Musikinstrumentenerkennung mit Hilfe der Hough-Transformation

Diplomarbeit im Fach Statistik

an der Universität Dortmund

eingereicht bei

Prof. Dr. Claus Weihs

vorgelegt von

Christian Röver

Herderstraße 69

44147 Dortmund

Dortmund im Juli 2003

Inhaltsverzeichnis

1 Einleitung **3**

2 Zugrundeliegendes Datenmaterial **5**

2.1 Die Audio-Rohdaten 5

2.1.1 Schall und Klang 5

2.1.2 Klangdigitalisierung 6

2.1.3 Der Datensatz 8

2.2 Die Hough-Transformation 8

2.2.1 Generelles Prinzip 8

2.2.2 Anwendung auf Audiodaten 12

2.2.3 Parametrisierung und Umsetzung 13

2.3 Resultierendes Datenformat 16

3 Klassifikation **19**

3.1 Das Klassifikationsproblem 19

3.2 Datenaufbereitung 21

3.2.1 Besetzungszahlen 21

3.2.2 Hough-Charakteristika 22

3.2.3 Clusteranalyse 24

3.3 Kurzer Datenüberblick 27

3.4 Diskriminanzanalyse 29

3.4.1 Lineare Diskriminanzanalyse (LDA) 29

3.4.2 Quadratische Diskriminanzanalyse (QDA) 32

3.4.3 Naive Bayes 33

3.4.4 Regularisierte Diskriminanzanalyse (RDA) 34
3.5 Support Vector Machines 37
3.6 Klassifikationsbäume 38
3.7 k-Nearest-Neighbour 40
3.8 Poisson-Modell 41
3.9 Variablenselektion 44
3.10 Benutzte Software 46

4 Ergebnisse **47**
4.1 Die Fehlerraten 47
4.2 Erster Ansatz: Besetzungszahlen 48
4.3 Zweiter Ansatz: Hough-Charakteristika 50
4.3.1 Variablenselektion 50
4.3.2 Fehlerraten 54
4.4 Zur Center-Frequency 56

5 Zusammenfassung **58**

A Tabellen und Abbildungen **60**

B Mathematischer Anhang **70**

C Literaturverzeichnis **76**

1 Einleitung

Diese Diplomarbeit entstand im Rahmen der Zusammenarbeit des Fachbereichs Statistik (speziell des Lehrstuhls für computergestützte Statistik) mit dem Fraunhofer Institut für Integrierte Schaltungen in Ilmenau (hier genauer die Arbeitsgruppe Elektronische Medientechnologie AEMT). Eine Zusammenarbeit findet seit Ende 2002 statt und bezieht sich auf das gemeinsame Forschungsgebiet, die mathematische Erfassung von Musik-Audiodaten.
Auf Dortmunder Seite läuft Forschung in dieser Richtung seit etwa 1999 und beschäftigt sich beispielsweise mit der statistischen Modellierung der Charakteristika von Gesangsstimmen; in diesem Kontext entstanden auch schon mehrere weitere Diplomarbeiten. Am Fraunhofer Institut wird momentan u.a. an der Extraktion von Metadaten aus Musikdateien gearbeitet, also beispielsweise Rhythmus- oder Melodieerkennung und der Abgleich mit entsprechenden Datenbanken.
Grundlage dieser Arbeit sind Daten, die mit Hilfe eines neu entwickelten Computerchips (ein ASIC = *application specific integrated circuit*, „anwendungsspezifische integrierte Schaltung“) aus digitalen Tonaufnahmen gewonnen werden können. Der Chip setzt ein Verfahren um, das klassischerweise aus der Bilderkennug stammt, das aber prinzipiell ebenso auf Audiodaten angewandt werden kann. Bei dem Verfahren handelt es sich um die *Hough-Transformation*, die im Jahre 1959 ursprünglich zum Aufspüren von Spuren von Elementarteilchen entwickelt wurde, und die in ihrer verallgemeinerten Form zur Erkennung von Kanten, Umrissen etc. in digitalisierten (insbesondere auch in verrauschten) Bildern angewandt wird.
In bezug auf die Anwendung auf Audiodaten soll nun die Eignung des Verfahrens zur Erkennung von Musikinstrumenten anhand ihres (digitalisierten) Klanges untersucht

werden. Nachdem ein digitalisierter Klang (beispielsweise ein Flötenton) vom Chip verarbeitet wurde, soll also aus den hierdurch gelieferten Daten auf das Instrument (Flöte) rückgeschlossen werden. Es handelt sich damit um ein Klassifikationsproblem, d.h. anhand der vom Chip gelieferten Information soll eine Entscheidung für eines aus einer bestimmten Auswahl von Instrumenten getroffen werden.
Die zentralen Fragen sind hier:

- *Auf welche Weise* kann man mit Hilfe der Hough-Transformation verschiedene Instrumente unterscheiden?
- *Wie sicher* ist die Vorhersage; wie groß ist dabei die Fehlerrate?
- Ist das ein *erfolgversprechender Ansatz*?

Im folgenden Kapitel wird zunächst erklärt, wie Klänge digitalisiert werden, was die Hough-Transformation ist und auf welche Weise sie hier angewandt wird und wie letztlich die Daten aussehen, auf deren Basis die Klassifikation stattfinden soll.
In Kapitel 3 wird das Klassifikationsproblem ausgeführt und dargelegt, wie es angegangen werden soll. Dabei werden auch die verwendeten Klassifikationsverfahren und weitere notwendige Schritte erläutert.
Kapitel 4 stellt dann die Ergebnisse der einzelnen Ansätze in einiger Ausführlichkeit dar, und in Kapitel 5 wird auf das letztlich erfolgversprechendste Verfahren noch einmal eingegangen.

2 Zugrundeliegendes Datenmaterial

2.1 Die Audio-Rohdaten

2.1.1 Schall und Klang

Schall ist eine mechanische Schwingung, der Luft im allgemeinen, und ein (Instrumenten-) *Klang* ist ebenfalls eine Form von Schall. Der Klang ist dabei ein Sonderfall, nämlich eine *periodische* Schwingung (im Gegensatz zum *Geräusch*, das nichtperiodisch ist). Periodisch bedeutet, daß sich ein Schwingungsmuster in festen Zeitabständen wiederholt; dieser Zeitabstand legt wiederum durch seinen Kehrwert die *Frequenz* des Klanges fest. Beträgt die Periode beispielsweise 0.002 Sekunden, so wiederholt sich die Schwingung 500mal pro Sekunde, der Klang hat damit eine Frequenz von 500 Hertz (500 Hz). Hohe Töne haben große (hohe) Frequenzen und tiefe Töne haben kleine (niedrige) Frequenzen.
Abbildung 2.1 zeigt ein Beispiel für einen Klang, einen Trompetenton. Die Schwingung bildet ein wiederkehrendes Muster, dessen Periode durch die Linie oben links angedeutet ist.
In der Musik wird die Tönhöhe einer Note zunächst durch Buchstaben, die die Lage innerhalb einer Oktave beschreiben, und Zahlen, die die Oktave angeben, bezeichnet. Der kleinstmögliche Abstand zwischen zwei Tönen ist (zumindest in der traditionellen europäischen Musik und Notation) der Halbton. Eine Oktave umfaßt 12 Halbtonschritte und bedeutet eine Verdoppelung der Tonfrequenz. Vom (Kammerton) a4 mit der Frequenz 440 Hz gelangt man zum eine Oktave höher gelegenen a5 (mit 880 Hz) über die folgenden Schritte: a4–a♯4–b4–c5–c♯5–d5–d♯5–e5–f5–f♯5–g5–g♯5–a5. Mit jedem der 12 Halbtonschritte erhöht sich bei temperierter Stimmung dabei die Frequenz um

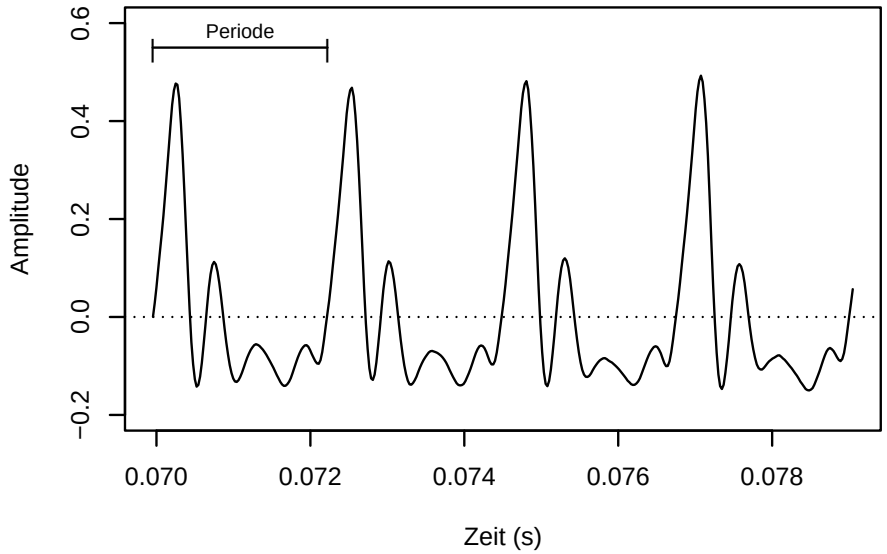

Abbildung 2.1: Die periodische Schwingung eines Klanges.

denselben Faktor $\sqrt[12]{2} \approx 1.06$, so daß man nach 12 Schritten bei Faktor $(\sqrt[12]{2})^{12} = 2$, also der doppelten Frequenz anlangt. Die Frequenz wächst damit exponentiell mit der Tonhöhe.

2.1.2 Klangdigitalisierung

Damit ein Klang mathematisch erfaßt werden kann, muß er zunächst *digitalisiert*, also in Zahlen umgewandelt werden. Die gebräuchliche Form der Digitalisierung, wie sie z.B. bei Audio-CDs und in einigen Klangdateien (z.B. `*.wav`-Dateien) zum Einsatz kommt, ist das *Sampling*. Hier wird die Schwingung durch eine Treppenfunktion angenähert, das heißt, in festen Zeitabständen wird der Schalldruck gemessen und aufgezeichnet, wie in Abbildung 2.2 angedeutet.
Die entscheidenden Parameter, die dabei die Tonqualität bestimmen, sind die *Abtastrate* und die *Auflösung*. Die Abtastrate gibt die Zeitabstände zwischen den aufgezeichneten Amplitudenwerten (den *Samples*) an, also die Länge der Stufen der Treppenfunktion. Sie wird in Hertz gemessen und beträgt bei Aufnahmen in CD-Qualität (und darum handelt es sich bei den hier behandelten Daten) 44100 Hz, d.h. die Zeitabstände

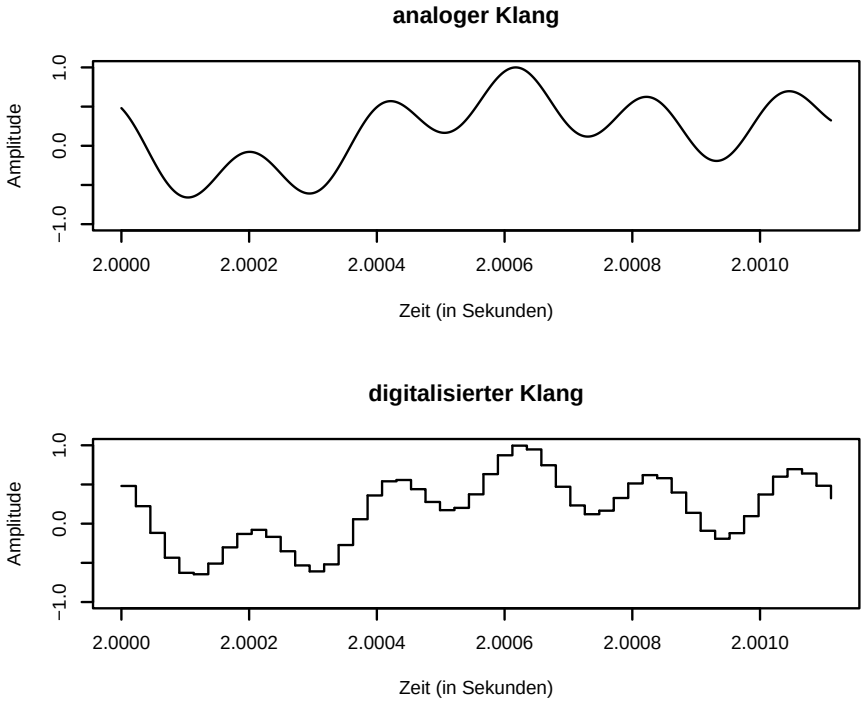

Abbildung 2.2: Digitalisierung eines Klanges.

betragen 1/44100 Sekunde. Die Auflösung gibt die Genauigkeit der aufgezeichneten Amplituden an, diese wird in Bit ausgedrückt und ist bei CD-Qualität wiederum 16 Bit (=2 Byte), d.h. jeder Wert hat eine von $2^{16} = 65536$ möglichen Ausprägungen im Intervall $[-1, 1]$. Dies führt insgesamt zu relativ großen Datenmengen, denn es ergeben sich hier pro Sekunde $44100 \times 2 = 88200$ Bytes, oder andersherum 12 Sekunden pro Megabyte. CDs werden in der Regel in Stereo aufgezeichnet, hier ist der Datenumfang dann wiederum doppelt so groß.

Eine solche Audiodatei listet also prinzipiell einfach die Amplituden in zeitlicher Reihenfolge auf und gibt zusätzlich Abtastrate, Auflösung und Anzahl der Kanäle (Mono/Stereo) an. Die Amplituden sind dann sogenannte „*PCM-Samples*“ (PCM=pulse code modulated). Statistisch ausgedrückt ist es eine Zeitreihe mit äquidistanten Zeitpunkten.

2.1.3 Der Datensatz

Die Daten, die in dieser Arbeit verwandt wurden, stammen aus einer käuflich erhältlichen Sammlung von digitalisierten Instrumentenklängen der McGill University in Montreal, Kanada (McGill, 1987).
Es handelt sich um 62 Sequenzen von Tönen, wobei eine Sequenz bedeutet, daß ein bestimmtes Instrument in einer Reihe von aufeinanderfolgenden Tonhöhen angespielt wurde. Jeder einzelne Ton ist dabei wie im vorigen Abschnitt beschrieben in Form einer Klangdatei gespeichert, die genauen Parameter sind 44.1 kHz, 16 Bit, Mono. Insgesamt ergeben sich Sequenzen mit Umfängen von 6 bis 88, im Mittel sind es etwa 32 Töne, und damit insgesamt 1987 Dateien (zu Details siehe auch Tabelle A.1, Seite 60).
Die Tonhöhe (und damit die Frequenz) ist zu jeder Datei ebenfalls bekannt.

2.2 Die Hough-Transformation

2.2.1 Generelles Prinzip

Die Hough-Transformation ist ein Verfahren, das seinen Ursprung in der Teilchenphysik hat; hier wurde es im Jahre 1959 von P. V. C. Hough entwickelt, um Teilchenspuren (Geraden) in den von entsprechenden Detektoren gemessenen Daten zu entdecken (Hough, 1959). Das Verfahren wurde verallgemeinert auf die Erkennung beliebiger Kurven oder Umrisse und wird heute generell zur Erkennung von Mustern insbesondere auch bei verrauschten Bilddaten verwendet.
Die Hough-Transformation nutzt die Beziehung zwischen Punkten auf einer Kurve und deren Parametern aus. Es werden aus den Bilddaten (Punkte im *Bildraum*) potentielle Parameterkombinationen (Punkte im *Parameterraum*) bestimmt; anschließend wird nach Häufungspunkten im Parameterraum gesucht und daraus die Parameterschätzung abgeleitet.

Die genaue Funktionsweise der Hough-Transformation soll nun am Beispiel der Erkennung einer Geraden (man könnte auch *Schätzung* oder *Anpassung* sagen) erläutert werden.

Es sei eine Menge von Punkten $(x_i, y_i)_{i=1,\ldots,n} \subset \mathbb{R}^2$ gegeben, die potentiell zu einer Geraden gehören. Die gesuchte Gerade hat die (unbekannten) Parameter α und β und alle Punkte (x, y), die auf der Geraden liegen, erfüllen also

$$y = \alpha x + \beta. \tag{2.1}$$

Die Bildpunkte (x_i, y_i) liegen im *Bildraum*, die Parameter (α, β) liegen im *Parameterraum*; beide Räume sind hier zweidimensional.
Für einen Bildpunkt (x_i, y_i) gibt es eine Menge von möglichen Lösungen für α und β, diese liegen wiederum auf einer Geraden im Parameterraum, die durch die Gleichung

$$\beta = -x_i\alpha + y_i \tag{2.2}$$

gegeben ist.
In Abbildung 2.3 sind drei Bildpunkte mit den drei zugehörigen Geraden im Parameterraum dargestellt. Jeder Schnittpunkt von zwei Geraden im Parameterraum be-

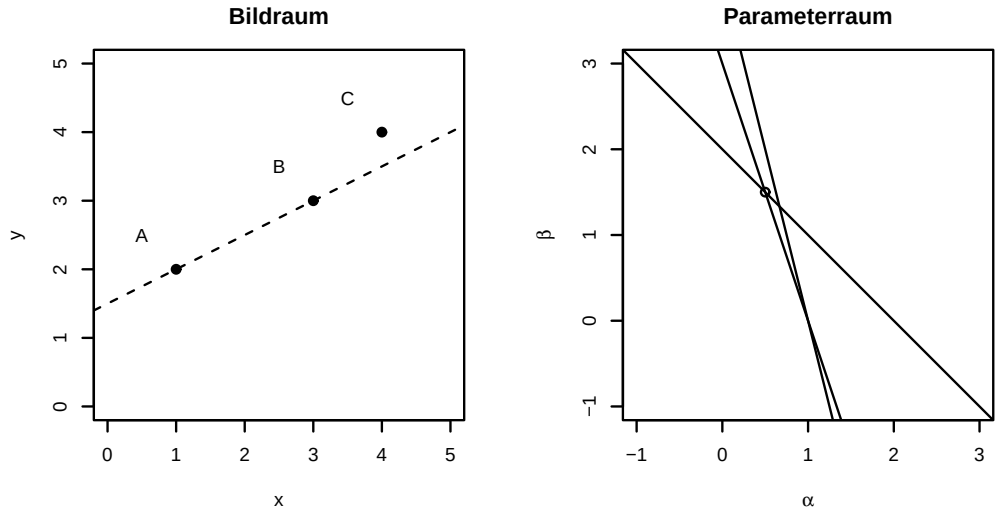

Abbildung 2.3: Hough-Transformation für drei Punkte.

zeichnet die Parameter derjenigen Geraden (im Bildraum), die durch die beiden entsprechenden Bildpunkte verläuft. Beispielsweise schneiden sich die zu den Punkten A und B gehörigen Geraden im Parameterraum im Punkt $(\alpha = 0.5, \beta = 1.5)$ (durch einen Kreis markiert). Die hieraus resultierende Gerade im Bildraum $y = 0.5x + 1.5$

(gestrichelte Linie) verläuft eben sowohl durch Punkt A als auch Punkt B.
Lägen *alle* Punkte auf einer Geraden, so würden sich wiederum alle Geraden im Parameterraum in *genau einem* Punkt schneiden, nämlich bei den wahren Parametern. Die Punkte liegen im allgemeinen aber nur *ungefähr* auf einer Geraden, daher vermutet man die wahren Parameter nun dort, wo die meisten Geraden verlaufen — und sich gegenseitig schneiden.
Über diese Schnittpunkte im Parameterraum (bei n unterschiedlichen Geraden gibt es $\frac{n\cdot(n-1)}{2}$ Schnittpunkte) kann nun die gesuchte Gerade ausfindig gemacht werden: es wird nach einem Häufungspunkt (Cluster) von Schnittpunkten im Parameterraum gesucht, und dieser wird dann als Parameterschätzer verwandt.
In Abbildung 2.4 sind 12 Bildpunkte und deren Hough-Transformation dargestellt; im ganz rechten Graph sind nur die Schnittpunkte der Geraden im Parameterraum eingezeichnet, und außerdem jeweils der Median (als ein ausreißerunempfindliches Lagemaß) der α und β (gestrichelte Linien). Nimmt man diese als Parameterschätzer, so erhält man die gestrichelte Gerade im linken Graphen.

