Vorgehens bei der Klassifikation werden Kriterien zur Beurteilung von Diskriminanzregeln formuliert und diese in den folgenden beiden Abschnitten näher erläutert. Hierbei wird zunächst auf die verschiedenen Arten von Merkmalen eingegangen und anschließend ein geeignetes Maß für die Beurteilung der Klassifikationsgüte definiert.

Im dritten Kapitel, dem eigentlichen Hauptteil der Arbeit, werden die aus Sicht des Autors wesentlichen Klassifikationsverfahren mit ihren jeweiligen Eigenschaften vorgestellt. Bei diesen Verfahren handelt es sich um den Bayes-Klassifikator, Neuronale Netze, k-Nearest-Neighbor und CART. Den Abschluss der Arbeit bildet eine Zusammenfassung der wesentlichen Aussagen aus den vorangegangenen Kapiteln.

⁴ Vgl. Trampisch, H.J, Diskriminanzanalyse, 1985, S. 57-132.

⁵ Vgl. Hupfeld, Jörg, Einführung in die Diskriminanzanalyse, 1998.

2. Grundlagen der Diskriminanzanalyse

Im Rahmen der Diskriminanzanalyse werden Verfahren gesucht, mit denen die Gruppenzugehörigkeit von bislang unbekanntem Objekten möglichst zuverlässig vorhergesagt werden kann. Dabei lässt sich das grundsätzliche Vorgehen in zwei Phasen unterteilen:

1. **Lernphase:** In der Lernphase wird aus der zu klassifizierenden Grundgesamtheit Ω zunächst eine Stichprobe entnommen und für jedes darin befindliche Objekt die tatsächliche Gruppenzugehörigkeit sowie die p Ausprägungen $x = (x_1, \dots, x_p)$ in den Merkmalen X_1, \dots, X_p ermittelt.⁶ Anschließend wird eine geeignete Zuordnungsregel λ gesucht, die die Objekte aus der Stichprobe anhand des objektspezifischen Merkmalsvektors x genau einer der k vordefinierten Gruppen G_1, \dots, G_k so zuordnet, dass diese Zuordnung mit der tatsächlichen Gruppenzugehörigkeit des Objekts übereinstimmt.
2. **Klassifikationsphase:** Die Klassifikationsphase zeichnet sich durch die Anwendung der ermittelten Diskriminanzregel auf die unbekanntem Objekte in der Grundgesamtheit aus, so dass jedes Objekt aus Ω klassifiziert wird.

Für die Auswahl eines Klassifikationsverfahrens spielen die zugrunde liegenden Merkmalsarten sowie die Genauigkeit, mit der die Zuordnung der Objekte zu den einzelnen Gruppen erfolgt, eine wichtige Rolle.⁷ Daneben sind u.a. die Geschwindigkeit der Klassifikation, die Stabilität gegenüber Ausreißern und die Interpretierbarkeit der Ergebnisse für die Beurteilung von Interesse, wobei diese Kriterien im Folgenden nicht weiter behandelt werden sollen.

2.1. Arten von Merkmalen

Die bei der Klassifikation betrachteten Merkmale X_1, \dots, X_p können hinsichtlich ihrer möglichen Ausprägungen in quantitative und qualitative Merkmale unterschieden werden.

⁶ Vgl. Trampisch, H.J., Diskriminanzanalyse, 1985, S.57.

⁷ Vgl. Nothnagel, Michael, Diskriminanzanalyse, 1999, S27.

2.1.1. Quantitative Merkmale

Quantitative Merkmale besitzen als Ausprägungen nur reelle Zahlen. Die Abstände zwischen den einzelnen Ausprägungen lassen sich berechnen und sinnvoll interpretieren (z.B. Semesterzahl, Körpergröße, Alter, Geldbeträge).⁸ Häufig erfolgt eine weitere Unterteilung der quantitativen Merkmale in diskrete und stetige Größen. Erstere können in einem bestimmten Intervall höchstens abzählbar viele Ausprägungen annehmen (z.B. Semesterzahl), letztere hingegen überabzählbar viele (z.B. Alter, Geldbeträge).⁹

2.1.2. Qualitative Merkmale

Die Ausprägungen von qualitativen Merkmalen sind keine reellen Zahlen (z.B. Geschlecht, Geburtsort), dementsprechend ist eine Interpretation der Abstände nicht sinnvoll. Qualitative Merkmale lassen sich in ordinale und nominale Größen unterscheiden.¹⁰ Bei nominalen Merkmalen existiert zwischen den Ausprägungen keine Ordnung (z.B. Geschlecht). Es ist lediglich feststellbar, ob sie gleich oder verschieden sind. Die Ausprägungen von ordinalskalierten Merkmalen können hingegen in eine Rangfolge gebracht werden, wobei ebenfalls keine Aussage über die Abstände möglich ist (z.B. Qualitätskategorien).

In Abhängigkeit von der Art der Merkmale ist eine Unterscheidung der Klassifikationsverfahren in Verfahren für ausschließlich qualitative bzw. quantitative Merkmale und in solche für gemischt skalierte (d.h. qualitative und quantitative) Merkmale möglich. Im Rahmen dieser Ausarbeitung werden nur Verfahren für gemischt skalierte Merkmale betrachtet.

2.2. Beurteilung der Klassifikationsgüte

Das Ziel einer Klassifikationsregel λ besteht darin, die unbekanntes Objekte aus der Grundgesamtheit möglichst korrekt zu klassifizieren, d.h. ihre tatsächliche Gruppenzugehörigkeit mit hoher Wahrscheinlichkeit richtig vorherzusagen. Dabei ist mit jeder Klassifikation das Risiko einer Fehlentscheidung verbunden. Im hier betrachteten Fall zweier Gruppen G_1 und G_2 sind zwei Arten von Fehl-

⁸ Vgl. Löwe, Matthias, Einführung in die elementare Stochastik, 2005, S5f.

⁹ Vgl. Weihs, Claus, Diskriminanzanalyse, 2000, S.24.

¹⁰ Vgl. Nothnagel, Michael, Diskriminanzanalyse, 1999, S.18f.

entscheidungen möglich: Die Regel λ ordnet ein Objekt aus der Gruppe G_1 irrtümlich der Gruppe G_2 zu oder umgekehrt.

2.2.1. Fehlerrate

Die Güte einer Klassifikationsregel bestimmt sich anhand ihrer Fehlerrate $E(\lambda)$, die als das Verhältnis zwischen der Anzahl der falsch klassifizierten Objekte und der Anzahl der insgesamt klassifizierten Objekte definiert ist:¹¹

$$E(\lambda) = \frac{\text{Anzahl falsch klassifizierter Objekte}}{\text{Anzahl klassifizierter Objekte}} \quad (2.1)$$

Die mit Hilfe von (2.1) berechnete empirische Fehlerrate ist stark von der Wahl der zugrunde liegenden Stichprobe abhängig. Für unterschiedliche Stichproben können sich demnach stark differierende Fehlerraten ergeben, die sich entsprechend auf die Beurteilung der Klassifikationsregel auswirken. Dabei wird die empirische Fehlerrate mit wachsendem Stichprobenumfang gegen die wahre Fehlerrate in der Grundgesamtheit streben. Um jedoch auch bei einer relativ kleinen Stichprobe eine Aussage über die wahre Fehlerrate einer Regel machen zu können, wurden eine Reihe von Schätzmethoden entwickelt, von denen im Folgenden drei erläutert werden sollen:

1. **Resubstitutionsmethode:** Die Resubstitutionsmethode stellt das einfachste Verfahren zur Fehlerratenschätzung dar. Dabei wird die Klassifikationsregel zunächst auf alle Objekte aus der Trainingsstichprobe angewandt und anschließend der Anteil der falsch klassifizierten Trainingsobjekte bestimmt. In der Praxis erweist sich die ermittelte Fehlerrate jedoch als schlechter Schätzer für die wahre Fehlerrate, insbesondere bei kleinen Stichproben.¹² Die Ursache hierfür liegt in der Verwendung der Trainingsstichprobe sowohl zur Bestimmung der Zuordnungsregel als auch zur Ermittlung der Fehlerrate. Durch die Überanpassung der Regel an die Trainingsstichprobe ergibt sich dadurch im Allgemeinen eine zu optimistische Schätzung für die wahre Fehlerrate.¹³

¹¹ Vgl. Weihs, Claus, Diskriminanzanalyse, 2000, S. 596.