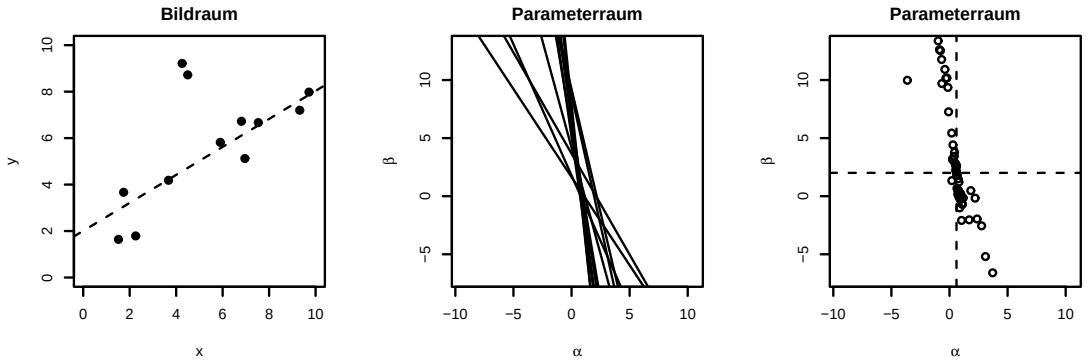

Abbildung 2.4: Hough-Transformation für 12 Punkte. Ganz rechts sind nur die Schnittpunkte und die Parameterschätzer eingezeichnet.

Eine andere Methode um mit Hilfe der Hough-Transformation zu Parameterschätzern zu gelangen funktioniert über eine Diskretisierung des Parameterraumes. Hier werden die einzelnen Parameter entlang ihrer Definitionsbereiche in Klassen eingeteilt; im Falle von zwei Parametern wie im Beispiel entstünde so ein zweidimensionales Raster. Dann werden nicht die Geradenschnittpunkte berechnet, sondern es wird (im soge-

nannten *Hough-Histogramm*) ausgezählt, welche Zellen im Raster von wievielen Geraden durchlaufen werden — dadurch wird wiederum bewertet, wie dicht die Geraden im Parameterraum liegen. Die Zelle mit den meisten Geraden (bzw. deren Mittelpunkt) dient dann zur Schätzung. Im obigen Beispiel zur Geradenanpassung könnte man die Parameter beispielsweise entlang der Grenzen $\{\ldots, \frac{1}{2}, 1\frac{1}{2}, 2\frac{1}{2}, \ldots\}$ aufteilen, und dann für alle (x_i, y_i) und ganzzahlige Werte von α nach (2.2) die zugehörigen β berechnen. In einer Tabelle trägt man laufend ein, welche Zellen von den Geraden geschnitten werden, und am Ende bestimmt man die am häufigsten gekreuzte Zelle und deren entsprechende Parameterwerte als Schätzer.
Ein offensichlicher Nachteil ist hier, daß in diesem Falle nur ganzzahlige Schätzer bestimmt werden können und allgemein eben nur eine diskrete Menge von Schätzern möglich ist. Allerdings ist diese Methode numerisch einfacher, da bei der ersteren Methode die Anzahl der zu betrachtenden Schnittpunkte im Quadrat mit der Anzahl der Bildpunkte wächst, während im zweiteren Fall „nur“ mit einer Matrix von konstanter Größe (die allerdings wiederum von Dimension und Klasseneinteilung des Parameterraumes abhängig ist) gearbeitet wird. Weiterhin ist die Suche nach Clustern (insbesondere wiederum bei größeren Datenmengen) relativ aufwendig im Vergleich zur Feststellung des Maximums in der Matrix.

Ein Vorteil der Hough-Transformation — im Gegensatz beispielsweise zur linearen Regression mit quadratischer Verlustfunktion — ist, daß Ausreißer (Punkte, die nicht auf der gesuchten Geraden liegen) einen geringen Einfluß auf die Parameterschätzung haben; sie liefert also vergleichsweise robuste Schätzungen. Außerdem können auch mehrere Geraden *gleichzeitig* gesucht werden, indem die Suche nicht auf einen Häufungspunkt beschränkt wird. Aktuelle Forschungen deuten an, daß möglicherweise Analogien zwischen Hough-Transformation und Mustererkennung auf neuronaler Ebene im Gehirn bestehen (Hopfield und Brody, 2000, 2001).
Zu Einzelheiten zur generellen Anwendung siehe auch Shapiro (1978) und Ballard (1981); zu statistischen Eigenschaften (wie z.B. Konvergenz und Robustheit) siehe Goldenshluger und Zeevi (2002).

2.2.2 Anwendung auf Audiodaten

Ehe die Daten durch den Computerchip verarbeitet werden, liegen diese zunächst in Form einer Zeitreihe $\{(t_i, y_i)\}_{i=1,\dots,N}$ vor. Die t_i sind hier die Zeitpunkte mit konstantem Abstand von $t_i - t_{i-1} = \frac{1}{44100}$ Sekunden, und die y_i sind die Samples. Jedes Element der Zeitreihe (t_i, y_i) stellt nun einen Bildpunkt dar, der Bildraum wird durch die Zeitachse und die Amplitudenachse aufgespannt.
Gesucht werden soll nach sogenannten *Signalflanken*. Der Begriff der Signalflanke entstammt der Physik und bezeichnet beispielsweise eine Sinusschwingung im Bereich $[0, \frac{\pi}{2}]$ (eine Viertelperiode) oder generell den über die Nullinie aufsteigenden Teil einer Schwingung. Eine komplexere Schwingung kann also auch mehrere Signalflanken haben.
Im vorliegenden Falle soll tatsächlich nach der Flanke einer Sinusschwingung mit unbekannter Amplitude und Phasenverschiebung (bei gegebener Frequenz) gesucht werden. Die Motivation dabei ist, daß sich der Klang durch die spezifische Aufeinanderfolge dieser Signalflanken charakterisieren läßt und so eine Zuordnung zu einem bestimmten Instrument möglich ist. Abbildung 2.5 zeigt einen Klang, dessen Signalflanken

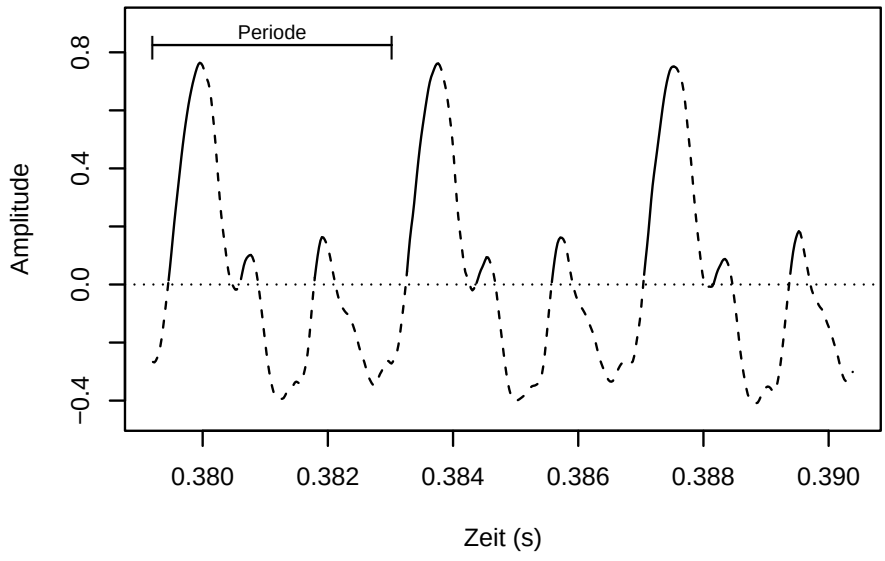

Abbildung 2.5: Charakteristische Signalflanken eines Klanges.

hervorgehoben sind; einen Flötenton (c4, Vibrato). Man sieht, daß sich innerhalb einer Periode jeweils drei Signalflanken mit unterschiedlicher Amplitude (und Steigung) abwechseln — diese Struktur soll zur Identifizierung ausgenutzt werden. Die Signalflanken sind hier offenbar keine reinen Sinuskurven, aber wenn sie einer solchen auch nur stückweise hinreichend ähneln, sollten sie trotzdem durch die Hough-Tranformation entdeckt werden.

2.2.3 Parametrisierung und Umsetzung

Die gesuchte Funktion hat zunächst die Form

$$y = A \cdot \sin(2\pi \cdot f \cdot t - \varphi) \tag{2.3}$$

(analog zur Gleichung (2.1) im Beispiel zur Geradenanpassung), wobei f, die sogenannte *Center-Frequency*, konstant und gleich 261 Hz ist. Dieser Wert wurde in der Diplomarbeit von Backes und Gerlach (2000) auf seinen Effekt hin untersucht und im Sinne einer möglichst zuverlässigen Erkennung der Signalflanken festgelegt. Freie Parameter sind die Amplitude $A \in [1, \infty[$ und die Phasenverschiebung $\varphi \in \mathbb{R}^+$. Als Signalflanke ist die Zielfunktion zunächst eine auf eine Viertelperiode (das Intervall $[0, \frac{1}{4f}]$) eingeschränkte Sinuskurve (siehe gestrichelte Linie in Abbildung 2.6). Die Amplitude (A) bewirkt eine Streckung der Signalflanke in y-Achsenrichtung und beeinflußt damit auch ihre Steigung; die Phasenverschiebung (φ) verschiebt sie in x-Achsenrichtung. Die Signalflanke ist auf einem Intervall der Länge $\frac{1}{4f}$ definiert, was hier (da $f = 261$) 0.001 Sekunden oder 42 Samples entspricht.
Um die Transformation für ein Sample (t_i, y_i) durchzuführen, muß man die obige Funktionsgleichung (2.3) nun folgendermaßen umformen:

$$\frac{1}{A} = \frac{1}{y_i} \cdot \sin(2\pi \cdot f \cdot t_i - \varphi) \tag{2.4}$$

so daß für gegebenes t_i, y_i und φ die Amplitude berechnet werden kann (entspricht Gleichung (2.2) im vorigen Beispiel).
Zur Bestimmungs der Häufungspunkte werden nun beide Parameter diskretisiert, also in Klassen aufgeteilt. Die Phasenverschiebung wird entlang der Klassengrenzen $\{\frac{1+2k}{88200}\}_{k=0,1,2,\ldots}$ aufgeteilt, womit die Klassenmittelpunkte gerade die $\{\frac{j}{44100}\}_{j=1,2,3,\ldots}$

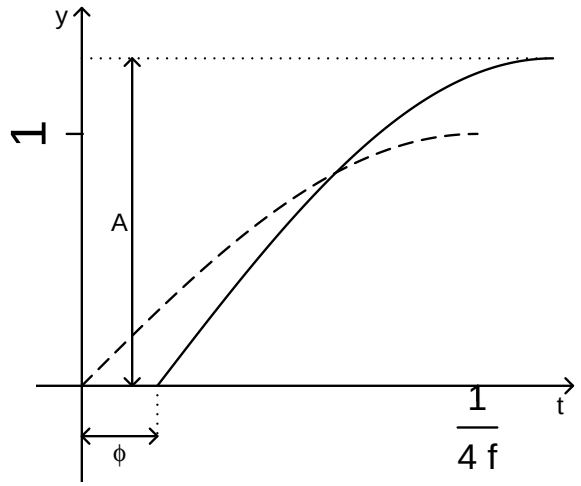

Abbildung 2.6: Wirkung der beiden Parameter Amplitude (A) und Phasenverschiebung (φ) auf die Form der Signalflanke.

sind; also entspricht ihre zeitliche Auflösung eben der Abtastrate der ursprünglichen Klangdatei. Die Amplitude ist in 32 Klassen mit den Grenzen $\{1, \frac{32}{31}, \ldots, \frac{32}{2}, 32, \infty\}$ aufgeteilt, wobei allerdings intern mit der inversen Amplitude $\frac{1}{A}$ und den entsprechenden inversen Klassengrenzen $\{0, \frac{1}{32}, \frac{2}{32}, \ldots, \frac{31}{32}, 1\}$ gerechnet wird, da sich die inverse Amplitude direkt aus (2.4) ergibt.

Das (zweidimensionale) Hough-Histogramm hat nun in einer Richtung 32 Klassen und in der anderen soviele wie der zu verarbeitende Klang Samples hat. Das bedeutet allerdings nicht, daß die zu bearbeitende Datenmenge beliebig groß werden kann: die Klangdatei wird von vorne nach hinten Sample für Sample abgearbeitet, und da die gesuchte Signalflanke nur eine begrenzte Länge (nämlich 42 Samples) hat, können zu einem gegebenen Zeitpunkt nur Signalflanken entdeckt werden, deren Phasenverschiebung im Bereich der vorangegangenen 42 Samples liegt. In den weiter zurückliegenden Zellen des Histogramms finden keine Veränderungen mehr statt; diese können sogar schon ausgewertet werden. Es müssen also „nur“ jeweils $32 \times 42 = 1344$ Histogrammzellen im Auge behalten werden.

Da das Ziel hier nicht ist, *eine* Signalflanke zu erkennen, sondern entlang der Klangda-

tei *viele* festzustellen, wird nun nach lokalen Maxima im Hough-Histogramm gesucht. Eine Signalflanke wird immer dann als vorhanden angenommen, wenn durch eine Histogrammzelle 4 oder mehr Kurven verlaufen. Ab 4 Kurven wird eine Signalflanke damit sozusagen als „signifikant“ bewertet; dieser Wert ist heuristisch gewählt und erkennt erfahrungsgemäß zuverlässig die Signalflanken.
Die Auswertung verläuft nun ähnlich wie in vorigen Beispiel (Abschnitt 2.2.1) beschrieben. Entscheidend ist, daß die geschätzten Parameter nur diskrete Werte annehmen können und die Auswertung laufend, Sample für Sample, in Echtzeit vorgenommen wird. Zur genauen Implementierung der Transformation siehe Epstein u. a. (2001); zur Anwendung auf Sinusschwingungen siehe auch Klefenz (1999) und Klefenz und Brandenburg (2003).
Abbildung 2.7 zeigt einen Klang mit den gefundenen Signalflanken; es sind die einzelnen Samples (t_i, y_i) als Kreise dargestellt und die entdeckten Signalflanken sind als Linien darübergelegt (vergleiche auch Abbildung 2.5). Die beiden größeren der drei Flanken pro Periode werden hier jeweils erkannt. An einer Stelle werden zwei dicht aufeinanderfolgende Flanken erkannt, was wahrscheinlich daran liegt, daß dieser „echte“ Klang natürlich keine reine Sinusform hat.

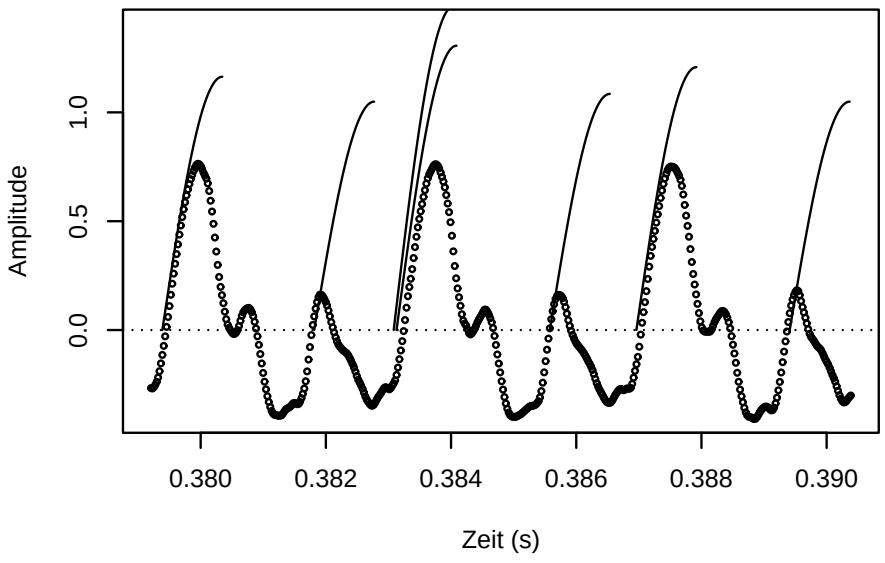

Abbildung 2.7: Gefundene Signalflanken am Beispiel.

2.3 Resultierendes Datenformat

Ergebnis der Hough-Transformation ist letztlich eine Liste von entdeckten Signalflanken, wobei jeweils die Phasenverschiebung und Amplitude als definierende Parameter der Signalflanke angegeben sind. Tabelle 2.1 zeigt einen Ausschnitt aus einem Da-

Tabelle 2.1: Das Datenformat nach der Transformation.