¹² Vgl. Weihs, Claus, Diskriminanzanalyse, 2000, S. 596.

¹³ Vgl. Trampisch, H.J., Diskriminanzanalyse, 1985, S. 102f.

2. **Hold-Out-Methode:** Bei der Hold-Out-Methode wird die Trainingsstichprobe in einen Lern- und einen Testdatensatz aufgeteilt. Ersterer dient zur Konstruktion der Klassifikationsregel und letzterer zur Ermittlung der Fehlerrate. Diese Trennung vermeidet die Verfälschung der Fehlerrate und führt bereits ab einem 1000 Objekte umfassenden Testdatensatz zu einer guten Approximation der wahren Fehlerrate, sofern die statistische Unabhängigkeit von Lern- und Trainingsdatensatz gegeben ist.¹⁴ Als Nachteil der Hold-Out-Methode erweist sich, dass nur ein Teil der Trainingsstichprobe als Lerndatensatz verwendet wird und somit „wertvolle Information für die Bestimmung der Klassifikationsregel“¹⁵ verloren geht.
3. **Leaving-one-out-Methode:** Der Leaving-one-out-Methode liegt ebenfalls eine Aufteilung der Trainingsstichprobe in einen Lern- und einen Testdatensatz zugrunde. Im Gegensatz zur Hold-out-Methode handelt es sich hierbei allerdings nicht um eine statische Aufteilung der n Objekte aus der Trainingsstichprobe, sondern vielmehr um einen dynamischen Prozess, der sich in drei Phasen unterteilt: In der ersten Phase werden aus der Stichprobe $n - 1$ Objekte als Lerndatensatz entnommen und zur Bestimmung der Klassifikationsregel eingesetzt. Das in der Stichprobe verbleibende n -te Objekt bildet den Testdatensatz, der in der zweiten Phase zur Überprüfung der Klassifikationsregel verwendet wird. Die dritte Phase stößt schließlich die Wiederholung der vorangegangenen Phasen an. Der Algorithmus terminiert, wenn jedes der n Trainingsobjekte als Testdatensatz verwendet wurde. Die Fehlerrate ergibt sich, indem die Anzahl der falsch zugeordneten Objekte durch n dividiert wird. *Trampisch* empfiehlt die Methode „als Standardmethode zur Fehlerratschätzung“¹⁶, da sie im Mittel einen richtigen Schätzwert liefert.¹⁷

2.2.2. Fehlklassifikationskosten

Bei der bisherigen Betrachtung wurde implizit davon ausgegangen, dass die Fehlentscheidung, ein Objekt aus der Gruppe G_1 der Gruppe G_2 zuzuordnen, die gleiche Bedeutung besitzt, wie die Fehlentscheidung, ein Objekt aus der Gruppe G_2

¹⁴ Vgl. Weihs, Claus, Diskriminanzanalyse, 2000, S. 597.

¹⁵ Weihs, Claus, Diskriminanzanalyse, 2000, S. 597.

¹⁶ Trampisch, H.J., Diskriminanzanalyse, 1985, S. 107.

der Gruppe G_1 zuzuordnen. *Bonne* und *Armingier* führen jedoch an, dass diese Annahme in der Praxis häufig nicht erfüllt ist.¹⁸

*„Beispielspielsweise 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“.*¹⁹

Um der unterschiedlichen Bedeutung der beiden Risikoarten Rechnung zu tragen, können so genannte Fehlklassifikationskosten $C(i, j) > 0$ für falsche und $C(i, j) = 0$ für richtige Zuordnungen eingeführt werden. Der Index i bezeichnet hierbei die tatsächliche und der Index j die prognostizierte Gruppenzugehörigkeit eines Objekts.²⁰ Das Ziel einer Klassifikationsregel besteht dann nicht mehr in der Minimierung der Fehlerrate, sondern vielmehr in der Minimierung der Fehlklassifikationskosten.²¹ Allerdings sind diese Kosten oftmals schwer zu bestimmen und entsprechende Quantifizierungen eher subjektiver Natur. Daher werden sie „in der Praxis häufig als identisch angenommen“²², obwohl die Ungleichheit der zugrunde liegenden Risikoarten offensichtlich ist.

¹⁷ Vgl. Trampisch, H.J, Diskriminanzanalyse, 1985, S. 107.

¹⁸ Vgl. Bonne, Thorsten / Armingier, Gerhard, Diskriminanzanalyse, 2001, S. 197.

¹⁹ Bonne, Thorsten / Armingier, Gerhard, Diskriminanzanalyse, 2001, S. 197.

²⁰ Vgl. Bonne, Thorsten / Armingier, Gerhard, Diskriminanzanalyse, 2001, S. 197.

²¹ Vgl. Bao, Ho Tu, Knowledge Engineering, o.J.

²² Weihs, Claus, Diskriminanzanalyse, 2000, S.590.

3. Diskriminanzanalytische Verfahren

Für die Klassifikation von Objekten existiert eine Vielzahl diskriminanzanalytischer Verfahren, die sich insbesondere in der Herangehensweise an die Problemstellung unterscheiden. Im Rahmen dieses Kapitels sollen vier dieser Verfahren näher erläutert werden, wobei deren grundsätzliche Funktionsweise und ihre wesentlichen Eigenschaften im Mittelpunkt der Betrachtungen stehen.

3.1. Bayes-Klassifikator

Das Bayes-Verfahren basiert auf wahrscheinlichkeitstheoretischen Überlegungen und ist besonders wegen seiner hohen Geschwindigkeit und Präzision für die Klassifikation großer Datenmengen von Interesse.²³ Allerdings setzt es bestimmte Kenntnisse über die Verteilung der Gruppen sowie der darin befindlichen Objekte voraus, die in der Praxis meist nicht vorhanden sind oder nur sehr aufwändig beschafft werden können.

3.1.1. Funktionsweise

Der Bayes-Klassifikator $\lambda_{\text{Bayes}}(x)$ ordnet ein Objekt derjenigen Gruppe G_i (mit $i = 1, \dots, k$) zu, welche für den objektspezifischen Merkmalsvektor x die größte Auftretswahrscheinlichkeit $P(G_i | x)$ besitzt. Im Falle zweier Gruppen ($k = 2$) gilt:²⁴

$$\lambda_{\text{Bayes}}(x) = \begin{cases} G_1 & \text{wenn } P(G_1 | x) \geq P(G_2 | x) \\ G_2 & \text{wenn } P(G_1 | x) < P(G_2 | x) \end{cases} \quad (3.1)$$

Die Auftretswahrscheinlichkeit $P(G_i | x)$ kann mit Hilfe des Theorems von Bayes bestimmt werden:

$$P(G_i | x) = \frac{\pi_i P(x | G_i)}{\sum_{j=1}^k \pi_j P(x | G_j)} \quad (3.2)$$

²³ Vgl. Pretzer, Michael, Clustering und Klassifikation, 2003, S. 22.

²⁴ Vgl. Bonne, Thorsten / Arminger, Gerhard, Diskriminanzanalyse, 2001, S. 196.