	Phasenverschiebung φ		Amplitude A	
Nr.	Sample	Sekunden	Klassen-Nr.	Wert
⋮	⋮	⋮	⋮	⋮
104	16731	0.3793881	28	1.163636
105	16838	0.3818141	31	1.049180
106	16894	0.3830841	22	1.488372
107	19896	0.3831291	25	1.306122
108	17004	0.3855781	30	1.084746
109	17065	0.3869611	27	1.207547
110	17173	0.3894101	31	1.049180
⋮	⋮	⋮	⋮	⋮

tensatz (die Daten zu Abbildung 2.7). Die Phasenverschiebung kann in Samples oder auch in Sekunden ausgedrückt werden, die Amplitude nimmt nur diskrete Werte an, von daher reicht die Angabe der Klassennummer prinzipiell aus. Man beachte, daß die Klassennummer antiproportional zur Amplitude ist; Klasse 32 entspricht also einer kleinen Amplitude und Klasse 1 einer großen. Aus diesen zwei Variablen können weitere abgeleitet werden, insbesondere beispielsweise die Zeitdifferenz zur vorhergehenden Signalflanke und ähnliche; durch Phasenverschiebungen und Amplituden ist jedoch die eigentliche Information komplett gegeben.
Abbildung 2.8 zeigt den kompletten Datensatz zum Klang aus Abbildung 2.7 und Tabelle 2.1 (Amplituden aufgetragen gegen die Phasenverschiebung). In diesem Falle sind es etwa 300 Datenpunkte. Die ersten Signalflanken wurden nach etwas mehr als 0.1 s festgestellt, und die Amplituden liegen zum größten Teil in den Klassen 20–32.

Die Daten stammen von zunächst 62 Intrumenten, von denen jeweils Tonsequenzen verschiedenen Umfanges eingespielt wurden. Genauere Einzelheiten dazu sind im

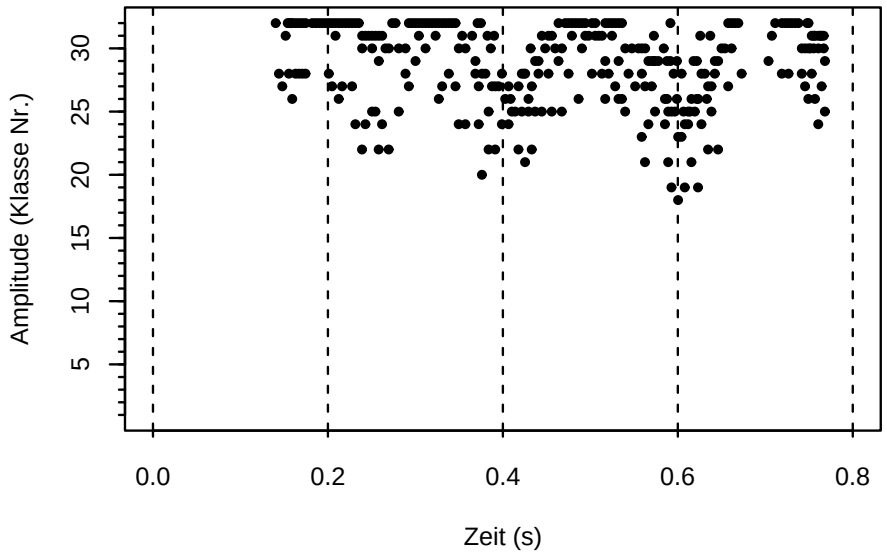

Abbildung 2.8: Gefundene Signalflanken über die Zeit.

Anhang in Tabelle A.1 und Abbildung A.1 (Seiten 60 und 63) dargestellt. Die Daten wurden dann zu größeren Gruppen angeordnet, indem sehr ähnlich klingende Instrumente jeweils zusammengefaßt wurden; zum Beispiel wurden die unterschiedlich laut angespielten Klaviertöne in einer Gruppe vereinigt, ebenso Fagott und Kontrafagott. Danach bleiben noch 25 unterschiedliche Instrumenten-Klassen übrig (siehe Tabelle A.2, Seite 62).

Die einzelnen Klänge sind naturgemäß auch von unterscheidlicher Dauer, im Mittel 2.8 Sekunden; kürzester und längster Ton dauern 0.03 und 12 Sekunden, 95% der Töne bewegen sich zwischen 0.17 und 7.4 Sekunden. Die Hough-transformierten Daten beziehen sich jeweils nur auf die ersten 0.77 Sekunden (34000 Samples) jedes Klanges; alle folgenden Ergebnisse der Klassifikation stützen sich also nur auf die Information, die nach den besagten ersten 0.77 Sekunden eines Klanges gegeben ist.

Der Umfang der Hough-transformierten Daten ist sehr unterschiedlich und hängt letztlich von der Sinusähnlichkeit der Signalform, der Frequenz und der Dauer des Klanges ab. Bei einigen Klängen wurden *keine* Signalflanken ausgelöst (bei 32 Klängen $\hat{=}$ 1.6%); Grund hierfür ist wahrscheinlich eine zu geringe Sinusähnlichkeit oder eine zu kleine Amplitude. 95% der Daten haben dann allerdings 8 oder mehr Signalflanken.

Im Mittel sind es etwa 400 Signalflanken, allerdings selten mehr als 1500. In Abbildung A.2 (Anhang, Seite 64) sind die Verteilungen der Anzahlen von Signalflanken noch einmal nach Intrumenten aufgeschlüsselt dargestellt.

3 Klassifikation

3.1 Das Klassifikationsproblem

Das Ziel dieser Arbeit ist es, Methoden zu untersuchen, mit denen man den im vorigen Kapitel beschriebenen Daten ein Instrument zuordnen kann. Methoden dieser Art sind *Klassifikationsverfahren*. Die vorherzusagende Variable ist *qualitativ*, nämlich das Instrument, das den jeweiligen Klang hervorgebracht hat. Das Instrument kann ein Klavier, eine Violine oder sonstige Vertreter aus einer endlichen Menge von Instrumententypen sein. Insbesondere besitzen die Instrumente auch keine natürliche Reihenfolge untereinander.
Das Klassifikationsverfahren soll nun also anhand der gegebenen Daten (die der Computerchip durch die Hough-Transformation der „rohen“ Audiodaten liefert) eine Entscheidung für eine aus einer begrenzten Menge von Klassen (die zur Auswahl stehenden Instrumente) liefern.
Das Problem kann man nun folgendermaßen formulieren: $I = \{i_1, i_2, \ldots, i_g | g \in \mathbb{N}\}$ ist die Menge der Klassen, wobei g, die Anzahl der Klassen, mindestens 2 beträgt. Ein Objekt ω gehört einer der Klassen an und an ihm können d Merkmale beobachtet werden. Der Merkmalsvektor X ist eine d-dimensionale Zufallsvariable. Die Verteilung der Zufallsvariablen hängt von der Klasse ab. Die Grundgesamtheit Ω läßt sich durch die Klassenzugehörigkeit ihrer Elemente in disjunkte Teilgesamtheiten $\Omega_1, \ldots, \Omega_g$ zerlegen.
Gesucht ist nun eine Entscheidungsfunktion der Form $f_\vartheta : \mathbb{R}^d \to I$, die vom Raum, der durch den Wertebereich der d-dimensionalen Zufallsvariablen aufgespannt wird, auf die Menge der Klassen abbildet. Die Parameter $\vec{\vartheta}$ der Funktion werden aus einem

„Trainingsdatensatz“ abgeleitet („gelernt“) (Hastie u. a., 2001).
Nach der Anpassung der Entscheidungsfunktion an die Daten kann man dieser also schließlich einen Merkmalsvektor übergeben und erhält dessen Klassifizierung.

Zum „Trainieren“ steht hier der Datensatz zur Verfügung (siehe auch Abschnitt 2.1), der aus den transformierten Daten zu einem Satz von Klängen besteht, bei denen das Instrument jeweils bekannt ist.
Nach der Transformation liegen die Daten zunächst als Zeitreihe vor, deren Länge durch den ursprünglichen Klang selber (und dessen Dauer) bestimmt ist. Vor der Klassifikation müssen die Daten deshalb noch auf eine feste Anzahl (d) von charakterisierenden Variablen reduziert werden.
Vom Klang bis zum klassifizierten Instrument sind es damit 4 Zwischenschritte, wie in Abbildung 3.1 dargestellt. Die Eingabe ist hier der „rohe“ Klang an sich, der im

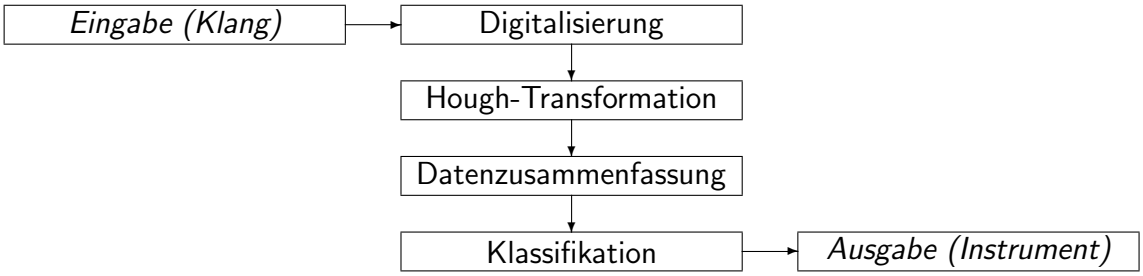

Abbildung 3.1: Die Schritte vom Klang zur Klassifikation.

ersten Schritt zunächst aufgenommen und digitalisiert werden muß. Danach folgt die Hough-Transformation, deren Resultat wiederum eine Zeitreihe variabler Länge ist. Im folgenden Schritt wird diese auf wenige Variablen kondensiert, mit deren Hilfe letztlich die Klassifikation durchgeführt werden kann. Ausgegeben wird schließlich das Instrument, welches dem Klang zugeordnet wird.
Die offenen Probleme sind nun noch, wie man die Daten am sinnvollsten zusammenfaßt und wie man schließlich anhand dieser Daten klassifiziert.

In den folgenden Abschnitten 3.2 und 3.3 wird die Zusammenfassung der Daten beschrieben, in den anschließenden Abschnitten (3.4 – 3.9) werden dann die einzel-

nen Klassifikationsverfahren vorgestellt. Deren Ergebnisse werden dann wiederum im nächsten Kapitel (4) diskutiert.

3.2 Datenaufbereitung

3.2.1 Besetzungszahlen

Eine einfache Art, die Hough-transformierten Daten eines Klanges zusammenzufassen, ist die Betrachtung der *Besetzungszahlen* der Amplitudenklassen. Abbildung 3.2 zeigt

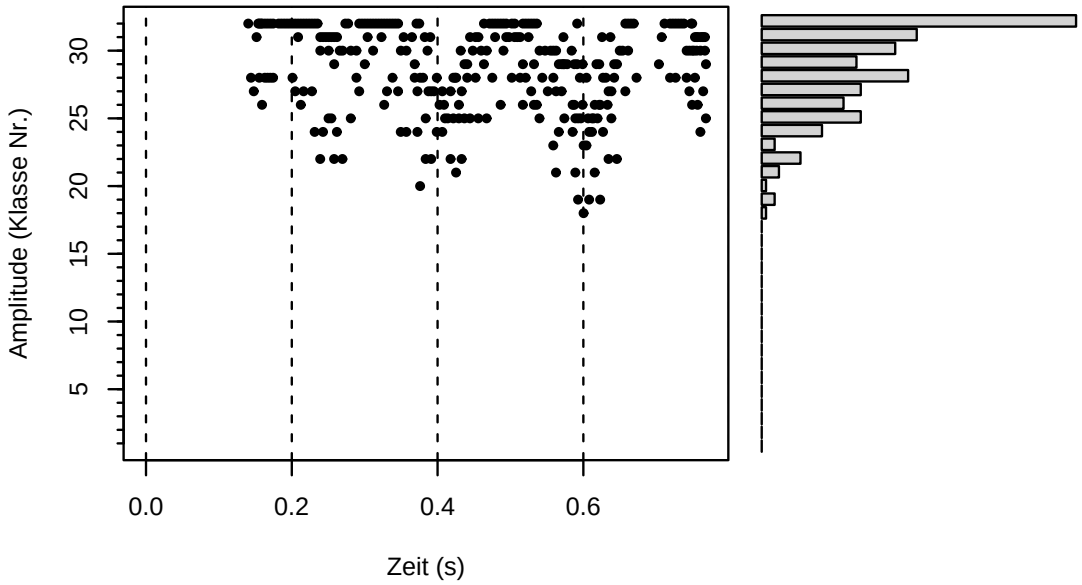

Abbildung 3.2: Besetzungszahlen der Amplitudenklassen.

die transformierten Daten zu einem Klang und dazu (an der rechten Seite) ein Balkendiagramm, das die Häufigkeit des Auftretens der 32 möglichen Amplitudenwerte über die Zeit insgesamt wiedergibt. Die Bildung der Besetzungszahlen ist wie gesagt sehr einfach, nur geht alle Information über die zeitliche Abfolge der einzelnen Signalflanken verloren.

Ergebnis dieser Aufbereitung sind für jeden einzelnen Hough-transformierten Klang (entsprechend der Anzahl der Amplituden-Klassen) 32 Variablen, die die einzelnen Häufigkeiten der Amplituden-Klassen angeben:

	Amplituden-Klasse				
Klang	**1**	**2**	$\cdots$	**31**	**32**
Flöte vibrato c4	0	0	$\cdots$	36	73
Flöte vibrato c♯4	0	0	$\cdots$	58	85
Flöte vibrato d4	0	0	$\cdots$	52	33
$\vdots$					

Äquivalent wäre auch die Angabe von *relativen Häufigkeiten* (also Häufigkeiten der Klassen geteilt durch die Gesamt-Anzahl von Signalflanken) und deren Gesamtsumme.

3.2.2 Hough-Charakteristika

Mit „Hough-Charakteristika“ sind Maßzahlen gemeint, die die Eigenheiten eines Hough-transformierten Klanges — und damit die Gemeinsamkeiten und Unterschiede zu anderen Klängen — widerspiegeln sollen. Die Daten aus der Transformation stellen sich zunächst dar wie in Abbildung 2.8 (Seite 17). Charakteristischen Eigenschaften sind dann zum Beispiel Maßzahlen, die Lage und Streuung der Amplitude beschreiben, wie etwa die *mittlere Amplitude* oder deren *Varianz*. Maßzahlen, die die zeitliche Abfolge der Signalflanken beschreiben, sollen ebenfalls konstruiert werden (zur Motivation siehe Abbildung 2.7, Seite 15: hier wechseln sich Signalflanken verschiedener Amplituden in bestimmten Zeitabständen ab).
Um die zeitliche Komponente zu erfassen, wird der „ursprüngliche“ Datensatz (wie in Abschnitt 2.3 vorgestellt) um zusätzliche Variablen erweitert. Aus den Phasenverschiebungen (Zeitpunkten) der Signalflanken wird z.B. jeweils die Zeitdifferenz zur vorhergehenden Signalflanke ($d_i := \varphi_i - \varphi_{i-1}$) abgeleitet und diese schließlich in eine Frequenz ($f_i := \frac{1}{d_i}$) umgerechnet. Tabelle 3.1 zeigt Beispieldaten mit diesen beiden neuen Variablen. Wenn nun bei einem Klang bei jeder Periode eine Signalflanke ausgelöst würde, so sollte die hier gemessene Frequenz gleich der Tonfrequenz sein. Sind es mehrere pro Periode, so steigen damit auch die Frequenzen.
Zusätzlich zur Zeitreihe der Amplituden (wie schon in Abbildung 2.8, Seite 17) ergibt sich so also eine Frequenzen-Zeitreihe (Abbildung 3.3).
Nun wird noch jede Signalflanke mit der jeweils Vorhergehenden in Beziehung gesetzt, indem das Verhältnis der Amplituden $\alpha_i := \frac{A_i}{A_{i-1}}$ und das Verhältnis der Zeitdifferenzen $\delta_i := \frac{d_i}{d_{i-1}} = \frac{f_{i-1}}{f_i}$ betrachtet wird. Ändert sich beispielsweise die Amplitude

Tabelle 3.1: Die transformierten Daten mit weiteren abgeleiteten Variablen.

	Phasenverschiebung φ		Amplitude A		**Zeit-**	**Frequenz**
Nr.	Sample	Sekunden	Klassen-Nr.	Wert	**differenz** d	f
⋮	⋮	⋮	⋮	⋮	⋮	⋮
104	16731	0.3793881	28	1.163636	0.001326	753.86
105	16838	0.3818141	31	1.049180	0.002426	412.20
106	16894	0.3830841	22	1.488372	0.001270	787.40
⋮	⋮	⋮	⋮	⋮	⋮	⋮

nicht über die Zeit, so ist der Amplitudenquotient konstant 1, ansonsten wechseln sich größere und kleinere Werte ab.

Charakteristisches Verhalten der Signalflanken über längere Zeiträume sollen ebenfalls berüchsichtigt werden. So werden z.B. Veränderungen der Amplitude über die Gesamtzeit durch die Veränderung der mittleren Amplituden von der ersten zur zweiten Hälfte erfaßt.

Die hieraus abgeleiteten Variablen sind Maßzahlen der univariaten Verteilungen der Amplituden und Frequenzen (Lage-, Streuungs-, Schiefemaße usw.), sowie Maße, die deren gemeinsame Verteilung oder die zeitliche Aufeinanderfolge beschreiben (Kovarianzmaße).

Eine genaue Auflistung der betrachteten Variablen findet sich in Tabelle A.3 auf Seite 65 im Anhang; die dazugehörigen Formeln sind in Abschnitt B.1 (Seite 70, ebenfalls im Anhang) zusammengestellt. Einige Variablen haben eine sehr schiefe Verteilung, die insbesondere der unterstellten Normalverteilung bei der Diskriminanzanalyse (wird in Abschnitt 3.4 eingeführt) widersprechen würde. Diese Schiefe konnte in einigen Fällen durch Logarithmieren der betreffenden Variablen „repariert“ werden, daher sind einige Variablen entsprechend transformiert.

Die Hough-Charakteristika können nur berechnet werden, wenn dafür genügend Signalflanken (z.B. zur Varianzschätzung) zur Verfügung stehen. Bei 88 Klängen (≙4.4%) war dies nicht der Fall; diese wurden daher aussortiert.

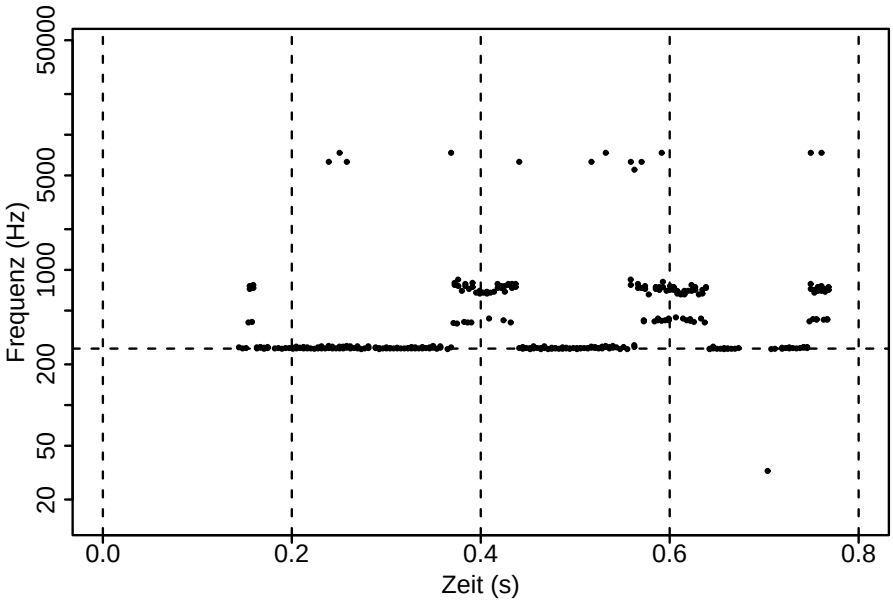

Abbildung 3.3: Frequenzen über die Zeit. Die horizontale Linie zeigt die Tonfrequenz an.

3.2.3 Clusteranalyse

In Abbildung 2.7 (Seite 15) konnte man sehen, daß sich hier in jeder Periode in bestimmten Zeitabständen Signalflanken bestimmter Amplituden abwechseln. In diesem Falle war es jeweils eine Signalflanke hoher Amplitude, der eine kürzere Zeitspanne vorausging, nach einer längeren Zeitspanne wiederum gefolgt von einer Signalflanke kleinerer Amplitude. Demnach müßte es hier zwei Gruppen von Signalflanken geben: einerseits mit großer Amplitude und kleiner Zeitdifferenz und andererseits kleine Amplidude und große Differenz. Diese Gruppen sollten dann charakteristisch sein für das Instrument.

In Abbildung 3.4 sind (für ein anderes Instrument) die beiden Merkmale Amplitude und Zeitdifferenz gegeneinander aufgetragen, und es sind tatsächlich drei verschiedene Gruppen (*Cluster*) von Signalflanken zu unterscheiden; außerdem sind noch einige „Ausreißer“ zu erkennen, die nicht in diese Gruppen fallen. Der eingespielte Ton war ein „d♯4“ und hat damit eine Frequenz von 311 Hz. Bei 311 Hz beträgt die Periode des Klanges 0.0032 Sekunden oder 142 Samples; die waagerechten Linien im Graphen

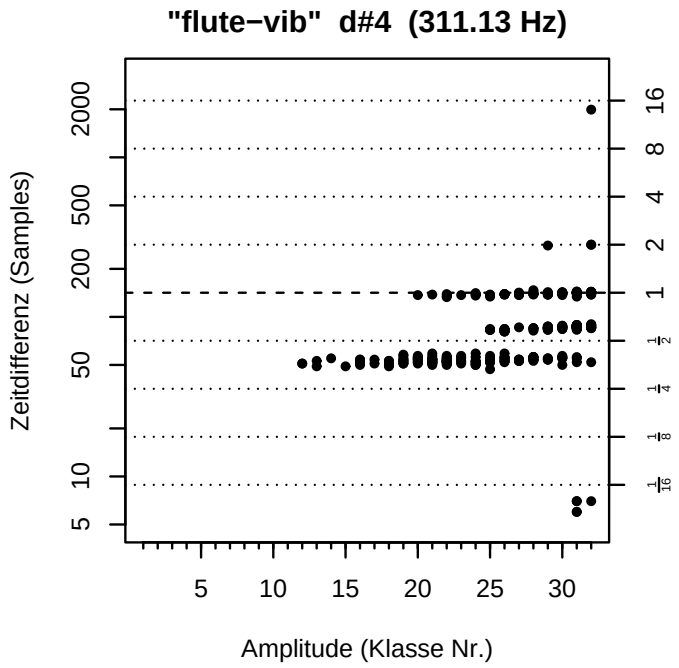

Abbildung 3.4: Clusterstruktur in den Daten.

zeigen diese Periode sowie Vielfache (Viertel, Halbe, Doppelte,...) an. Von daher liegt ein Cluster ziemlich genau bei der Tonfrequenz, die anderen beiden liegen jeweils etwas oberhalb und unterhalb der halben Periode, und alle drei unterscheiden sich auch in der Verteilung der Amplituden.

Cluster dieser Art finden sich auch bei anderen Klängen in verschiedener Anzahl, charakteristisch ist dabei auch ihre oft längliche Form. Um diese Cluster automatisch zu trennen (und außerdem zwischen „echten“ Clustern und Ausreißern zu unterscheiden), wurde daher ein *hierarchisches Clusterverfahren* mit *complete linkage* benutzt (Mardia u. a., 1979). Cluster, die im Verhältnis zur Gesamtzahl von Signalflanken „zu klein“ sind (kleiner als $\sqrt{N}$), werden als Ausreißer betrachtet.

Wendet man dies auf die obigen Daten an, so werden diese auch entsprechend zerlegt (siehe Abbildung 3.5). Die Ausreißer werden als solche erkannt und sind als Kreuze markiert. Die übrigen Signalflanken sind ihrem entsprechenden Cluster zugeordnet und jeweils durch Dreieck, Kreis bzw. Raute dargestellt. Hieraus kann man nun Kennzahlen ableiten, die Anzahl, Lage und Größe der Cluster beschreiben:

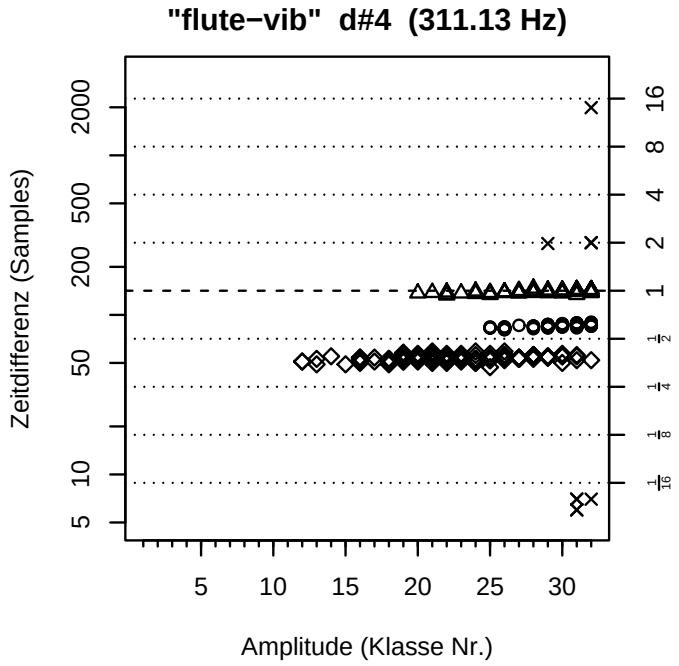

Abbildung 3.5: Gefundene Cluster.

Cluster	**Umfang**	**mittlere Amplitude**	**...**
1	143	22.69	...
2	111	29.56	...
3	106	29.35	...
Ausreißer	9	–	...

Betrachtete Variablen sind hier jeweils Mittel und Standardabweichung von Amplitude und Differenz, wobei die Differenzen d_i vorher noch auf die Tonfrequenz f normiert und logarithmiert werden: $\Delta_i := \log_2(d_i f)$, so daß die Δ_i nun die Lage der Signalflanken auf den Linien im Graphen wiedergeben.

Ein entscheidender Nachteil der Datenaufbereitung durch Clustering ist, daß Clusterverfahren immer sehr rechenaufwendig sind. Es sind auch bei weitem nicht bei allen Klängen Cluster in den Daten vorhanden; und wenn, so sind diese so unterschiedlich in ihrer Struktur (Anzahl, Größe, Form, Lage zueinander, ...), daß kein Clusterverfahren diese (optisch erkennbaren) Cluster hinreichend zuverlässig trennen konnte. So wurde dieser Ansatz schließlich fallengelassen.

3.3 Kurzer Datenüberblick

Abbildung 3.6 zeigt die Ähnlichkeiten und Unterschiede von Klängen gleicher und verschiedener Instrumente untereinander. Es handelt sich um jeweils 2 aufeinanderfolgende Töne von Klavier und Trompete. Beide weisen charakteristische Muster auf, die allerdings auch in einem gewissen Spielraum variieren.

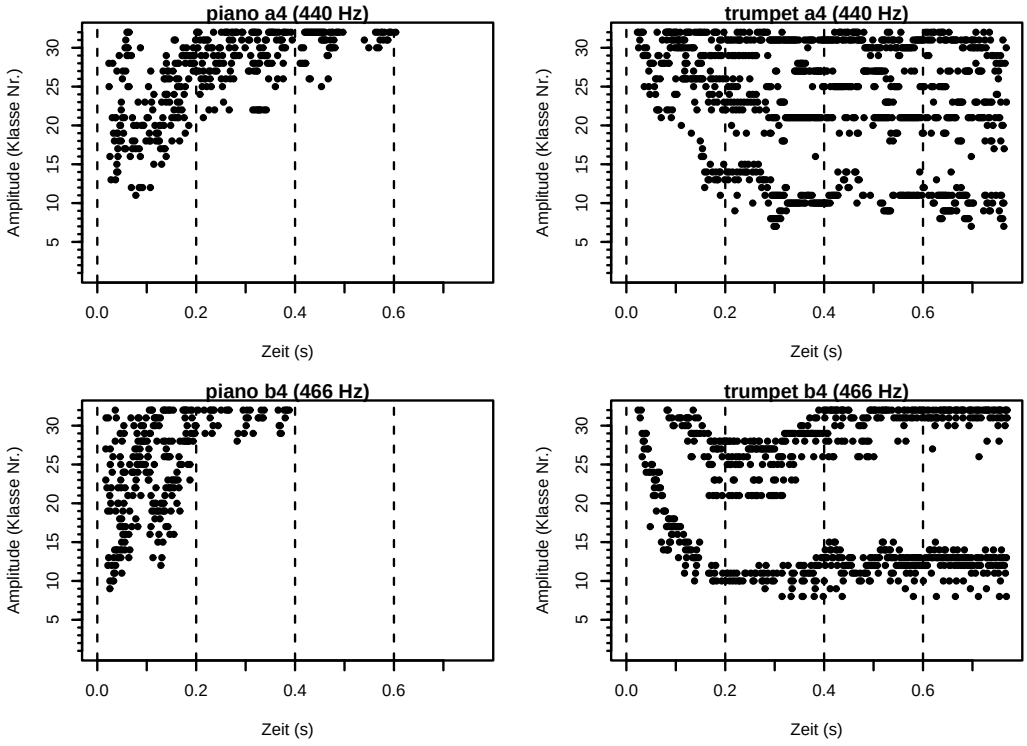

Abbildung 3.6: Unterschiede in den Hough-transformierten Daten.

Abbildung 3.7 zeigt die Besetzungszahlen der 4 Klänge aus der vorigen Abbildung als Histogramme. Auch hier sind Gemeinsamkeiten und Unterschiede zwischen Klängen und Instrumenten zu erkennen; es ist deutlich, daß sich die Besetzungszahlen *zwischen* den Instrumenten stärker unterscheiden als *innerhalb* der Instrumente. Die jeweiligen Gesamtsummen von Signalflanken sind ein weiteres Indiz: bei den Klavierklängen betragen sie jeweils 370 und 272, bei den Trompetentönen 981 und 740.
Es sind hier die „kleinen“ Amplitudenklassen (entsprechend den großen Amplituden)

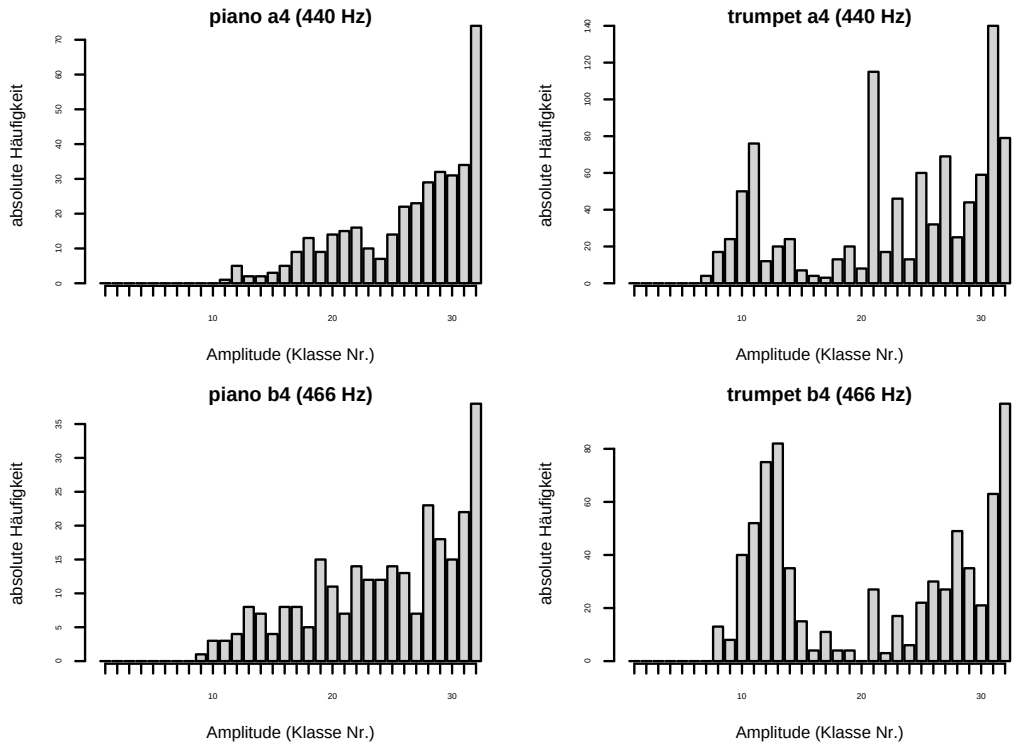

Abbildung 3.7: Amplituden-Histogramme der 4 Klänge.

unbesetzt — das ist keine Einzelerscheinung, sondern durchgängig der Fall. Die Klassen 1–3 sind bei den gegebenen Daten in keinem Falle besetzt.

In Abbildung 3.8 sind nun zwei der abgeleiteten Variablen für *alle* Trompeten- und Klavierklänge gegeneinander abgetragen: die schwarzen Punkte sind die Trompetenklänge, die weißen Punkte sind die Klavierklänge. Die Variablen sind die (logarithmierte) Tonfrequenz und die Veränderung der mittleren Amplitude über die Zeit. Man sieht zunächst, daß sich die Trompetentöne in einem engeren Frequenzbereich bewegen als das Klavier. Weiterhin ist die Amplituden-Mittelwertverschiebung bei der Trompete eher negativ und beim Klavier eher positiv — das deckt sich mit Abbildung 3.6: beim Klavier steigt die Amplitude über die Zeit, bei der Trompete sinkt sie ein wenig.

Anhand dieser Beobachtungen deutet sich schon eine Klassifikationsregel für Klavier und Trompete an: ein neuer Ton unbekannter Herkunft wird als Klavier klassifiziert,

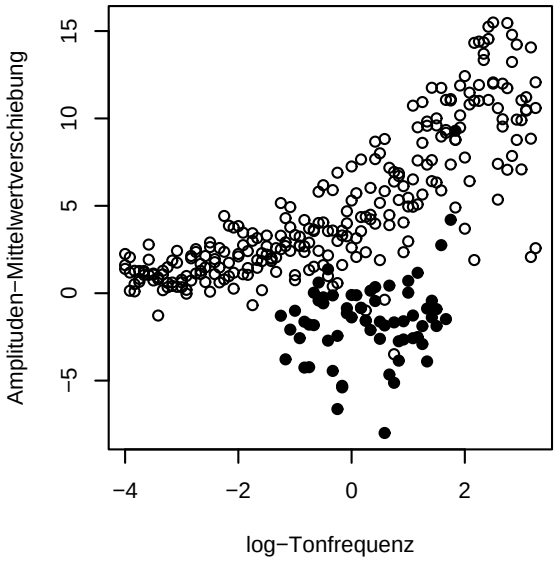

Abbildung 3.8: Unterschiede in 2 abgeleiteten Variablen
(schwarz: Trompete, weiß: Klavier).

wenn er eine positive Mittelwertverschiebung oder eine extreme Frequenz hat. Hat er dagegen eine negative Mittelwertverschiebung und eine mittlere Frequenz, handelt es sich wahrscheinlich um eine Trompete.

3.4 Diskriminanzanalyse

3.4.1 Lineare Diskriminanzanalyse (LDA)

Bei der Diskriminanzanalyse wird grundsätzlich unterstellt, daß die gemessenen Merkmale für jede einzelne Klasse einer (multivariaten) Normalverteilung folgen.
Im Falle der Linearen Diskriminanzanalyse wird diese Annahme wie folgt formuliert:

$$X|k = i \quad \sim \quad \mathrm{N}(\mu_i, \Sigma). \tag{3.1}$$

Das heißt, bei gegebener Klasse $k = i$ sind die Merkmale normalverteilt um einen Mittelwertvektor μ_i, der von der Klasse i abhängt. Die Kovarianzmatrix Σ ist für alle

Klassen gleich. Die bedingte Dichte von X ergibt sich dann folgendermaßen:

$$f(x|k=i) = \frac{1}{(2\pi)^{\frac{d}{2}}\sqrt{|\Sigma|}} \cdot \exp\left(-\frac{1}{2}(x-\mu_i)'\Sigma^{-1}(x-\mu_i)\right). \tag{3.2}$$

Die Parameter $\mu_1, \ldots, \mu_g$ und Σ sind im vorhinein nicht bekannt und werden daher aus den Trainingsdaten geschätzt, bei denen die Klassenzugehörigkeit bekannt ist. Zur Schätzung werden dabei das arithmetische Mittel und die empirische Kovarianz benutzt.