Da es sich bei dem Nenner um eine konstante Größe für alle Gruppen handelt, ist für die Zuordnung eines Objekts allein der Zähler $\pi_i P(x | G_i)$ ausschlaggebend. Um diesen berechnen zu können, müssen die beiden Faktoren π_i und $P(x | G_i)$ anhand der Trainingsstichprobe T geschätzt werden. Der Faktor π_i bezeichnet hierbei die Wahrscheinlichkeit für das Auftreten der Gruppe G_i in der Grundgesamtheit und kann durch einfaches Auszählen von T ermittelt werden:²⁵

$$\pi_i = \frac{\text{Anzahl Objekte aus } G_i \text{ in } T}{\text{Anzahl aller Objekte in } T} \quad (3.3)$$

Demgegenüber gestaltet sich die Schätzung von $P(x | G_i)$ deutlich schwieriger. Da sich der Merkmalsvektor x aus den p Ausprägungen der Merkmale X_1, \dots, X_p zusammensetzt, muss die Wahrscheinlichkeit für jede Kombination der Merkmalsausprägungen bestimmt werden. Dies ist aufgrund des hierfür erforderlichen Stichprobenumfangs in der Praxis kaum möglich. Um dennoch $P(x | G_i)$ berechnen zu können, wird im Allgemeinen die Unabhängigkeit der Merkmale unterstellt, so dass sich $P(x | G_i)$ aus der Multiplikation der bedingten Auftretswahrscheinlichkeiten für die Merkmalsausprägungen x_1, \dots, x_p in der Gruppe G_i ergibt:²⁶

$$P(x | G_i) = \prod_{j=1}^p P(x_j | G_i) \quad (3.4)$$

Die Wahrscheinlichkeit $P(x_j | G_i)$ lässt sich wiederum durch Auszählen der Trainingsstichprobe schätzen:²⁷

$$P(x_j | G_i) = \frac{\text{Anzahl Objekte aus } G_i \text{ in } T \text{ mit Ausprägung } x_j}{\text{Anzahl Objekte aus } G_i \text{ in } T} \quad (3.5)$$

²⁵ Vgl. Wiehr, Stephan, Automatische Klassifikation, 2002; ebenso: Pretzer, Michael, Clustering und Klassifikation, 2003, S. 23.

²⁶ Vgl. Wiehr, Stephan, Automatische Klassifikation, 2002; ebenso: Pretzer, Michael, Clustering und Klassifikation, 2003, S. 24.

²⁷ Vgl. Wiehr, Stephan, Automatische Klassifikation, 2002; ebenso: Pretzer, Michael, Clustering und Klassifikation, 2003, S. 24.

Bei kleinen Stichproben besteht allerdings die Gefahr, dass keines der Objekte aus der Gruppe G_i die Ausprägung x_j besitzt. In diesem Fall gilt $P(x_j | G_i) = 0$ und somit nimmt auch $P(x | G_i)$ den Wert 0 an. Um dieses Problem zu vermeiden, wird $P(x_j | G_i)$ in der Praxis häufig noch geglättet.²⁸

Der Bayes-Klassifikator besitzt eine minimale Fehlerrate, da er ein Objekt stets der Gruppe mit der größten Auftretswahrscheinlichkeit für den objektspezifischen Merkmalsvektor zuordnet.²⁹ Die Fehlerrate lässt sich im Falle zweier Gruppen wie folgt berechnen:³⁰

$$E(\lambda_{\text{Bayes}}) = \pi_1 P(\lambda_{\text{Bayes}} = G_2 | G_1) + \pi_2 P(\lambda_{\text{Bayes}} = G_1 | G_2) \quad (3.6)$$

Dabei bezeichnet $P(\lambda_{\text{Bayes}} = G_2 | G_1)$ die Wahrscheinlichkeit, mit der ein Objekt aus der Gruppe G_1 falsch klassifiziert wird und $P(\lambda_{\text{Bayes}} = G_1 | G_2)$ die analoge Wahrscheinlichkeit für die fehlerhafte Zuordnung eines Objekts aus der Gruppe G_2 .

3.1.2. Eigenschaften

Die wesentlichen Eigenschaften des Bayes-Klassifikators wurden bereits in der Einleitung zu diesem Abschnitt erläutert. Hierzu zählen insbesondere die hohe Geschwindigkeit des Verfahrens und dessen Präzision bei der Klassifikation von großen Datenmengen.³¹ Dem stehen die Nachteile gegenüber, dass die Annahme der Merkmalsunabhängigkeit im Allgemeinen nicht zutrifft (z.B. ist das Merkmal „Einkommen“ vom Merkmal „Beruf“ abhängig) und für die verlässliche Schätzung der Wahrscheinlichkeiten ein sehr großer Stichprobenumfang erforderlich ist.³²

²⁸ Vgl. Wiehr, Stephan, Automatische Klassifikation, 2002.

²⁹ Vgl. Bonne, Thorsten / Arminger, Gerhard, Diskriminanzanalyse, 2001, S. 196.

³⁰ Vgl. Trampisch, H.J., Diskriminanzanalyse, 1985, S. 67.

³¹ Vgl. Pretzer, Michael, Clustering und Klassifikation, 2003.

³² Vgl. Weihs, Claus, Diskriminanzanalyse, 2000, S. 591.

3.2. Neuronale Netze

Neuronale Netze simulieren die Funktionsweise des Gehirns und bestehen aus einer Vielzahl einfacher Einheiten, den so genannten Neuronen, die über gerichtete und gewichtete Verbindungen untereinander Informationen austauschen.³³

Im Hinblick auf die zulässige Flußrichtung der Informationen zwischen den Neuronen können Netze ohne Rückkoppelung (Feed-Forward-Netze) und Netze mit Rückkoppelung (Feed-Backward-Netze) unterschieden werden. Bei ersteren dürfen Informationen stets nur in eine Richtung fließen, bei letzteren ist der Informationsfluss in beide Richtungen erlaubt.³⁴ Im Rahmen dieses Abschnitts werden ausschließlich Feed-Forward-Netze betrachtet.

3.2.1. Funktionsweise

Feed-Forward-Netze zeichnen sich durch eine Anordnung der Neuronen in mehreren Schichten aus, wobei jedes Neuron einer vorgelagerten mit jedem Neuron einer nachgelagerten Schicht verbunden ist. Die oberste Schicht wird als Ausgabe- und die unterste Schicht als Eingabeschicht bezeichnet. Dazwischen liegen eine oder mehrere versteckte Schichten, die von außen nicht zugänglich sind und in denen die eigentliche Informationsverarbeitung stattfindet.³⁵ Die folgenden Ausführungen beschränken sich auf Netze mit einer versteckten Schicht („One Hidden Layer“), deren prinzipieller Aufbau (Topologie) in Abbildung 1 veranschaulicht wird:

³³ Vgl. Lippe, Wolfram, Einführung in Neuronale Netze, 2000.

³⁴ Vgl. Strecker, Stefan, Künstliche Neuronale Netze, 1997, S. 16.

³⁵ Vgl. Strecker, Stefan, Künstliche Neuronale Netze, 1997, S. 17f.