Um eine *optimale* Entscheidungsregel festzulegen, müßte man prinzipiell zunächst eine *Verlustfunktion* (oder auch Kostenfunktion) definieren, die den „Schaden“ einer Fehlentscheidung bemißt. In medizinischen Fragestellungen ist es beispielsweise in der Regel so, daß eine Fehldiagnose eines kranken Patienten als gesund einen sehr schwerwiegenden Fehler darstellt. Die fälschliche Einstufung eines gesunden Patienten als krank ist dagegen weniger gravierend.
Außerdem wäre eine *a-priori-Verteilung* über die Klassen notwendig; es müßte also bekannt sein, mit welchen Wahrscheinlichkeiten die verschiedenen Klassen auftreten. Im medizinischen Beispiel hieße dies, daß man bei einer Routineuntersuchung im vorhinein schon weiß, daß der Patient mit wesentlich größerer Wahrscheinlichkeit gesund ist als krank.
Eine (in gewissem Sinne) optimale Entscheidungsregel wäre dann diejenige Regel, die den *erwarteten Verlust* (also den Erwartungswert des Verlustes oder den mittleren Verlust auf lange Sicht) minimiert.
Beides erübrigt sich allerdings, wenn man sowohl Verlustfunktion als auch a-priori-Wahrscheilichkeiten als konstant annimmt; wenn man also Fehlentscheidungen in allen Richtungen als gleich schwerwiegend beurteilt und außerdem keine Klasse als wahrscheinlicher als eine andere annimmt. Auf die Instrumentenerkennung bezogen bedeutet das, daß eine fälschliche Klassifikation eines Flötentons als eine Geige genauso schwer wiegt wie eine Fehlklassifikation eines Klaviers als Glockenspiel und so weiter. Außerdem wäre nicht ein Saxophon von vornherein wahrscheinlicher als eine Trompete oder ähnliches. In diesem Falle ist dann die optimale Entscheidungsregel gleich der *Maximum-Likelihood-Entscheidungsregel*, die dem beobachteten Merkmalsvektor x

diejenige Klasse $\hat{k}$ zuordnet, für die gilt:

$$f(x|\hat{k}) \geq f(x|i) \quad \text{für } i = 1, \ldots, g, \tag{3.3}$$

die also die *Likelihood* $L(k|x) = f(x|k)$ maximiert (Fahrmeir u. a., 1996).

Um unter dem unterstellten Normalverteilungsmodell zu klassifizieren, müssen also zunächst Schätzer für die unbekannten Parameter $\mu_1, \ldots, \mu_g$ und Σ bestimmt werden. Für eine neue Beobachtung x wird dann die Likelihood $L(k|x)$ für $k = 1, \ldots, g$ bestimmt und diejenige Klasse k gewählt, für die die Likelihood am größten ist.
Abbildung 3.9 zeigt zwei bivariate Normalverteilungen: es ist jeweils eine Höhenlinie

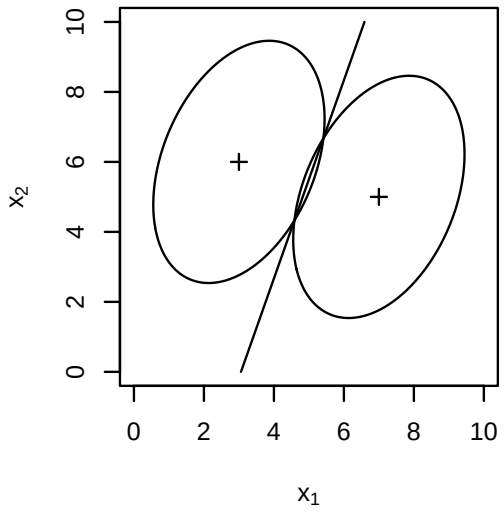

Abbildung 3.9: Modell und Diskriminanzfunktion bei der LDA.

der Dichten zu sehen; entsprechend dem Modell der LDA unterscheiden sich die beiden Verteilungen in der Lage, das Streuungsverhalten ist jedoch gleich. Die Gerade beschreibt die Linie, entlang derer die beiden Dichten (und damit die Likelihoods) gleich groß sind (die *Diskriminanzfunktion*). Sie illustriert die Entscheidungsregel: alle zu klassifizierenden Merkmalsvektoren x, die rechts oder links der Linie fallen, werden der entsprechenden Klasse zugeordnet. Im Falle der LDA ist die Diskriminanzfunktion immer linear, bzw. bei mehr als zwei Klassen stückweise linear (Fahrmeir u. a., 1996).

3.4.2 Quadratische Diskriminanzanalyse (QDA)

Wie bei der LDA wird auch bei der Quadratischen Diskriminanzanalyse eine Normalverteilung unterstellt. Die Annahme ist hier:

$$X|k = i \quad \sim \quad \mathrm{N}(\mu_i, \Sigma_i). \tag{3.4}$$

Die Merkmale innerhalb einer Klasse sind demnach normalverteilt, wobei hier sowohl Mittelwert μ_i als auch Kovarianz Σ_i von der jeweiligen Klasse i abhängen. Die Dichte ergibt sich analog zu Gleichung (3.2), nur daß anstelle der gemeinsamen Kovarianz Σ jede Klasse eine individuelle Kovarianz Σ_i besitzt.
Die Klassifikation verläuft wiederum analog zur LDA anhand der Likelihood. Wie man in Abbildung 3.10 sieht, ist die Diskriminanzfunktion allerdings nicht mehr linear, sondern (stückweise) quadratisch (Fahrmeir u. a., 1996).

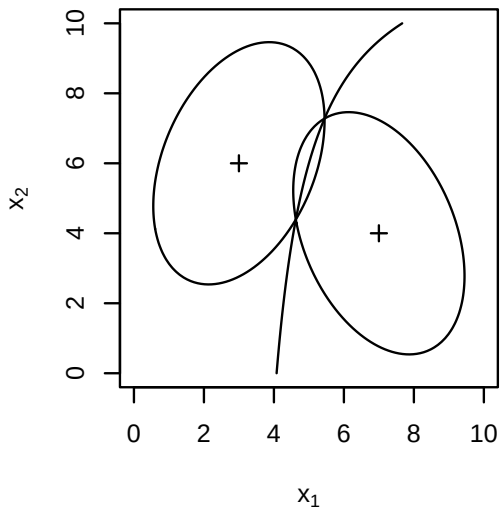

Abbildung 3.10: Modell und Diskriminanzfunktion bei der QDA.

Im Vergleich zur LDA zeichnet sich die QDA durch weniger restriktive Annahmen aus, ein großer Nachteil ist allerdings die erheblich größere Anzahl zu schätzender Parameter. Waren es bei der LDA noch eine symmetrische $(d \times d)$-Kovarianzmatrix und für jede Klasse ein d-dimensionaler Mittelwertvektor, so kommen für die QDA $(g - 1)$ weitere Kovarianzmatrizen dazu. Und wärend für die Schätzung der gemeinsamen Kovarianz bei der LDA die kompletten Trainingsdaten zur Verfügung standen,

werden die Klassenkovarianzen natürlich nur aus den Beobachtungen der jeweiligen Klasse geschätzt.
Diese beiden Faktoren (mehr Parameter, dabei weniger Beobachtungen zur Schätzung) führen dazu, daß dieses Modell oft schlechter funktioniert als das der LDA, da hier die Varianzen der Parameterschätzer zu groß sind; die geschätzten Parameter haben also eine größere Abweichung von den wahren Parametern. Im Extremfall vieler Variablen, weniger Beobachtungen und hoher Korrelation zwischen den Variablen tritt sogar oft das Problem auf, daß die Schätzungen nicht nur ungenau sind, sondern zu nicht invertierbaren (singulären) Kovarianzmatrizen führen und damit eine Klassifikation unmöglich machen.
Zwei Ansätze um die Nachteile der QDA gegenüber der LDA zu beheben, sind *Naive Bayes* und die *Regularisierte Diskriminanzanalyse (RDA)*, die in den folgenden Abschnitten beschrieben werden.

3.4.3 Naive Bayes

Das Naive-Bayes-Modell ist zunächst prinzipiell das gleiche wie bei der QDA, also Normalverteilung mit individuellen Klassenkovarianzen, nur unterliegen die Kovarianzen weiteren Restriktionen. Die zusätzliche („naive“) Annahme ist die der *bedingten Unabhängigkeit* der Merkmale gegeben die Klasse (Hastie u. a., 2001).
Bei der QDA hat eine Kovarianzmatrix für eine Klasse folgendes Aussehen:

$$\Sigma_k^{\mathrm{QDA}} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1d} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{d1} & \sigma_{d2} & \cdots & \sigma_{dd} \end{pmatrix} \tag{3.5}$$

Auf der Hauptdiagonalen stehen die Varianzen der einzelnen Variablen, also σ_{ii} $(= \sigma_i^2)$ für Variable i. Auf den Nebendiagonalen stehen die paarweisen Kovarianzen zwischen den Variablen, also σ_{ij} für die Kovarianz zwischen Variablen i und j, wobei $\sigma_{ij} = \sigma_{ji}$. Unter Normalverteilung folgt aus der angenommenen Unabhängigkeit auch Unkorreliertheit der Variablen, womit $\sigma_{ij} = 0$ für alle $i \neq j$. Die Klassen-Kovarianzmatrix

vereinfacht sich beim Naive Bayes so zur folgenden Form:

$$\Sigma_k^{\mathrm{NB}} = \begin{pmatrix} \sigma_{11} & 0 & \cdots & 0 \\ 0 & \sigma_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_{dd} \end{pmatrix} \tag{3.6}$$

Es bleiben also (pro Klasse) nur noch d zu schätzende Parameter übrig von ursprünglich $\frac{d(d+1)}{2}$ bei der QDA. Abbildung 3.11 zeigt die Diskriminanzfunktion des Naive Bayes, die wie bei der QDA quadratisch ist. Die Restriktion macht sich anschaulich

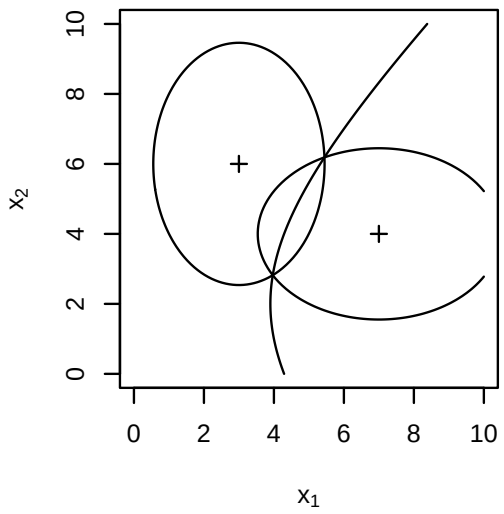

Abbildung 3.11: Modell und Diskriminanzfunktion beim Naive Bayes.

dadurch bemerkbar, daß sich die Ellipsen konstanter Dichte nur noch entlang der Hauptachsen ausrichten, also nicht mehr in beliebige Richtungen geneigt sein können.

3.4.4 Regularisierte Diskriminanzanalyse (RDA)

Die Regularisierte Diskriminanzanalyse stellt eine Erweiterung der QDA dar, die aber sowohl QDA als auch LDA mit einschließt. Das Modell ist prinzipiell zunächst wiederum das der QDA (Normalverteilung, individuelle Gruppenkovarianzen), nur werden hier die Kovarianzmatrizen noch mit Hilfe zweier Parameter manipuliert. Die Motivation hierbei ist einerseits, die große Varianz der Schätzer bei der QDA zu vermindern,

ohne dabei diesen Ansatz ganz fallenzulassen. Andererseits wird versucht, eine mögliche Singularität der Matrizen zu „reparieren“.
Ausgegangen wird vom Schätzer der gemeinsamen (*gepoolten*) Kovarianz $\hat{\Sigma}$ (wie bei der LDA) und den individuellen Klassenkovarianzschätzern $\hat{\Sigma}_k$ $(k = 1, \ldots, g)$ (wie bei der QDA). Hieraus wird anhand des zusätzlichen Parameters $\lambda \in [0, 1]$ ein gewichtetes Mittel

$$\hat{\Sigma}_k(\lambda) = (1 - \lambda)\hat{\Sigma}_k + \lambda\hat{\Sigma} \tag{3.7}$$

aus beiden Schätzern gebildet. λ gibt hier also das Gewicht der gepoolten Varianz an. Der zweite Parameter $\gamma \in [0, 1]$ erlaubt dann weiterhin eine „Verschiebung“ dieser Kovarianzmatrizen in Richtung der Einheitsmatrix (bzw. eines Vielfachen):

$$\hat{\Sigma}_k(\lambda, \gamma) = (1 - \gamma)\hat{\Sigma}_k(\lambda) + \gamma\frac{1}{d}\mathrm{tr}[\hat{\Sigma}_k(\lambda)]\mathrm{I} \tag{3.8}$$

Der Vorfaktor $\hat{\sigma}^2 := \frac{1}{d}\mathrm{tr}[\hat{\Sigma}_k(\lambda)]$ der Einheitsmatrix ist dabei das arithmetische Mittel der Hauptdiagonalelemente von $\hat{\Sigma}_k(\lambda)$ und damit die gemittelte Varianz der einzelnen Variablen unter Annahme der Klassenkovarianz $\hat{\Sigma}_k(\lambda)$. In diesem gewichteten Mittel ist γ das Gewicht der (skalierten) Einheitsmatrix (Friedman, 1989).

Tabelle 3.2: Die vier Extremfälle der RDA.

	Parameter		**Form der**	**Anzahl**
Fall	λ	γ	**Kovarianz**	**Parameter**
I	0	0	$\hat{\Sigma}_k$	$g \cdot \frac{d(d+1)}{2}$
II	1	0	$\hat{\Sigma}$	$\frac{d(d+1)}{2}$
III	0	1	$\hat{\sigma}_k^2\mathrm{I}$	g
IV	1	1	$\hat{\sigma}^2\mathrm{I}$	1 (bzw. 0)

Für die extremen Werte der Parameter λ und γ reduziert sich die Form der Klassenkovarianz jeweils auf einen der folgenden Spezialfälle (vgl. Tabelle 3.2):

I *QDA*: jede Klasse hat eine individuelle Kovarianzmatrix.

II *LDA*: alle Klassen haben dieselbe Kovarianzmatrix.

III *bedingt unabhängige Variablen*: innerhalb jeder Klasse sind die Variablen bedingt unabhängig (gegeben die Klasse) ähnlich wie beim Naive Bayes — nur sind

die Varianzen der Variablen innerhalb einer Klasse (die Diagonalelemente der Klassen-Kovarianzmatrix) gleich.

IV *Klassifikation anhand euklidischem Abstand*: wie Fall III, nur sind zusätzlich die Varianzen für alle Gruppen gleich. Dies führt bei der Klassifikation dazu, daß eine neue Beobachtung derjenigen Klasse zugeordnet wird, zu dessen Mittel sie den geringsten *euklidischen* Abstand hat (Fahrmeir u. a., 1996); und das *unabhängig* von $\hat{\sigma}^2$, womit in diesem Falle also die Varianzschätzung komplett hinfällig ist.

Von Fall I ($\lambda = \gamma = 0$) zu Fall IV ($\lambda = \gamma = 1$) reduziert sich also die Anzahl der zu schätzenden Varianzparameter von $g \cdot \frac{d(d+1)}{2}$ auf 1 (bzw. 0, da der Wert des verbleibenden Parameters $\hat{\sigma}^2$ bedeutungslos ist). Es bleiben allerdings in jedem Falle die Klassenmittel $\mu_1, \ldots, \mu_g$ (jeweils d-dimensional) zu schätzen.
λ regelt letztlich die Gleichheit/Verschiedenheit der Klassenkovarianzen, und γ regelt den Grad der Korreliertheit der Variablen (gegeben eine Klasse).
Die Parameterwahl kann letztlich anhand der geschätzten Fehlklassifikationsrate getroffen werden. In diesem Falle wurde die optimale Parameterkombination mit Hilfe eines Nelder-Mead-(Simplex-)Algorithmus (Press u. a., 1992) bestimmt.

3.5 Support Vector Machines

Die *Support Vector Machine* („Stützvektormaschine“) ist eine Erweiterung des sogenannten *Support Vector Classifiers*, bei dem versucht wird, zwei Klassen durch eine Hyperebene zu trennen. Abbildung 3.12 zeigt Daten, die zwei Klassen angehören (weiße und schwarze Punkte) und eine Gerade (die durchgezogene Linie), die die beiden Klassen trennt. Es sind hier viele Geraden möglich, die die Klassen trennen; beim

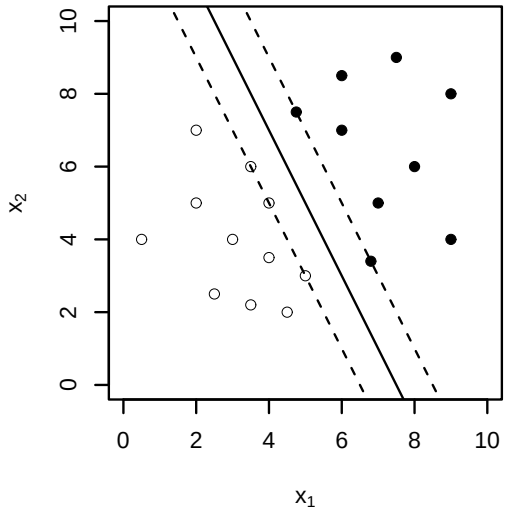

Abbildung 3.12: Trennende Hyperebene beim Support Vector Classifier.

Support Vector Classifier wird diejenige Gerade bestimmt, die den größten *Rand* zu beiden Seiten (durch die gestrichelten Linien angedeutet) freiläßt. Die Beobachtungen, die genau auf dem Rand liegen, sind dann die Stützvektoren.
Der Support Vector Classifier funkioniert allerdings nur, solange

a) die Klassengrenzen linear sind,

b) die Klassen sich nicht überlappen und

c) es genau 2 Klassen gibt.

Das Problem nichtlinearer Klassengrenzen wird gelöst, indem die Daten in einen höherdimensionalen Raum projiziert werden, in dem sie dann linear trennbar sind. Um auch

mit überlappenden Klassen arbeiten zu können, werden auch Beobachtungen innerhalb des Randes und jenseits der trennenden Hyperebene (auf der „falschen“ Seite) zugelassen, deren Einfluß auf die Bestimmung der Ebene dann herabgewichtet wird. Für mehr als 2 Klassen werden jeweils paarweise Klassifikatoren bestimmt und die Klassifikation schließlich durch einen Abstimmechanismus zwischen diesen ermittelt. Ein Algorithmus, das all das leistet, ist dann eine Support Vector Machine. Das Problem läßt sich schließlich als ein *quadratisches Optimierungsproblem* formulieren, das mit bekannten Methoden gelöst werden kann. Variierbare Parameter sind noch die *Kernfunktion*, mit deren Hilfe die Projektion in den hochdimensionalen Raum stattfindet, und der *Kostenparameter*, der die „Kosten“ einer Restriktionsverletzung bemißt (Meyer, 2001; Hastie u. a., 2001).

3.6 Klassifikationsbäume

Bei Klassifikationsbäumen wird der Raum wiederholt entlang der Hauptachsen geteilt, so daß der Raum letztlich in Rechtecke aufgeteilt wird. Abbildung 3.13 zeigt Beispieldaten (zwei Variablen x_1 und x_2 und zwei Klassen A und B, die als weiße und

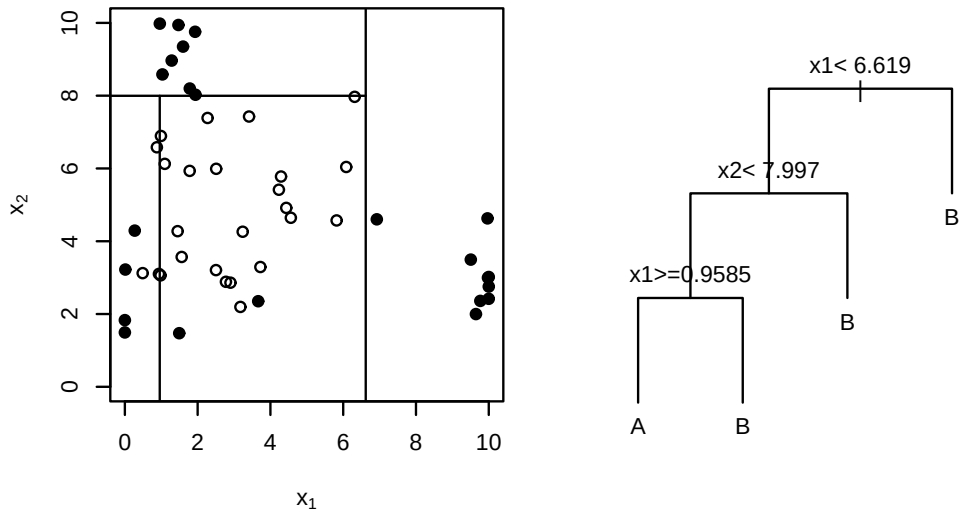

Abbildung 3.13: Partitionierung beim Klassifikationsbaum.

schwarze Punkte dargestellt sind) und eine hieraus hergeleitete Partitionierung (links).

Die Klassifikationsregeln lassen sich als ein Entscheidungsbaum darstellen (rechts).
Die Partitionierung entsteht dadurch, daß Schritt für Schritt jeweils eine Partition entlang einer der Variablen in zwei Klassen aufgeteilt wird. Diese „Spaltpunkte“, an denen die Partitionen getrennt werden, werden ermittelt, indem in jedem Schritt der „beste“ Spaltpunkt über alle Variablen bestimmt wird. Es kommt in jedem Schritt jeweils nur eine begrenzte Anzahl neuer Partitionierungen in Frage, da das Partitionieren nur *zwischen* zwei beobachteten Realisierungen einer Variablen sinnvoll ist, und das für alle d Variablen. Für diese potentiellen Spaltpunkte wird jeweils eine Maßzahl berechnet, die die Vermischung der verbleibenden Partitionen beschreibt und es wird diejenige Spaltung gewählt, die zu den „reinsten“ Partitionen führt.
Vorteil der Klassifikationsbäume ist wiederum, daß keine Verteilungssannahmen gemacht werden müssen, die abgeleiteten Klassifikationsregeln sind einfach und leicht interpretierbar, und die Variablenselektion (dazu mehr in Abschnitt 3.9) erübrigt sich.
Nachteil ist, daß die Klassengrenzen nur entlang der Hauptachsen gezogen werden — das verkompliziert die Klassentrennung bei Klassen, die sich durch eine Gerade beliebiger Orientierung möglicherweise einfach trennen ließen, durch einen Klassifikationsbaum allerdings nur durch viele Einzelschritte; dies kann insbesondere bei korrelierten Variablen zum Problem werden.
Um eine Überanpassung (*Overfitting*, siehe auch Abschnitt 3.9, Seite 44) zu verhindern, muß die Feinheit der Aufteilung reguliert werden. Überanpassung tritt auf, wenn die Partitionierung zu genau an die Trainingsdaten angepaßt wird, so daß diese dann nicht mehr repräsentativ für deren Grundgesamtheit ist. Bei den Daten in Abbildung 3.13 ließe sich durch weitere Partitionierung ein Baum konstruieren, der die Daten *perfekt*, also ohne Fehler, klassifiziert. Dieser würde dann aber wahrscheinlich die Klassen *neuer Daten* aus derselben Grundgesamtheit schlechter vorhersagen.
Ausgehend von der (vermeintlich) perfekten Partitionierung wird der Klassifikationsbaum daher „zurückgeschnitten“ (*Pruning*), dies geschieht in der verwandten Implementation anhand der (durch Kreuzvalidierung) geschätzten Fehlerrate (Venables und Ripley, 2002).

3.7 k-Nearest-Neighbour

Die Klassifikationsregel beim k-Nearest-Neighbour ist sehr einfach, es muß nur der Parameter $k \in \mathbb{N}$ festgelegt werden. Beim 1-Nearest-Neighbour ($k = 1$) wird für eine neue, zu klassifizierende Beobachtung die *ähnlichste* Beobachtung aus den Trainingsdaten gesucht; diejenige, die den geringsten euklidischen Abstand zu der neuen Beobachtung hat, das ist dann der „nächste Nachbar“. Deren Klasse wird festgestellt, und der neuen Beobachtung wird dann dieselbe Klasse zugeordnet. Für $k > 1$ werden die k nächstliegenden Beobachtungen bestimmt und festgestellt, welcher Klasse die Mehrheit dieser k nächsten Nachbarn angehört, die Nachbarn dürfen sozusagen „abstimmen“, und bei Stimmengleichheit wird zufällig zugewiesen.
Abbildung 3.14 zeigt die Diskriminanzfunktion beim k-Nearest-Neighbour für ein paar Beispieldaten und variierendes k. Es gibt zwei Klassen „A“ und „B“; die Klassengrenzen verlaufen sehr unregelmäßig.

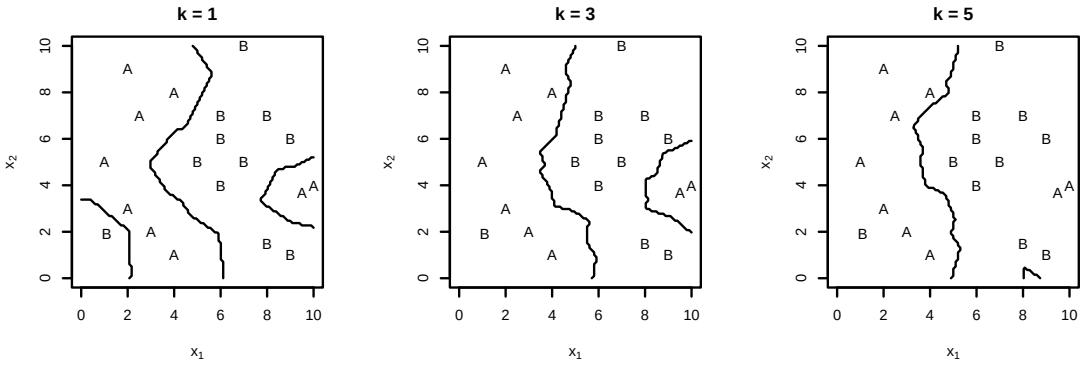

Abbildung 3.14: Diskriminanzfunktion beim k-Nearest-Neighbour für verschiedene Werte von k.

Vorteil dieser Methode ist, daß sie ohne Modellannahmen (Normalverteilung oder ähnliches) auskommt und daher auch die Form der Klassengrenzen nicht restringiert ist, die Klassen müssen nicht einmal zusammenhängend sein. Es muß lediglich der Wert von k festgelegt werden.
Ein Nachteil ist, daß für diese Entscheidungsregel jeweils die gesamten Daten betrachtet werden müssen (es muss zu jedem einzelnen Datenpunkt die Entfernung bestimmt werden). Bei anderen Verfahren beschränkt sich die zur Klassifikation notwendige In-

formation meistens auf wenige Parameter (wie z.B. Mittel und Varianz bei der LDA). Entscheidend für das Verfahren ist allerdings noch die Skalierung der Variablen: Da für die Klassifikation euklidische Abstände bestimmt werden, müssen die Skalen der Variablen so aufeinander abgestimmt werden, daß die Metrik nicht von einer Variablen dominiert wird. In der Regel werden die Variablen normiert, indem der Mittelwert subtrahiert und anschließend durch die Standardabweichung dividiert wird (Hastie u. a., 2001).

3.8 Poisson-Modell

Ein weiterer Ansatz zur Klassifikation soll über den direkten Vergleich der Besetzungszahlen laufen, wobei auch das Wissen um die Tonfrequenz ausgenutzt werden soll. Dabei wird ein neuer Klang mit denjenigen Trainingsdaten verglichen, die eine ähnliche Frequenz wie der neue Klang haben. Abweichungen in den Besetzungszahlen werden im Verhältnis zu ihrer Größenordnung betrachtet.
Hierzu werden die Besetzungszahlen als Zufallsvariablen modelliert: Es wird zunächst angenommen, daß die Gesamtsumme N der Signalflanken in einem bestimmten Zeitraum einer Poissonverteilung folgt. Die Poissonverteilung ist eine diskrete Verteilung mit einem Parameter (die „Rate“ $\lambda \in \mathbb{R}^+$), und ihre Dichte ist (für $x \in \mathbb{N}_0$) gegeben durch

$$f_\lambda(x) = \frac{1}{x!}\lambda^x \exp(-\lambda) \tag{3.9}$$

Die Poisson-Verteilung wird aufgrund ihrer Eigenschaften oft zur Modellierung von Zählvariablen (Variablen, die das Auftreten bestimmter Ereignisse innerhalb eines Zeitraumes zählen) benutzt; einige dieser Eigenschaften sind (Mood u. a., 1974):

- $X \sim \text{Poisson}(\lambda) \quad \Rightarrow \quad \text{E}(X) = \text{Var}(X) = \lambda$
- Die Poissonverteilung ist die Grenzverteilung der Hypergeometrischen und der Binomialverteilung
- Für $\lambda \to \infty$ konvergiert die Poissonverteilung gegen eine Normalverteilung mit Parametern $\mu = \sigma^2 = \lambda$

- Es gibt eine eine „Rate“ ν, so daß
  P(Ereignis in Intervall der Länge h) $= \nu h + o(h)$
- Die Wartezeit zwischen zwei Ereignissen ist exponentialverteilt

Angewandt auf die Signalflanken als Ereignisse ist der letzte Punkt offensichtlich nicht erfüllt, denn die Wartezeiten zwischen den Signalflanken folgen (abhängig vom Instrument) oft sogar multimodalen Verteilungen, wie man z.B. in Abbildung 3.4 sieht. Sehr anschaulich ist allerdings die Vorstellung von einer bestimmten Rate, mit der die Signalflanken auftreten; und außerdem die zum Erwartungswert proportionale Varianz.
Nun sollen die Besetzungszahlen der verschiedenen Amplituden (nach Abschnitt 3.2.1) modelliert werden. Es wird angenommen, daß die Gesamtzahl von Signalflanken N in einer Zeitspanne der Länge t poissonverteilt ist mit einer gewissen Rate (λt), und die einzelnen Signalflanken jeweils mit einer gewissen Wahrscheinlichkeit $(p_1, \ldots, p_{32}; \sum p_i = 1)$ einer der 32 Amplituden-Klassen angehören, d.h.:

$$N \sim \text{Poisson}(\lambda t), \tag{3.10}$$

$$n_1, \ldots, n_{32} | N \sim \text{Multinomial}(N, p_1, \ldots, p_{32}) \tag{3.11}$$

Ein vermeintlich einfacheres Modell wäre, daß die Besetzungszahlen n_i der einzelnen Amplituden-Klassen unabhängig poissonverteilt sind mit Rate $\lambda t \cdot p_i$:

$$n_i \sim \text{Poisson}(\lambda t \cdot p_i) \quad i = 1, \ldots, 32 \tag{3.12}$$

Beide Modelle sind allerdings äquivalent (siehe Seite 72 ff. im Anhang), d.h. sie implizieren die gleichen Verteilungen und damit auch die gleichen Klassifikationsregeln.
Diese Modellierung soll nun dazu dienen, in einer Art modifiziertem k-Nearest-Neighbour-Verfahren die Metrik darzustellen:
Ein neu zu klassifizierendes Histogramm wird mit einer Teilmenge der Trainingsdaten verglichen, und zwar nur mit den Klängen, die in einem gewissen Frequenzband ($\pm m$ Halbtöne) um den neuen Klang herum liegen. Für diese Auswahl von Klängen werden jeweils Schätzer für obiges Modell bestimmt:

$$\hat{\lambda} := \frac{N}{t} \quad \text{und} \tag{3.13}$$

$$\hat{p}_i := \frac{n_i + \frac{1}{2}}{N + 16} \quad \text{für } i = 1, \ldots, 32 \tag{3.14}$$

Bei der Schätzung der p_i (3.14) wird jeweils eine „halbe Beobachtung“ hinzugezählt, um zu verhindern, daß einer der Schätzer gleich Null wird. Dies kann man auch als Einbindung von a-priori-Information interpretieren: Für $N = 0$, also ohne eine Beobachtung, sind die Signalflanken gleichverteilt auf die 32 Klassen. Für große N wird der Einfluß dieser Korrektur immer geringer.
Für die neue Beobachtung x wird nun für jedes dieser geschätzten Modelle die Likelihood $L(x|\lambda, p_1, \ldots, p_{32}) = \mathrm{P}_{\lambda,p_1,\ldots,p_{32}}(X = x)$ berechnet und dasjenige bestimmt, unter dem diese maximal ist. Dann wird die neue Beobachtung der entsprechenden Klasse zugeordnet (entsprechend der Maximum-Likelihood-Entscheidungsregel wie bei den linearen Verfahren, Abschnitt 3.4). Der Unterschied zum k-Nearest-Neighbour ist, daß die Likelihood nicht die *Distanz* mißt, sondern (antiproportional) ein Maß für die *Ähnlichkeit* ist; insbesondere ist sie keine Metrik.
Abbildung 3.15 illustriert die Likelihoods zweier Modelle: Die beiden Koordinatenach-

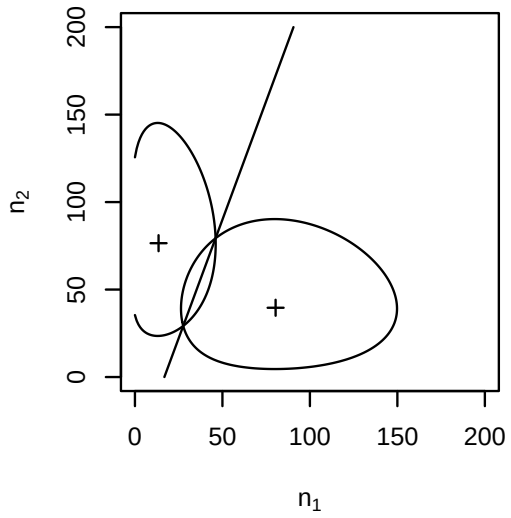

Abbildung 3.15: Diskriminanzfunktion im Poisson-Modell.

sen entsprechen zwei Besetzungszahlen n_1 und n_2 und die beiden Kreuze bezeichnen jeweils den Erwartungswert $(\lambda p_1, \lambda p_2)$ nach einem der Modelle. Für beide Likelihoodfunktionen ist jeweils eine Höhenlinie eingezeichnet, und zusätzlich die Gerade, entlang derer beide gleich sind (die Diskriminanzfunktion also), analog zu den Graphen aus Abschnitt 3.4. Man beachte, daß der Definitionsbereich hier diskret ist ($\mathbb{N}_0^2$).

Dieses Modell ähnelt in seinen Annahmen sehr dem Modell, das auch dem χ^2-Test zugrundeliegt — dort werden auch Häufigkeiten bei klassierten Daten verglichen. Mehr dazu im Anhang, Seite 74.

3.9 Variablenselektion

Bei Klassifikationsverfahren (wie auch bei Regressionsverfahren) steht oft eine große Zahl von Variablen zur Auswahl, die *potentiell* zur Klassifikation nützlich sind. Die Betrachtung *zu vieler* Variablen, also von Variablen, die keine oder wenig zusätzliche Information liefern, führt bei manchen Verfahren zu Problemen wie verzerrten Schätzern und Überanpassung.

Überanpassung (*Overfitting*) bedeutet, daß sich das Klassifikationsverfahren zu sehr an Eigenheiten der Trainingsauswahl anpaßt, und damit nicht mehr repräsentativ für die Grundgesamtheit ist.

Die Anzahl der Variablen muß also auf eine notwendige Auswahl reduziert werden. Das Problem stellt sich nicht bei den Entscheidungsbäumen, da hier die entscheidenden Variablen „automatisch“ ausgewählt werden; hier wird stattdesen das Zurückstutzen (*Pruning*, siehe Seite 38) notwendig. Bei den restlichen hier betrachteten Verfahren (Diskriminanzanalyse, k-Nearest-Neighbour und Support Vector Machines) ist die Reduktion jedoch notwendig.

Theoretisch wäre es am besten, alle möglichen Variablenkombinationen durchzuprobieren, um dann die „beste“ auszuwählen, nur scheitert dieses Ansinnen an der meist astronomischen Zahl von möglichen Zusammenstellungen der Variablen (bei d Variablen gibt es 2^d mögliche Teilmengen).