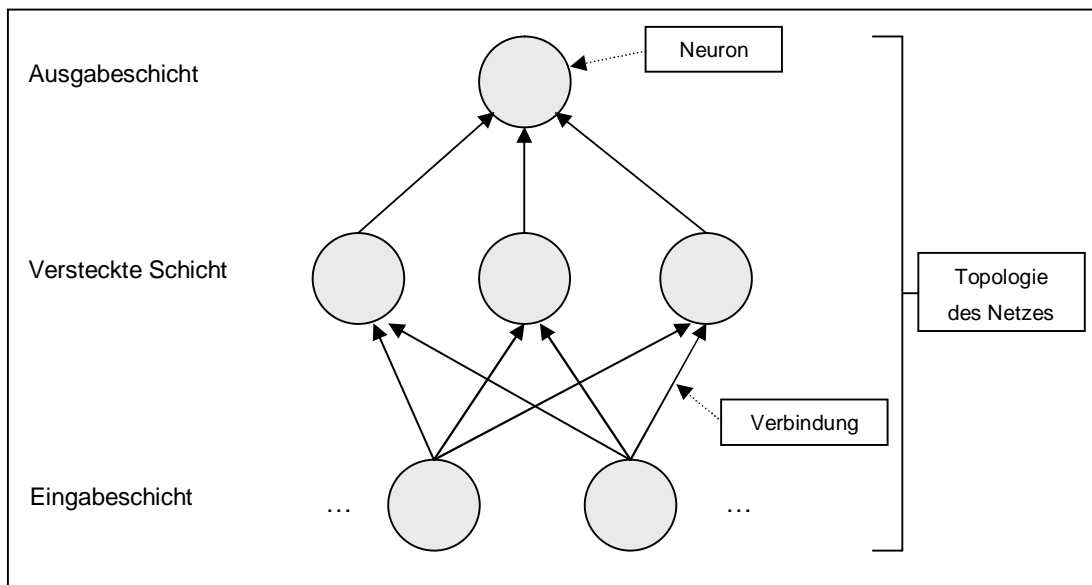


Abbildung 1: Beispielstruktur eines Neuronalen Netzes³⁶

Grundsätzlich besitzt jedes der in Abbildung 1 dargestellten Neuronen p Eingangskanäle, über die es Informationen von anderen Neuronen oder der Netzumgebung empfängt, sowie m Ausgangskanäle, über die es seine berechneten Output o_i an die Neuronen der nachgelagerten Schicht bzw. an die Netzumgebung abgibt.³⁷ Die Berechnung des Outputs erfolgt nach dem in Abbildung 2 gezeigten Prozess:

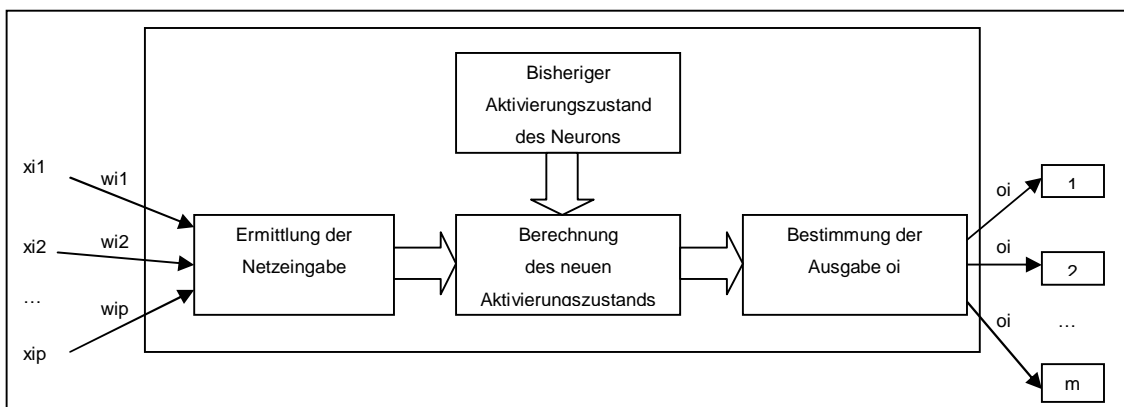


Abbildung 2: Informationsverarbeitungsprozess innerhalb eines Neurons³⁸

³⁶ In Anlehnung an: Strecker, Stefan, Künstliche Neuronale Netze, 1997, S. 12.

³⁷ Vgl. Strecker, Stefan, Künstliche Neuronale Netze, 1997, S. 14.

³⁸ In Anlehnung an: Strecker, Stefan, Künstliche Neuronale Netze, 1997, S. 14.

1. **Ermittlung der Netzeingabe:** Im ersten Schritt ermittelt das Neuron i aus den empfangenen Inputdaten x_{i1}, \dots, x_{ip} und den zugehörigen Verbindungsgewichten $w_{i1}, w_{i2}, \dots, w_{ip}$ seine spezifische Netzeingabe:³⁹

$$net_i = \sum_{j=1}^p w_{ij} x_{ij} \quad (3.7)$$

Die Gewichte $w_{i1}, w_{i2}, \dots, w_{ip}$ entsprechen hierbei reellen Zahlen, die die Stärke einer Verbindung zwischen dem Neuron i und einem Neuron j (mit $j = 1, \dots, p$) der vorgelagerten Schicht beschreiben. Sie werden im Rahmen des Lernprozesses modifiziert und repräsentieren in ihrer Gesamtheit das erlernte Wissen des Netzes.

2. **Berechnung des neuen Aktivierungszustands:** Aus der Netzeingabe net_i und seiner bisherigen Aktivität $a_i(t)$ berechnet das Neuron anschließend über eine Aktivierungsfunktion f seine neue Aktivität $a_i(t+1) = f(net_i, a_i(t))$.⁴⁰ In der Praxis weist f häufig eine lineare, sprunghafte (Schwellenwertfunktion) oder stetige (Sigmoidfunktion) Gestalt auf, wobei aus Vereinfachungsgründen auf eine explizite Berücksichtigung des bisherigen Aktivierungszustands des Neurons verzichtet wird:⁴¹

Lineare Funktion: $f(net_i) = net_i \quad (3.8)$

Schwellenwertfunktion: $f(net_i) = \begin{cases} 1: & \text{wenn } net_i \geq S \\ 0: & \text{wenn } net_i < S \end{cases} \quad (3.9)$

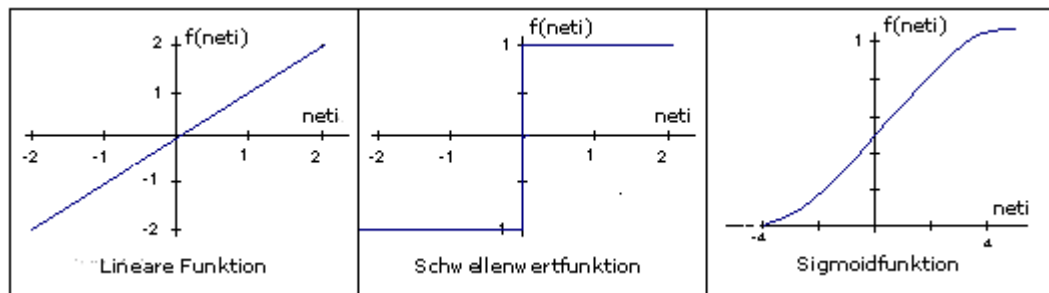
Sigmoidfunktion: $f(net_i) = \frac{1}{1 + e^{-net_i}} \quad (3.10)$

Die Verläufe der oben genannten Funktionen sind in Abbildung 3 dargestellt:

³⁹ Vgl. Lippe, Wolfram-M., Prinzipien Künstlicher Neuronaler Netze, 2000.

⁴⁰ Vgl. Strecker, Stefan, Künstliche Neuronale Netze, 1997, S. 13.

⁴¹ Vgl. Lippe, Wolfram-M., Prinzipien Künstlicher Neuronaler Netze, 2000.