In der Regel wird deshalb eine Schrittweise Auswahl (*Stepwise Selection*) von Variablen durchgeführt: Es werden zunächst alle Modelle ausprobiert, die nur eine Variable berücksichtigen, und anhand eines bestimmten Kriteriums wird das beste Modell bestimmt. In den folgenden Schritten wird von den verbleibenden Variablen probehalber jeweils eine dazugenommen, und diejenige wird schließlich in das Modell übernommen, die die größte Verbesserung mit sich bringt (Fahrmeir u. a., 1996).

Um ein einheitliches Kriterium für alle Verfahren zu haben, wird hier der geschätzte Vorhersagefehler benutzt. Andere mögliche Kriterien stützen sich oft auf Parameter der

Verfahren und sind dadurch nicht zwischen den Verfahren übertragbar. Zur Schätzung des Vorhersagefehlers wird hier die *10-fache Kreuzvalidierung* genutzt, d.h. der gesamte Datensatz wird zufällig in 10 gleichgroße Teile zerlegt und die Klassenzugehörigkeiten für jeden einzelnen Teil (als Teststichprobe) werden anhand der verbleibenden 9 Teile (als Trainingsstichprobe) geschätzt; anschließend werden die 10 resultierenden Fehlerraten gemittelt.
In diesem Falle wird jeweils anfangs von dem Modell ausgegangen, das nur die Tonfrequenz als Variable enthält. Dann werden weitere hinzugenommen, bis insgesamt 20 Variablen ausgewählt sind. Aus der so entstandenen Folge von 20 größer werdenden Modellen soll dann anhand der Fehlerrate ein sinnvoller Modellumfang bestimmt werden.

3.10 Benutzte Software

Für alle Berechnungen wurde R-1.4.0 benutzt. R ist eine Programmiersprache und -umgebung für Datenanalyse und Grafik. Sie kann im Internet unter `http://www.r-project.org` kostenlos (unter der „General Public License“ (GPL): `http://www.gnu.org/copyleft/gpl.html`) heruntergeladen werden (Ihaka und Gentleman, 1996). Zum Einlesen von Klangdateien wurde das R-Package `sound` verwandt, und für die Clusteranalyse das Package `cluster`. Für die Klassifikation wurden die Packages

Tabelle 3.3: Verwandte R-Packages.

Verfahren	**Package**	**Funktion**
LDA	`MASS`	`lda`
Entscheidungsbaum	`rpart`	`rpart`
Support Vector Machine	`e1071`	`svm`
k-Nearest-Neighbour	`class`	`knn`

aus Tabelle 3.3 benutzt, die übrigen Klassifikationsverfahren (QDA, Naive Bayes, RDA) sowie Nelder-Mead-Algorithmus zur Parameterbestimmung bei der RDA und die Variablenselektion sind selbstprogrammiert. Bis auf die letztere waren die Programme jedoch größtenteils schon vorhanden (Theis u. a., 2002). Das Package zur RDA kann unter der URL `http://www.statistik.uni-dortmund.de/de/content/einrichtungen/lehrstuehle/personen/wtheis_de.html` heruntergeladen werden.

4 Ergebnisse

4.1 Die Fehlerraten

Die Fehlerraten in den folgenden Abschnitten sind jeweils durch Simulation ermittelt. Dafür wird jeweils der komplette Datensatz zufällig in zwei Teile aufgeteilt; der größere Teil enthält dabei jeweils $\frac{3}{4}$ der Klänge jedes Instrumentes, der kleinere $\frac{1}{4}$. Mit dem größeren Teil wird dann das jeweilige Modell angepaßt, um damit die Klassenzugehörigkeiten für den kleineren Teil vorherzusagen. Für einen solchen Simulationslauf werden dann die einzelnen Fehlerraten aufgeschlüsselt nach den wahren Instrumentenklassen berechnet und anschließend gemittelt. Der letzte Schritt ist notwendig, da die einzelnen Instrumentenklassen in unterschiedlichem Umfang im Datensatz vertreten sind und damit ansonsten zu einer unterschiedlichen Gewichtung der Intrumente führen würden. Die so bestimmte Schätzung der Gesamtfehlerrate beruht dann auf der Annahme, daß alle Instrumentenklassen gleichhäufig vorkommen.
Die angegebenen geschätzten Fehlerraten sind dann Mittelwerte über viele solcher Simulationsläufe. Die in den folgenden Tabellen und Graphen angegebenen Fehlerraten sind jeweils über 100 wiederholte Klassifizierungen gemittelt, mit Ausnahme der Fehlerraten in der Variablenselektion (s.d., Seite 44), die durch Kreuzvalidierung ermittelt sind (betrifft Abbildung 4.1 und Tabelle A.5).

Die schlechtestmögliche Fehlerrate (entsprechend der Anzahl Klassen) ist $\frac{24}{25} = 96\%$, das ist die Fehlerrate, die man durch pures Raten oder auch durch sture Klassifikation als immer dasselbe Instrument erreichen würde.
Fehlerraten, die der Mensch erreicht, wurden in anderen Experimenten untersucht und

sind z.B. bei Bruderer (2003) zusammengestellt: In zwei möglicherweise vergleichbaren Szenarien (27 Instrumente) wurden Fehlerraten von 54% bzw. 44% festgestellt. In der ersteren Untersuchung lag sogar derselbe Datensatz wie in dieser Arbeit zugrunde; allerdings ist nicht klar, um welche Auswahl von Instrumenten es sich dabei handelte. In derselben Arbeit sind auch Fehlerraten zusammengetragen, die bisher durch automatische Klassifikation erreicht wurden; hier wurden in der Regel Spekrum- und Hüllkurvencharakteristika verwendet. Bei vergleichbarer (eher schwierigerer) Ausgangssituation, was Anzahl der Klassen und Datenumfang betrifft, sind hier Fehlerraten von 19 und 7.2% angegeben. Allerdings ist zu bedenken, daß dort jeweils der komplette Klang zur Verfügung stand, während in dieser Arbeit nur anhand der ersten 0.77 Sekunden klasssifiziert wird.

4.2 Erster Ansatz: Besetzungszahlen

Bei diesem Ansatz wurden neben den Besetzungszahlen zwei weitere Variablen berücksichtigt: die Tonfrequenz und die Dauer des Tones. Die Tondauer ist insbesondere notwendig, um beim Poisson-Modell jeweils die Anzahl von Signalflanken pro Zeiteinheit (pro Sekunde) zu bestimmen. Bei den übrigen Verfahren geht die Dauer als weitere diskriminierende Variable ein.

Die ersten 3 Amplitudenklassen, die *nie* besetzt waren (also in jedem Falle konstant 0 sind), wurden auch nicht als Variablen berücksichtigt, da das auch zu Problemen bei der Diskriminanzanalyse geführt hätte. Insgesamt sind es damit $2 + 32 - 3 = 31$ Variablen.

Nicht verwendet werden konnten diejenigen Klänge, die nur eine oder weniger Signalflanken ausgelöst hatten, da hier zumindest die Dauer des Klanges nicht definiert wäre. Nach Entfernen dieser 44 Beobachtungen bleiben 1943 übrig.

Demnach ergeben sich die Fehlerraten in Tabelle 4.1. Am besten schneidet das k-Nearest-Neighbour (mit $k = 1$) mit einer Fehlerrate von 55.8% ab. Im Anhang ist die Fehlerrate für dieses Verfahren in einer Fehlklassifikationsmatrix detailliert aufgeschlüsselt (Tabelle A.4, Seite 67).

Bemerkenswert ist, daß die Fehlerraten beim Poisson-Modell und beim Naive Bayes praktisch gleich sind. Das kann Zufall sein, oder es könnte an ihrer Gemeinsamkeit

Tabelle 4.1: Fehlerraten beim reinen Besetzungszahlenvergleich.

Verfahren	**Fehlerrate (%)**
LDA	67.0
QDA	–
Naive Bayes	69.7
RDA	62.0
SVM	73.6
Entscheidungsbaum	80.9
k-NN ($k = 1$)	55.8
Poisson-Modell ($m = 2$)	69.6

liegen: Bei beiden Verfahren werden die Besetzungszahlen als unabhängig voneinander modelliert. Unterschiede sind jedoch, daß beim Naive Bayes die Varianzen und Erwartungswerte separat geschätzt werden, während sie beim Poisson-Modell als identisch angenommen werden; außerdem werden Tonfrequenz und -dauer auf unterschiedliche Weise eingebunden.

QDA funktioniert hier nicht, da für einige Klassen die geschätzten Kovarianzen nicht invertierbar sind. Die RDA wurde mit den Parametern $\lambda = 0.1$ und $\gamma = 0$ durchgeführt. Minimiert man die Fehlerrate über beide Parameter, so wird sehr deutlich, daß $\gamma = 0$ sein muß. Für λ sind die Unterschiede nicht so eindeutig, aber bestimmt man wiederholt die Fehlerrate für verschiedene Werte von λ bei $\gamma = 0$, so zeigt sich, daß der optimale Wert tatsächlich etwa bei 0.1 liegt (siehe dazu auch Abbildung A.3 auf Seite 68 im Anhang).

Entsprechend sind die Parameter $k = 1$ für das k-NN und $m = 2$ beim Poisson-Modell diejenigen, die zur kleinsten Fehlerrate führten.

Bei der Support Vector Machine wurde die „*radial basis*“-Kernfunktion mit den Parametern $C = 100$ und $\gamma = \frac{1}{d}$ benutzt.

Die Tatsache, daß die RDA wesentlich besser als das Naive Bayes funktioniert, legt nahe, daß die Besetzungszahlen für die verschiedenen Klassen nicht unabhängig sind.

4.3 Zweiter Ansatz: Hough-Charakteristika

4.3.1 Variablenselektion

Für LDA, Naive Bayes, RDA, Support Vector Machine und 1-Nearest-Neighbour wurde nun die Variablenselektion durchgeführt. Abbildung 4.1 zeigt die dabei erreichten Fehlerraten in Abhängigkeit von der Anzahl der Variablen im Modell. *Welche* Va-

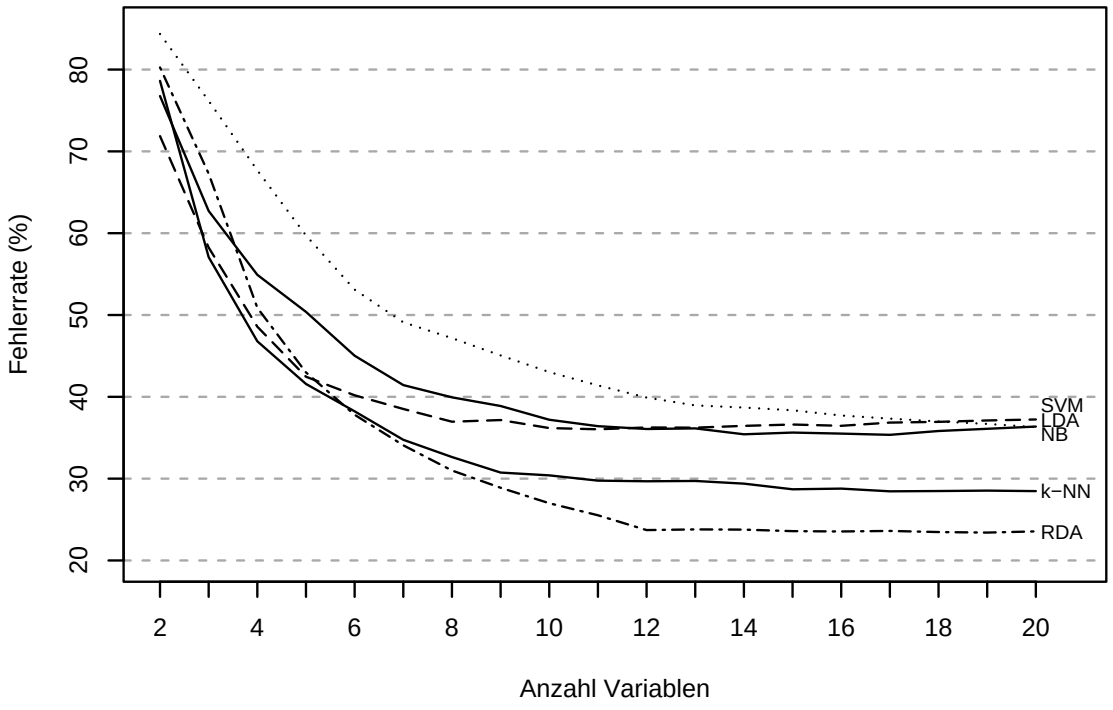

Abbildung 4.1: Fehlerraten in der Variablenselektion.

riablen jeweils ausgewählt wurden, ist von Verfahren zu Verfahren verschieden; die genauen Auswahlen sind in Tabelle A.5, Seite 69 im Anhang ausführlich dargestellt. Die beste Fehlerrate insgesamt wird mit der RDA erreicht, die ab 6 Variablen die übrigen Verfahren hinter sich läßt; ab 12 Variablen wird die Fehlerrate durch Hinzunahme weiterer Variablen dann allerdings nicht mehr wesentlich verbessert.
Bemerkenswert ist auch der Verlauf der Fehlerrate bei der Support Vector Machine (gestrichelte Linie): sie funktioniert zunächst (mit 2 Variablen) am besten, erreicht

dann relativ schnell ihr Optimum bei 8–11 Variablen und wird anschließend wieder schlechter.
Am schlechtesten schneidet zunächst die LDA ab (punktierte Linie), allerdings erreicht sie mit weiteren Variablen dann auch das Niveau von Support Vector Machine und Naive Bayes.
Auf die RDA als bestes Verfahren wird im Folgenden noch näher eingegangen.

Die RDA wurde (wie schon bei den Besetzungszahlen) wiederum jeweils mit Parametern ($\lambda = 0.1, \gamma = 0$) durchgeführt. Die Optimierung der Fehlerrate über beide Parameter zeigte auch hier, daß γ in jedem Falle $= 0$ sein muß; Abbildung 4.2 zeigt die Fehlerrate in Abhängigkeit von λ (für $\gamma = 0$ und 12 Variablen). Es ist $\lambda \in \{0.02, 0.05, 0.1, 0.15, \ldots, 1\}$. Das Optimum liegt wiederum tatsächlich bei $\lambda = 0.1$.

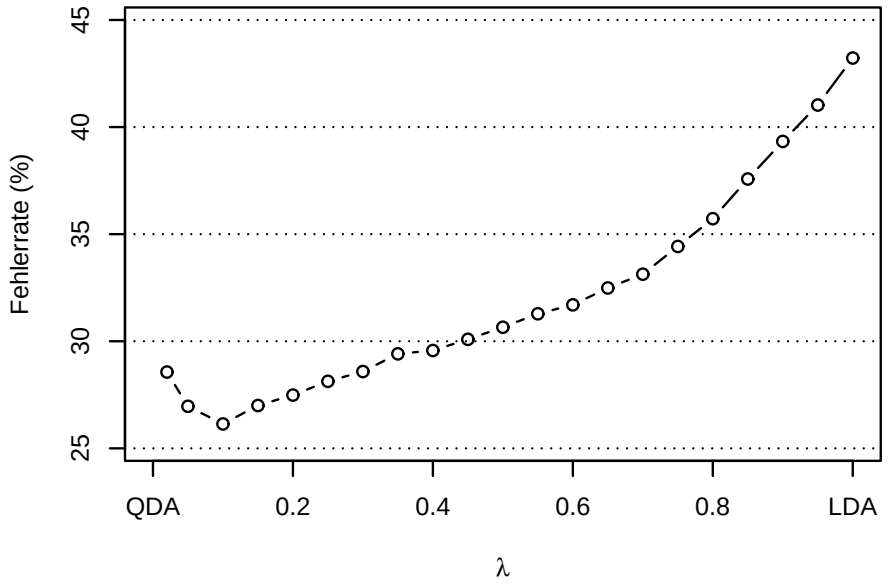

Abbildung 4.2: Fehlerrate bei RDA auf den Hough-Charakteristika in Abhängigkeit von λ (wobei $\gamma = 0$ und $\lambda \in [0.02, 1]$). Das Optimum liegt bei $\lambda = 0.1$.

Die ersten 12 selektierten Variablen sind:

i. logarithmierte Tonfrequenz (1),

ii. Differenz von 5%- und 95%-Quantil der δ's (2)

iii. logarithmierter Zeitpunkt der ersten (3) und letzten (8) Signalflanke

iv. logarithmierte Kurtosis (Wölbung) der Amplitudenverteilung (4)

v. Median (9) und mittlere Abweichung vom Median (5) der Amplituden

vi. Mittelwertverschiebung der Amplituden (6)

vii. Mittel (7) und Interquartilsabstand (12) der Frequenz

viii. Rate (10) der Signalflanken (pro Sekunde)

ix. Kendall's τ (11) für Korrelation zwischen Amplitude und Frequenz

Die Zahl in Klammern gibt jeweils an, an wievielter Stelle die entsprechende Variable in das Modell aufgenommen wurde.
Die Variablen sind auch sämtlich gut interpretierbar; demnach sind die zur Klassifikation entscheidenden Merkmale eines Klanges:

i: die Tonhöhe

ii: Art der zeitlichen Aufeinanderfolge der Signalflanken (Streuung der δ's)

iii: Wartezeit vom Tonbeginn bis zur ersten Signalflanke sowie Tondauer

iv, v: Lage, Streuung und Form der Amplitudenverteilung

vi: Veränderung der mittleren Amplitude über die Zeit

vii: Lage und Streuung der Frequenzenverteilung

viii: Rate der Signalflanken (pro Sekunde)

ix: Korrelation von Amplitude und Frequenz

Es tauchen mehrfach Merkmale auf, die Lage oder Streuung von Variablen bezeichnen; allerdings kommen dabei verschiedene Maße zum Einsatz, wie Mittel, Median und Interquartilsabstand. Unter bestimmten Verteilungsannahmen (etwa Unabhängigkeit und Normalverteilung) sind die in gewissem Sinne *besten* Maßzahlen (die sogenannten

suffizienten Statistiken) hierfür jedoch das arithmetische Mittel und die Standardabweichung. Ersetzt man nun an den betreffenden Stellen die Lagemaße jeweils durch das arithmetische Mittel und die Streuungsmaße durch die Standardabweichung, so ergibt sich keine wesentlich verschiedene (tendenziell sogar eine bessere) Fehlerrate als bei den obigen Variablen.
Die 4 eingewechselten Variablen sind dann: Standardabweichung der δ's (ii), Mittel und Standardabweichung der Amplituden (v) und die Standardabweichung der Frequenzen (vii). Im Folgenden sind arithmetisches Mittel und Standardabweichung dann die einzigen Lage- und Streuungsmaße im Modell.