Abbildung 3: Kurvenverläufe ausgewählter Aktivierungsfunktionen⁴²

Bei der Schwellenwertfunktion nimmt das Neuron den Zustand Eins an, wenn seine Netzeingabe eine definierte Schranke S übersteigt, andernfalls den Zustand Null. Der Wertebereich der Funktion ist demzufolge diskret und auf die beiden Werte $\{0,1\}$ beschränkt. Im Unterschied zur Schwellenwertfunktion besitzt die Sigmoidfunktion einen stetigen Wertebereich im Intervall von $[0,1]$. Sie lässt sich zudem differenzieren, was für die spätere Lernphase eine wesentliche Voraussetzung ist.⁴³

3. **Bestimmung der Ausgabe:** Den Abschluss des Informationsverarbeitungsprozesses bildet die Bestimmung des neuronalen Outputs o_i , der an die Neuronen in der nachgelagerten Schicht bzw. an die Netzumgebung weitergeleitet wird. Dazu wendet das Neuron auf seinen zuvor ermittelten Aktivierungszustand $a_i(t+1)$ die Ausgabefunktion f_{Out} an:

$$o_i = f_{Out}(a_i(t+1)) \quad (3.11)$$

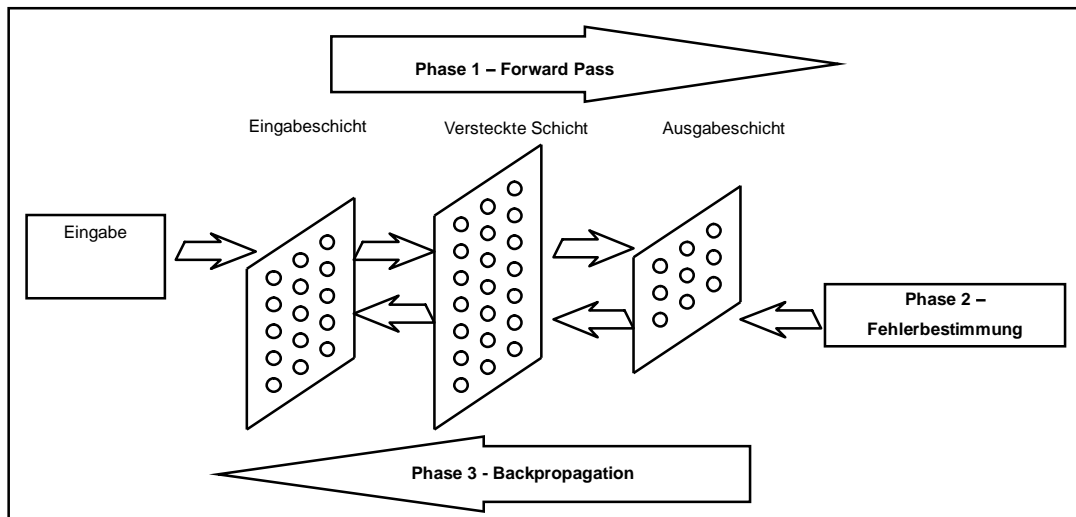
In der Praxis besitzt f_{Out} allerdings häufig keine weitere Funktionalität, so dass direkt $o_i = a_i(t+1)$ gilt.⁴⁴

Das Lernen in einem Neuronalen Netz vollzieht sich durch die Modifikation der Verbindungsgewichte zwischen den Neuronen. Diese werden in der Initialisierungsphase des Netzes zunächst per Zufall bestimmt und anschließend auf Basis einer Trainingsstichprobe so angepasst, dass das Netz die gewünschten Ausgaben erzeugt. Dabei lässt sich der Ablauf des Lernens in die folgenden drei Phasen unterteilen:

⁴² In Anlehnung an: Lippe, Wolfram-M., Prinzipien Künstlicher Neuronaler Netze, 2000.

⁴³ Vgl. Nothnagel, Michael, Diskriminanzanalyse, 1999, S. 77.

⁴⁴ Vgl. Strecker, Stefan, Künstliche Neuronale Netze, 1997, S. 13.

Abbildung 4: Phasen des Lernprozesses in einem NN⁴⁵

1. **Forward Pass:** In der ersten Phase des Lernprozesses wird ein beliebiges Objekt aus der Trainingsstichprobe ausgewählt und die einzelnen Komponenten seines Merkmalsvektors x_1, \dots, x_p an den entsprechenden Neuronen der Eingabeschicht angelegt. Diese leiten die Informationen direkt an die Neuronen in der versteckten Schicht weiter, wo die Verarbeitung der Daten nach dem in Abbildung 2 dargestellten Schema erfolgt. Der jeweils berechnete Output wird den Neuronen in der Ausgabeschicht zur Verfügung gestellt, dort erneut verarbeitet und in Form der prognostizierten Gruppenzugehörigkeit für das angelegte Objekt an die Netzumgebung gesendet.
2. **Fehlerbestimmung:** Die Ausgabe des Netzes wird anschließend mit der erwünschten Ausgabe, in diesem Fall der tatsächlichen Gruppenzugehörigkeit des Trainingsobjekts, verglichen und der Netzwerkfehler bestimmt. Übersteigt dieser einen festgelegten Schwellwert, so erfolgt im Rahmen der Backpropagation-Phase eine Anpassung der Verbindungsgewichte, andernfalls wird die Lernphase für dieses Objekt beendet.⁴⁶
3. **Backpropagation-Phase:** Die Anpassung der Verbindungsgewichte geschieht innerhalb eines iterativen Prozesses. Zunächst werden die Gewichte zwischen der Ausgabeschicht und der ihr vorgelagerten versteckten Schicht modifiziert, anschließend erfolgt die Anpassung der Gewichte zwischen der versteckten

⁴⁵ In Anlehnung an: Werner, H., Backpropagation, 1997, S. 1.

⁴⁶ Vgl. Lippe, Wolfram-M., Backpropagation Learning, 2000.

Schicht und der Eingabeschicht.⁴⁷ Die Backpropagation-Phase führt schließlich zur Ermittlung von Gewichtskombinationen, die eine bestmögliche Klassifikation der Trainingsobjekte erlauben.

Nach Abschluss der Lernphase wird die Generalisierungsfähigkeit des trainierten Netzes anhand einer weiteren Stichprobe überprüft. Diese Teststichprobe enthält neue Objekte, die mit denen aus der Trainingsstichprobe nicht identisch sind. Auch für diese Fälle wird der Netzwerkfehler ermittelt und eine erneute Trainingsphase initiiert, falls der Schwellwert aus Phase zwei überschritten wird.⁴⁸ Andernfalls wird die Testphase abgeschlossen und dem Netz eine genügend hohe Vorhersagegenauigkeit für zukünftige, neu zu klassifizierende Objekte unterstellt.

3.2.2. Eigenschaften

Neuronale Netze zeichnen sich insbesondere durch ihre hohe Lernfähigkeit, Parallelität bei der Informationsverarbeitung und Fehlertoleranz aus.⁴⁹ Ihre Struktur ist durch die Hinzunahme weiterer Schichten erweiterbar, so dass theoretisch beliebig komplexe Funktionen approximiert werden können.⁵⁰ Allerdings besteht der große Nachteil eines Neuronalen Netzes in der mangelnden Transparenz seiner Ergebnisse, da sich die Verbindungsgewichte kaum interpretieren lassen. Zudem weist *Nothnagel* darauf hin, dass das Setup eines Netzes eine Vielzahl von Entscheidungen erfordert (z.B. die Anzahl der Schichten, die Anzahl der Neuronen pro Schicht, die Wahl der Aktivierungsfunktion, etc.), die „eine automatisierte Bestimmung der Diskriminanzregel“⁵¹ nur begrenzt zulassen.⁵²

3.3. k-Nearest-Neighbor

3.3.1. Funktionsweise⁵³

Dem k-Nearest-Neighbor-Verfahren liegt die Überlegung zugrunde, dass benachbarte Objekte häufig ähnliche Merkmalsausprägungen besitzen und

⁴⁷ Vgl. Lippe, Wolfram-M., Backpropagation Learning, 2000.

⁴⁸ Vgl. Lippe, Wolfram-M., Backpropagation Learning, 2000.

⁴⁹ Vgl. Lippe, Wolfram-M., Einführung in Neuronale Netze, 2000.

⁵⁰ Vgl. Nothnagel, Michael, Diskriminanzanalyse, 1999, S. 80.

⁵¹ Nothnagel, Michael, Diskriminanzanalyse, 1999, S. 80.

⁵² Vgl. Nothnagel, Michael, Diskriminanzanalyse, 1999, S. 80.

⁵³ In diesem Abschnitt bezeichnet die Variable k die Anzahl der Nachbarn.

derselben Gruppe angehören. Daher ordnet das k-Nearest-Neighbor-Verfahren ein neues Objekt derjenigen Gruppe G_i zu, aus der die Mehrzahl seiner k Nachbarn stammt.⁵⁴ Diese werden auf Basis einer Trainingsstichprobe ermittelt, wobei sich das grundsätzliche Vorgehen in drei Phasen unterteilt:

1. **Abstandsberechnung:** Zunächst wird der Abstand zwischen dem Merkmalsvektor x des zu klassifizierenden Objekts und dem Merkmalsvektor y_i (mit $i = 1, \dots, n$) jedes Objekts aus der Trainingsstichprobe bestimmt. Als Abstandsmaß kann hierfür z.B. der Euklidische Abstand verwendet werden, wobei auf die Skalierung der Merkmale zu achten ist.⁵⁵

$$D(x, y_i) = \sqrt{\sum_{j=1}^p (x_j - y_{ij})^2} \quad (4.1)$$

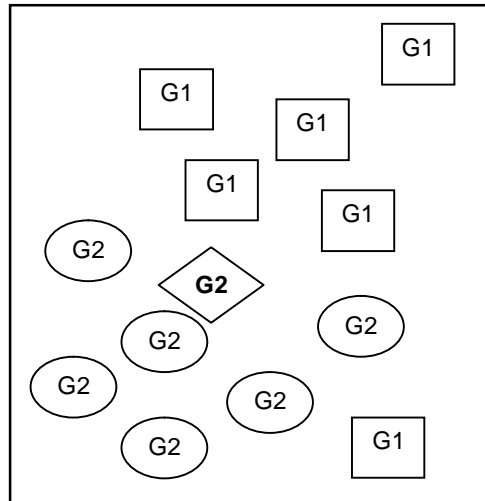
2. **Auswahl der k Nachbarn:** Aus der Trainingsstichprobe werden anschließend diejenigen k Objekte ausgewählt, die den kleinsten Abstand $D(x, y_j)$ zum Merkmalsvektor x aufweisen.⁵⁶
3. **Klassifikation:** In der abschließenden Klassifikationsphase wird die Auftrittshäufigkeit der einzelnen Gruppen anhand der k Objekte bestimmt und das neu zu klassifizierende Objekt der Gruppe mit der größten Häufigkeit zugeordnet.

In Abbildung 5 wird die Klassifikation eines Objekts nach dem k-Nearest-Neighbor-Verfahren für den Fall $k = 1$ noch einmal beispielhaft dargestellt. Dabei werden die Objekte aus der Gruppe G_1 durch Rechtecke und die Objekte aus der Gruppe G_2 durch Kreise repräsentiert. Das neu zu klassifizierende Objekt (Raute) weist zu einem Objekt aus der Gruppe G_2 den geringsten Abstand auf und wird daher dieser Gruppe zugeordnet:

⁵⁴ Vgl. Pretzer, Michael, Clustering und Klassifikation, 2003, S. 30.

⁵⁵ Vgl. Nothnagel, Michael, Diskriminanzanalyse, 1999, S. 60.

⁵⁶ Vgl. Pretzer, Michael, Clustering und Klassifikation, 2003, S. 30.

Abbildung 5: k-Nearest-Neighbor-Verfahren (k=1)⁵⁷

3.3.2. Eigenschaften

Im Gegensatz zu den bisher betrachteten Klassifikatoren entfällt beim k-Nearest-Neighbor-Verfahren prinzipiell die Lernphase. Sämtliche Informationen aus der Trainingsstichprobe werden lediglich zwischengespeichert und erst ausgewertet, wenn ein neues Objekt klassifiziert werden soll.⁵⁸ Dem stehen die wesentlichen Nachteile gegenüber, dass die vollständigen Trainingsdaten für jede Klassifikation verfügbar gehalten werden müssen und die Berechnungen zur Klassifikationszeit deutlich mehr Aufwand erfordern.⁵⁹

3.4. CART

Der Begriff CART steht für *Classification and Regression Tree* und bezeichnet ein von Breimann et al.⁶⁰ 1984 entwickeltes Klassifikationsverfahren. Dabei wird auf Basis der Trainingsstichprobe zunächst ein Entscheidungsbaum konstruiert und dieser anschließend zur Klassifikation neuer Fälle eingesetzt.

⁵⁷ In Anlehnung an: Simard, Philippe / Mitran, Marcel, The Nearest Neighbor Rule, 1999.

⁵⁸ Vgl. Pretzer, Michael, Clustering und Klassifikation, 2003, S. 31f.

⁵⁹ Vgl. Weihs, Claus, Diskriminanzanalyse, 2000, S. 584.

⁶⁰ Vgl. Breiman, L. / Friedman, J.H. / Olshen, R.A. / Stone, C.J., Classification and Regression Trees, 1984.

3.4.1. Funktionsweise

In Abbildung 6 wird der prinzipielle Aufbau eines CART am Beispiel einer stark vereinfachten Kreditwürdigkeitsprüfung verdeutlicht. Die Zuordnung eines Kunden zur Gruppe G1 (kreditwürdig) bzw. G2 (nicht kreditwürdig) erfolgt durch die kundenindividuelle Beantwortung bestimmter Fragen (z.B. „Alter < 18“) und der damit verbundenen Verzweigung im Entscheidungsbaum. Das Ergebnis der Klassifikation steht fest, wenn ein Blattknoten erreicht ist, d.h. ein Knoten, ohne weitergehende Verzweigungen. Dessen Beschriftung gibt die vom Baum prognostizierte Gruppenzugehörigkeit des Neukunden an.

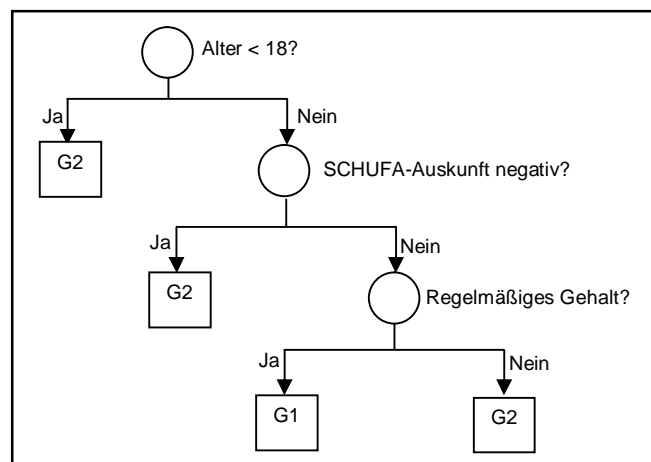


Abbildung 6: Beispiel eines CART
für eine Kreditwürdigkeitsprüfung

Der Algorithmus zur Konstruktion eines CART beruht auf dem Prinzip der binären rekursiven Partitionierung.⁶¹ Den Ausgangspunkt bildet hierbei die Trainingsstichprobe, die zunächst in zwei disjunkte Teilmengen zerlegt wird. Dieser Prozess setzt sich anschließend für jede der entstandenen Teilmengen solange rekursiv fort, bis jede Teilmenge nur noch Objekte derselben Gruppe beinhalten oder ihr Umfang einen vorgegebenen Grenzwert unterschreitet.⁶² Im Kontext des CART entspricht jede Teilmenge einem Knoten und die ursprüngliche Gesamtstichprobe der Wurzel des Entscheidungsbaumes.

⁶¹ Vgl. Schillinger, Marco Patrick, Classification and Regression Trees, 2002, S. 47.

⁶² Vgl. Schillinger, Marco Patrick, Classification and Regression Trees, 2002, S. 48.

Die Aufteilung eines Knotens t erfolgt durch einen so genannten Split s , der alle Objekte von t anhand ihrer Merkmalsausprägung x_j (mit $j = 1, \dots, p$) entweder einem linken Kindknoten t_l oder einem rechten Kindknoten t_r zuordnet. In Abhängigkeit von der Skalierung des zugrunde liegenden Merkmals X_j besitzt s die nachstehende Form.⁶³

$$X_j \text{ qualitativ: } s(x_j) = \begin{cases} t_l & \text{wenn } x_j \in C; C \subseteq \{a_1, \dots, a_n\} \\ t_r & \text{wenn } x_j \notin C; C \subseteq \{a_1, \dots, a_n\} \end{cases} \quad (4.2)$$

$$X_j \text{ quantitativ: } s(x_j) = \begin{cases} t_l & \text{wenn } x_j \leq c; c \in (-\infty, \infty) \\ t_r & \text{wenn } x_j > c; c \in (-\infty, \infty) \end{cases} \quad (4.3)$$

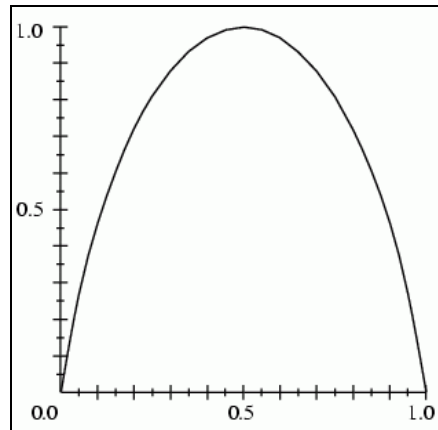
Das Ziel einer Zerlegung von t durch einen Split s besteht darin, möglichst homogene Kindknoten zu erzeugen, die jeweils nur Objekte mit derselben Gruppenzugehörigkeit beinhalten. Als Maß für die Homogenität eines Knotens kann der so genannte Gini-Index verwendet werden:⁶⁴

$$gini(t) = 1 - \sum_{i=1}^k p_i^2 \quad (4.4)$$

Dabei bezeichnet k die Anzahl der Gruppen und p_i die Wahrscheinlichkeit, mit der ein Objekt des Knotens t der Gruppe G_i angehört. In Abbildung 7 ist der Verlauf des Gini-Index für $k = 2$ dargestellt:

⁶³ Vgl. Nothnagel, Michael, Diskriminanzanalyse, 1999, S. 66f.

⁶⁴ Vgl. Bonne, Thorsten / Arminger, Gerhard, Diskriminanzanalyse, 2001, S. 216.

Abbildung 7: Verlauf des Gini-Index (k=2)⁶⁵

Der Gini-Index erreicht für den Knoten t sein Maximum, wenn die beiden Gruppen G_1 und G_2 in diesem Knoten gleichverteilt auftreten ($p_1 = 0,5$ und $p_2 = 0,5$). Ist die Wahrscheinlichkeit für eine der beiden Gruppen hingegen Null ($p_1 = 0$ oder $p_2 = 0$), so nimmt er seinen minimalen Wert an.⁶⁶

Die Aufgabe des Teilungsalgorithmus besteht darin, aus der Menge aller für einen Knoten in Frage kommender Splits S diejenigen auszuwählen, der die größtmögliche Veränderung der Homogenität im Entscheidungsbaum bewirkt. Diese Veränderung lässt sich mit Hilfe des Gini-Index für einen Split s wie folgt berechnen:⁶⁷

$$\Delta gini(s,t) = gini(t) - P(t_l)gini(t_l) - P(t_r)gini(t_r) \quad (4.5)$$

Die Terme $P(t_l)$ und $P(t_r)$ in (4.5) stehen hierbei für die bedingten Wahrscheinlichkeiten, mit denen ein Objekt aus dem Knoten t dem linken Kindknoten t_l bzw. dem rechten Kindknoten t_r zugeordnet wird. Diese Wahrscheinlichkeiten können ermittelt werden, indem die Anzahl der in den jeweiligen Kindknoten einsortierten Objekte durch die Gesamtzahl der Objekte im Ausgangsknoten dividiert wird.⁶⁸

⁶⁵ In Anlehnung an : Bonne, Thorsten / Arminger, Gerhard, Diskriminanzanalyse, 2001, S. 216.

⁶⁶ Vgl. Bonne, Thorsten / Arminger, Gerhard, Diskriminanzanalyse, 2001, S. 217.

⁶⁷ Vgl. Bonne, Thorsten / Arminger, Gerhard, Diskriminanzanalyse, 2001, S. 217.

⁶⁸ Vgl. Nothnagel, Michael, Diskriminanzanalyse, 1999, S. 69.

$$P(t_l) = \frac{|t_l|}{|t|} \text{ bzw. } P(t_r) = \frac{|t_r|}{|t|} \quad (4.6)$$

Der auf Basis des beschriebenen Algorithmus erzeugte Entscheidungsbaum wird als Maximalbaum T_{MAX} bezeichnet. Jeder Endknoten t repräsentiert darin eine der beiden Gruppen G_1 bzw. G_2 , wobei die Gruppenzuordnung von t über die Bayes-Regel erfolgt (gleiche Klassifikationskosten vorausgesetzt):⁶⁹

$$\lambda_{Bayes}(t) = \begin{cases} G_1 : \text{ wenn } P(G_1 | t) \geq P(G_2 | t) \\ G_2 : \text{ wenn } P(G_1 | t) < P(G_2 | t) \end{cases} \quad (4.7)$$

Im Hinblick auf die Fehlerrate bzw. die Fehlklassifikationskosten bildet T_{MAX} eine optimale Zerlegung der zugrunde liegenden Trainingsstichprobe. Allerdings weist der Baum im Allgemeinen eine zu starke Anpassung an die zufällige Struktur dieser Stichprobe auf („Overfitting“), so dass er nur eine sehr eingeschränkte Generalisierungsfähigkeit für neue Objekte besitzt.⁷⁰ Um diese Abhängigkeit von der Trainingsstichprobe zu reduzieren, wird der Maximalbaum daher auf seine optimale Größe gestutzt („Pruning“). Dazu werden aus dem Baum sukzessive schwach besetzte Knoten entfernt und Teilbäume mit abnehmender Komplexität erzeugt. Aus dieser Folge $T_{MAX} \succ T_1 \succ T_2 \succ \dots \succ \{t_0\}$ von Teilbäumen wird anschließend derjenige ausgewählt, der für eine Teststichprobe die kleinste Fehlerrate besitzt bzw. die Fehlklassifikationskosten minimiert.⁷¹ Eine detailliertere Beschreibung des Pruning-Algorithmus findet sich in den Arbeiten von *Bonne*⁷² oder *Schillinger*⁷³.

3.4.2. Eigenschaften

Der Vorteil des CART-Verfahrens liegt in seiner einfachen Anwendbarkeit: Nachdem der Entscheidungsbaum konstruiert ist, lassen sich neue Objekte schnell und effizient klassifizieren.⁷⁴ Die bei der Zuordnung getroffenen Entscheidungen sind transparent und können somit leicht interpretiert werden. Die Nachteile des

⁶⁹ Vgl. Nothnagel, Michael, Diskriminanzanalyse, 1999, S. 68.

⁷⁰ Vgl. Bonne, Thorsten / Arminger, Gerhard, Diskriminanzanalyse, 2001, S. 217

⁷¹ Vgl. Bonne, Thorsten / Arminger, Gerhard, Diskriminanzanalyse, 2001, S. 217

⁷² Vgl. Bonne, Thorsten / Arminger, Gerhard, Diskriminanzanalyse, 2001.

⁷³ Vgl. Schillinger, Marco Patrick, Classification and Regression Trees, 2002.

⁷⁴ Vgl. Nothnagel, Michael, Diskriminanzanalyse, 1999, S. 71.

Verfahrens liegen in dem relativ hohen Aufwand zur Konstruktion des CART sowie in der starken Abhängigkeit der Baumstruktur von der verwendeten Trainingsstichprobe.⁷⁵

⁷⁵ Vgl. Nothnagel, Michael, Diskriminanzanalyse, 1999, S. 71.

4. Zusammenfassung

Im Rahmen dieser Arbeit wurden die Grundlagen der Diskriminanzanalyse erläutert und die - aus der Sicht des Autors - wesentlichen Klassifikationsverfahren vorgestellt.

Den Ausgangspunkt bildete hierbei eine zu klassifizierende Grundgesamtheit Ω und eine Menge von vordefinierten Gruppen G_1, \dots, G_k . Gesucht wurde ein geeigneter Klassifikator, der jedes Objekt aus Ω anhand des objektspezifischen Merkmalsvektors x genau einer der k Gruppen so zuordnet, dass diese Zuordnung mit der tatsächlichen Gruppenzugehörigkeit des Objekts übereinstimmt. Dabei ergab sich das Problem, dass für die Konstruktion des Klassifikators im Allgemeinen nur eine begrenzte Teilmenge von Objekten aus Ω zur Verfügung steht und der Klassifikator somit auf Basis dieser Teilmenge für die Grundgesamtheit geschätzt werden muss.

Einige der hierfür geeigneten Verfahren wurden in Kapitel 3 näher erläutert, wobei drei der vorgestellten Verfahren in einer Lernphase zunächst die Teilmenge auswerten und explizites Wissen in Form von Wahrscheinlichkeitsverteilungen (Bayes), gewichteten Verbindungen (Neuronale Netze) oder Entscheidungsbäumen (CART) aufbauen. Bei dem vierten Verfahren (k-Nearest-Neighbor) erfolgt die Auswertung der Teilmenge hingegen erst zur Klassifikationszeit, so dass die Lernphase hier komplett entfällt.

Die Auswahl einer Diskriminanzregel ist von zahlreichen Kriterien abhängig, von denen im Rahmen der Ausarbeitung die Fehlerrate und die Art der zugrunde liegenden Merkmalsarten näher erläutert wurden. Weitere Kriterien sind u.a. die Geschwindigkeit der Klassifikation sowie die Interpretierbarkeit der Ergebnisse. Welche dieser Kriterien allerdings letztlich den Ausschlag für die Verfahrensauswahl geben, ist von Fall zu Fall verschieden.

Literaturverzeichnis

Bücher

Bonne, Thorsten / Arminger, Gerhard, 2001: [Diskriminanzanalyse], in: Hippner, H. et al. (Hrsg.): Handbuch Data Mining im Marketing, Wiesbaden: Vieweg-Verlag, S. 193-239

Breiman, L. / Friedman, J.H. / Olshen, R.A. / Stone, C.J., 1984: [Classification and Regression Trees], Belmont: Wadsworth, 1984

Fisher, R. A., 1936: [The use of multiple measurements] in taxonomic problem, Annals of Eugenics, 7, S. 179-188

Trampisch, H.J, 1985: [Diskriminanzanalyse], in: Deichsel, G., Trampisch, H.J.: Clusteranalyse und Diskriminanzanalyse, Stuttgart: Gustav Fischer Verlag, S. 57-132

Weihls, Claus, 2000: [Diskriminanzanalyse], in: Voß, Werner (Hrsg.): Taschenbuch der Statistik, Leipzig: Fachbuchverlag Leipzig, 2000, S. 583-608

Internet

Bao, Ho Tu, o.J.: [Knowledge Engineering], Internet <http://www.netnam.vn/unescocourse/knowlegde/71.htm>, Abruf 2005-03-20, Vietnam, o.J.

Diaz-Bone, Rainer, 2005: [Multiple Lineare Regression], Internet http://userpage.fu-berlin.de/~diazbone/Multiple%20lineare%20Regression_SS05.pdf, Abruf 2005-03-05, Berlin, 2005

Hupfeld, Jörg, 1998: Eine kurze [Einführung in die Diskriminanzanalyse], Internet <http://www.psy.unibe.ch/soz/team/pdf/hupfeld/Hupfeld1998.pdf>, Abruf 2005-03-10, Bern, 1998

Lippe, Wolfram-M., 2000: [Einführung in Neuronale Netze], Internet <http://wwwmath.uni-muenster.de/SoftComputing/lehre/material/wwwnscript/einl.html>, Abruf 2005-04-20, 2000

Lippe, Wolfram-M., 2000: [Backpropagation Learning], Internet <http://wwwmath.uni-muenster.de/SoftComputing/lehre/material/wwwnscript/backprop1.html>, Abruf 2005-04-20, 2000

Lippe, Wolfram-M., 2000: [Prinzipien Künstlicher Neuronaler Netze], Internet <http://wwwmath.uni-muenster.de/SoftComputing/lehre/material/wwwnscript/prin.html>, Abruf 2005-04-20, 2000

Löwe, Matthias, 2005: [Einführung in die elementare Stochastik], Internet <http://wwwmath.uni-muenster.de/math/inst/statistik/loewe/GHR/ghrkap1.pdf>, Abruf 2005-03-15, Münster, 2005

- Nothnagel, Michael, 1999: Klassifikationsverfahren der [Diskriminanzanalyse] – Eine vergleichende und integrierende Übersicht, Internet <http://capella.bioinf.mdc-berlin.de/da/mndalink.pdf>, Abruf 2005-03-20, Berlin, 1999
- Pretzer, Michael, 2003: [Clustering und Klassifikation], Internet <http://www.diko-project.de/dokumente/ausarbeitungen/pretzer.pdf>, Abruf 2005-03-25, Oldenburg, 2003
- 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/>, Abruf 2005-04-20, 2002
- Simard, Philippe / Mitran, Marcel, 1999: [The Nearest Neighbor Rule]: A Short Tutorial, Internet http://cgm.cs.mcgill.ca/~sooss/cs644/projects/simard/nn_theory.html, Abruf 2005-04-25, Montreal, 1999
- Stahel, Werner, o.J.: [Clusteranalyse], Distanzmethoden, Skalierung, Internet <http://stat.ethz.ch/~stahel/courses/multivariate/script/cla.pdf>, Abruf 2005-03-05, Zürich, o.J.
- Strecker, Stefan, 1997: [Künstliche Neuronale Netze] – Aufbau und Funktionsweise, Internet <http://geb.uni-giessen.de/geb/volltexte/2004/1697/>, Abruf 2005-05-01, Mainz, 1997
- Werner, H., 1997: [Backpropagation], Internet <http://www.neuro.informatik.uni-kassel.de/%20~werner/Adaptiv/Adapt03.ppt>, Abruf 2005-05-03, Kassel, 1997
- Wiehr, Stephan, 2002: [Automatische Klassifikation] semistrukturierter Daten, Internet <http://fsinfo.cs.uni-sb.de/~lynx/uni/diplom/node17.html>, Abruf 2005-03-25, Saarbrücken, 2002