Um den Beitrag der einzelnen Variablen zur Trennung der Klassen abzuschätzen, kann man einerseits die Reihenfolge betrachten, mit der die Variablen bei der Vorwärtsselektion in das Modell aufgenommen wurden (siehe die Aufzählung auf Seite 52). Einen weiteren Anhaltspunkt bietet die Verschlechterung der Fehlerrate, die jeweils eintritt, wenn man eine der Variablen aus dem Modell herausnimmt. In Abbildung 4.3 sind einmal diese Differenzen für alle 12 Variablen dargestellt.

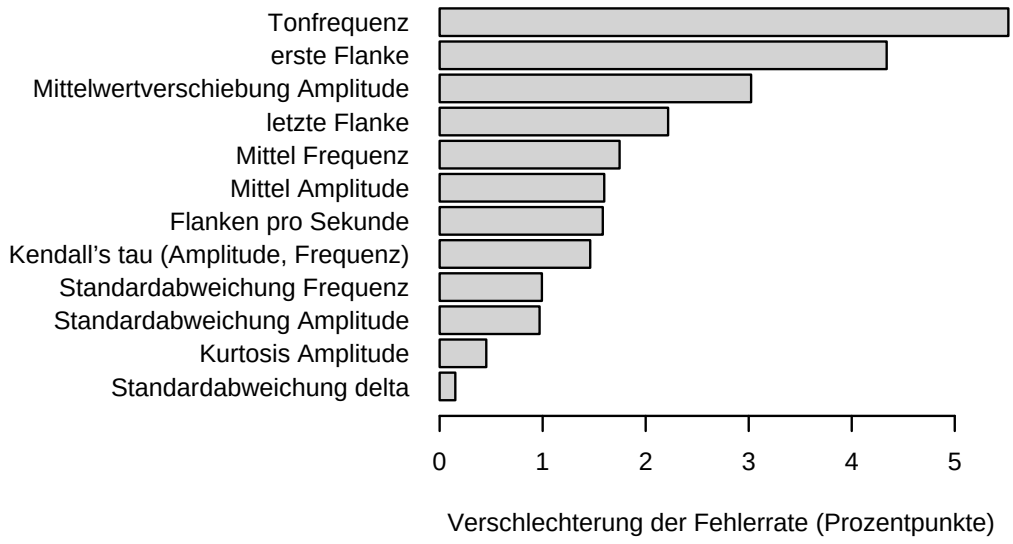

Abbildung 4.3: Differenzen der Fehlerraten beim Weglassen einzelner Variablen.

Die „wichtigste“ Variable wäre demnach die Tonfrequenz. Die Variable, die bei der

Selektion als nächste ausgewählt wurde, landet hier allerdings auf dem letzten Platz; die „Wichtigkeit“ einzelner Variablen ist also nur schwer zu bemessen. Allerdings ist zu überlegen, ob die beiden letzten Variablen, die die Fehlerrate am wenigsten verbessern, nicht weggelassen werden können.

4.3.2 Fehlerraten

Für die verschiedenen Verfahren ergeben sich nun die Fehlerraten nach Tabelle 4.2. Variablenanzahl und -auswahl sind jeweils im Hinblick auf die geschätzten Fehlerraten in der Variablenselektion festgelegt (Abbildung 4.1), bzw. bei der RDA auch nach Abbildung 4.3.

Tabelle 4.2: Fehlerraten bei Klassifikation anhand der Hough-Charakteristika.

Verfahren	**Variablen**	**Fehlerrate (%)**
LDA	20	37.9
QDA	–	–
Naive Bayes	14	36.4
RDA	10	26.6
RDA	11	26.1
RDA	12	26.0
SVM	8	39.2
SVM	11	38.3
Entscheidungsbaum	10.7*	72.3
k-NN ($k = 1$)	11	30.9
k-NN ($k = 1$)	15	31.5

* Durchschnitt

Die geringste Fehlerrate wird mit der RDA mit 12 Variablen erreicht. Nur unwesentlich schlechter ist die Fehlerrate, wenn man sich auf 11 Variablen beschränkt und die „unwichtigste “ (gemäß Abbildung 4.3) wegläßt. So erreicht man eine Fehlerrate von knapp über 26%. Das zweitbeste Verfahren ist das k-Nearest-Neighbour ($k = 1$) mit ebenfalls 11 Variablen und einer Fehlerrate von 31%. Am schlechtesten funktioniert der Klassifikationsbaum, der bei durchschnittlich 10.7 Variablen Fehlerraten um 72% erreicht. Das liegt wahrscheinlich an der Korreliertheit der Variablen innerhalb der Klassen,

die eine Trennung durch Ebenen senkrecht zu den Hauptachsen erschwert, und weiterhin an der relativ großen Anzahl der Klassen, die eine nennenswerte Erhöhung der Reinheit der Partitionen durch einzelne Trennungen verhindert. Die übrigen Verfahren erreichen 36–40% Fehlerrate, wobei die LDA dafür bei weitem die meisten Variablen benötigt. Die QDA kann nicht angewendet werden, da die Varianzschätzung hier wiederum zu nicht invertierbaren Koovarianzmatrizen führt.

Die Fehlerraten für die einzelnen Instrumente sind der Fehlklassifikationsmatrix (Tabelle 4.3) zu entnehmen: Die Zeilen geben jeweils an, wie oft das betreffende Instrument

Tabelle 4.3: Fehlklassifikationsmatrix für RDA mit 11 Variablen.

%	ba	be	ce	cl	cr	eb	eg	ed	ef	fl	fr	gk	ma	ob	pi	sx	sy	tb	tp	tp	tu	vb	vp	vi	xy	Σ
bassoon	78	0	2	1	0	1	0	0	0	0	0	0	0	1	0	0	2	9	0	0	6	0	0	0	0	22
bells	0	95	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	5
cello	6	0	72	3	0	0	4	3	0	0	0	0	0	1	0	4	0	0	2	0	5	0	0	0	0	28
clarinet	2	0	3	52	0	0	0	8	0	2	1	0	0	7	0	10	0	3	7	0	1	1	0	3	0	48
crota	0	0	0	0	97	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	3
elecbass	0	0	0	0	0	80	7	0	4	0	0	0	2	0	4	0	0	0	0	0	2	1	0	0	0	20
elecguitar	1	6	8	1	0	12	53	1	2	0	1	0	0	0	4	1	0	1	0	0	1	6	0	1	1	47
elecguitar-dist	0	0	0	1	0	3	0	95	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5
elecguitar-fh	0	0	0	0	0	12	3	0	73	0	0	0	0	0	3	0	0	0	0	0	0	1	8	0	0	27
flute	1	0	1	1	0	0	0	0	0	69	0	0	0	3	0	2	0	3	3	0	8	2	0	6	0	31
frenchhorn	0	0	0	2	0	0	0	0	0	0	90	0	0	4	0	2	0	0	2	0	0	0	0	0	0	10
gks	0	0	0	0	11	0	0	0	0	0	0	83	0	0	1	0	0	0	2	0	0	0	0	0	2	17
marimba	0	0	0	0	0	8	0	0	0	0	0	0	61	0	1	0	0	0	0	0	0	0	3	0	26	39
oboe-enghorn	0	0	0	9	0	0	0	0	0	5	2	0	0	70	0	2	0	2	7	1	0	0	0	2	0	30
piano	6	1	1	0	0	7	3	0	1	0	0	2	10	0	55	0	0	0	0	0	0	4	2	0	8	45
saxophone	8	0	10	11	0	0	0	0	0	0	7	0	0	6	0	46	0	3	6	0	0	2	0	0	0	54
synthbass	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	98	0	0	0	0	0	0	0	0	2
trombone	4	0	0	7	0	0	0	1	0	3	0	0	0	3	0	0	0	73	7	0	0	0	0	1	0	27
trumpet	0	0	1	2	0	0	0	0	0	4	5	0	0	8	0	2	0	7	68	0	0	3	0	0	0	32
trumpet-csto	0	0	0	3	0	0	0	3	0	0	0	0	0	3	0	0	0	0	0	90	0	0	0	0	0	10
tuba	3	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	95	0	0	0	0	5
vibraphone	0	2	1	1	0	5	8	0	2	9	1	0	3	0	0	0	0	0	1	1	7	57	0	0	1	43
violin-piz	0	0	0	0	0	2	0	0	5	0	0	1	6	0	2	0	0	0	0	0	0	0	84	0	0	16
violin-viola	2	0	2	6	0	0	0	0	2	7	1	0	3	16	0	1	0	7	1	2	1	0	1	48	1	52
xylo	0	0	0	0	0	0	0	0	1	0	0	5	23	0	2	0	0	0	0	0	0	2	0	0	66	34

Gesamtfehlerrate: 26.1%

den verschiedenen Klassen zugeordnet wurde; Die letzte Spalte gibt die Fehlerrate für das Instrument an (jeweils in Prozent, gerundet).

4.4 Zur Center-Frequency

Abbildung 4.4 zeigt die mittleren Häufigkeiten der Amplituden über alle Klänge. Die Verteilung ist ungleichmäßig: die kleinen Amplituden-Klassen sind selten oder gar nicht besetzt, während die großen Klassen häufiger besetzt sind. Gleichzeitig trat das

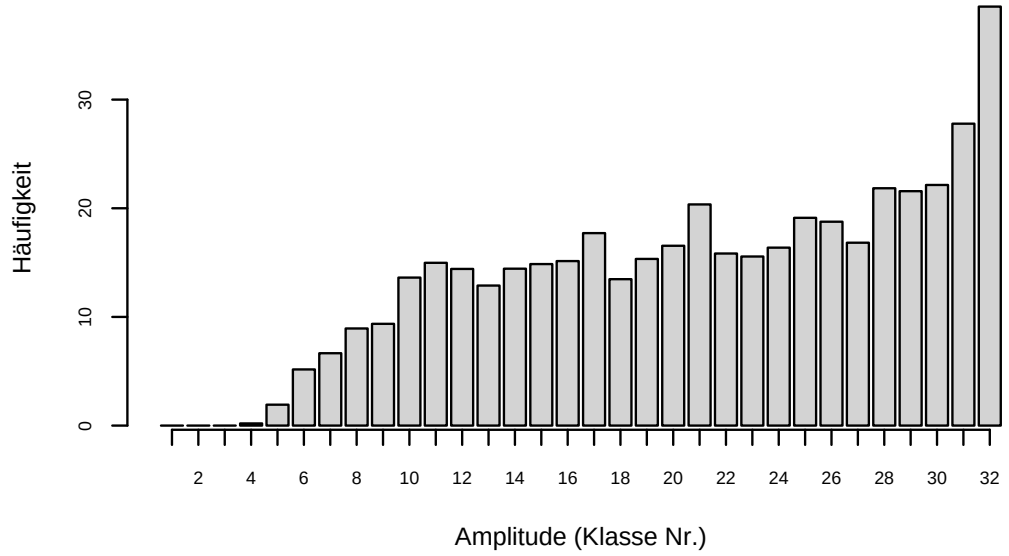

Abbildung 4.4: Die mittlere Häufigkeit der Amplituden über *alle* Klänge.
Große Amplituden-Klassen entsprechen kleinen Amplituden und umgekehrt.

Problem auf, daß bisweilen sehr wenige oder gar keine Signalflanken ausgelöst wurden, was wahrscheinlich oft darauf zurückzuführen ist, daß der jeweilige transformierte Klang eine zu kleine Amplitude hatte.
Beides ließe sich eventuell durch eine Änderung der Parametrisierung, nämlich eine Verringerung der Center-Frequency (f) beheben.
Bei einer kleineren Center-Frequency hätte die Referenz-Signalflanke eine größere Periode und wäre damit flacher. Dadurch wäre im Verhältnis zur jetzigen Parametrisierung jeweils eine größere Streckung in y-Achsenrichtung (durch den Amplitudenparameter A) nötig, womit dann die kleinen Amplitudenklassen (entsprechend den großen Amplituden) besetzt würden. Gleichzeitig wäre auch die Detektion flacherer

Signalflanken (tieferer und leiserer Klänge) als bisher möglich. Bei einer voraussichlich geringeren Konzentration der Signalflanken auf bestimmte Amplitudenklassen wäre auch der Informationsgehalt der transformierten Daten größer.

5 Zusammenfassung

Unter den untersuchten Ansätzen und Verfahren erziehlt man die besten Ergebnisse mit Hilfe der Regularisierten Diskriminanzanalyse (RDA), die man auf charakterisierende Variablen anwendet, die die Eigenheiten der Hough-transformierten Klänge beschreiben („Hough-Charakteristika“). Faßt man die vorliegenden Instrumente zu sinnvollen (= ähnlich klingenden) Klassen zusammen, so wird bei letztlich 25 Klassen und 11 betrachteten Merkmalen auf diesen Daten eine Fehlerrate von 26.1% erreicht. Die 11 Merkmale beschreiben dabei

- die Tonhöhe,
- die Wartezeit vom Tonbeginn bis zur ersten Signalflanke sowie Tondauer,
- die Rate der Signalflanken (pro Sekunde),
- Lage, Streuung und Form der Amplitudenverteilung,
- die Veränderung der mittleren Amplitude über die Zeit,
- Lage und Streuung der Frequenzenverteilung und
- Korrelation von Amplitude und Frequenz.

Das zur Klassifikation zugrundegelegte Modell unterstellt dabei, daß sich die verschiedenen Klassen (Instrumente) in den besagten Variablen durch ihre Mittelwerte und ihre Varianz- und Kovarianzstruktur unterscheiden. Die als optimal befundene Parameterkombination ($\lambda = 0.1, \gamma = 0$) stellt dabei einen Kompromiß zwischen Linearer und Quadratischer Diskriminanzanalyse (LDA und QDA) dar, der die verschiedenen Gruppenkovarianzschätzer durch den gemeinsamen (gepoolten) Kovarianzschätzer stabilisiert.

Weitere untersuchte Verfahren auf diesen Variablen waren Lineare und Quadratische Varianzanalyse, naive Bayes, Support Vector Machines, Klassifikationsbäume sowie k-Nearest-Neighbour.
Außerdem wurde versucht, die Klänge alleine anhand der Randverteilungen der Amplituden sowie Frequenz und Tondauer zu klassifizieren. Hier wurden wiederum die obigen Verfahren, jedoch ohne die Klassifikationsbäume, angewandt; außerdem wurde versucht, die Randverteilungen als poissonverteilte Variablen zu modellieren. Auf diesen Variablen war das k-Nearest-Neighbour mit 56% Fehlerrate das beste Verfahren.
Ein weiterer Ansatz, die Daten der einzelnen Klänge durch Clusteranalyse aufzubereiten erwies sich als nicht erfolgversprechend.

Die hier geschätzten Fehlerraten sind zunächst einmal repräsentativ für den vorliegenden Datensatz; bei Übertragung des Verfahrens auf „neue“ Daten, können also Abweichungen auftreten. Weiteren Aufschluß über die Fehlerraten gibt die Fehlklassifikationsmatrix auf Seite 55.

Eine weitere Verbesserung der Fehlerrate ist wahrscheinlich möglich, sobald der zeitliche Umfang der Hough-transformierten Daten über die 0.77 Sekunden hinaus ausgedehnt wird, auf die sich die hier bearbeiteten Daten ausschließlich bezogen. Neben besseren Schätzungen der Hough-Charakteristika aufgrund der größeren Datengrundlage erhielte auch das Merkmal „Zeitpunkt der letzten Signalflanke“ eine größere Bedeutung: nur etwa $\frac{1}{3}$ der Klänge endet vor 0.77 Sekunden, für die Übrigen nimmt diese Variable bisher also nahezu konstant den Wert 0.77 an.
Eine Änderung in der Parametrisierung der Transformation (eine Verringerung der Center-Frequency) würde möglicherweise eine weitere Verbesserung mit sich bringen.

Vom Menschen erreichte Fehlerraten liegen bei ähnlicher Klassenanzahl bei etwa 44%, und bei automatischer Klassifikation werden 19 bis 7.2% angegeben — wobei jeweils offen ist, inwieweit die Ausgangslage wirklich vergleichbar ist.
Zieht man in Betracht, daß hier nur die Daten der Hough-Transformation der jeweils ersten 0.77 Sekunden der Klänge zugrunde lagen, so erscheint die dabei erreichte Fehlerrate von 26.1% durchaus beachtlich.

A Tabellen und Abbildungen

Tabelle A.1: Der Datensatz.

Instrument	**interner Name**	**Anzahl Dateien**	**Tonumfang** *Noten*	**Tonumfang** *Frequenz (Hz)*
Alt-Flöte *Vibrato*	aflute-vib	30	g3–c6	196.0–1046.5
Fagott	bassoon	32	a♯1–f4	58.3– 349.2
Rohrglockenspiel	bells	20	c4–g5	261.6– 784.0
Baß-Flöte *'flutter-tongued'*	bflute-flu	16	c3–d6	196.0–1046.5
Baß-Flöte *Vibrato*	bflute-vib	26	c3–c♯5	130.8– 554.4
Kontrafagott	cbassoon	32	a♯0–f3	29.1– 174.6
Cello *Vibrato*	cello-bv	47	c2–g5	65.4– 784.0
Klarinette	clari-ba	25	c♯2–c♯4	69.3– 277.2
B-Klarinette	clari-bfl	37	d3–d6	146.8–1174.7
Kontrabaß-Klarinette	clari-cb	25	f♯1–f♯3	46.2– 185.0
Es-Klarinette	clari-efl	32	g3–d6	146.8–1174.7
Becken ('crotales')	crota	13	c6–c7	1046.5–2093.0
E-Baß	elecbass 1	42	d1–e4	36.7– 329.6
E-Baß *'slap'*	elecbass 2	35	d1–a3	36.7– 220.0
E-Baß *'pop style'*	elecbass 3	29	a♯1–c4	58.3– 261.6
E-Baß *'deadnotes, pops'*	elecbass 4	4	–	–
E-Baß *'bright'*	elecbass 5	39	e1–d♯4	41.2– 311.1
E-Baß *'bright, harmonics'*	elecbass 6	29	e2–g♯4	82.4– 415.3
E-Gitarre	elecguitar 1	52	e2–d6	82.4–1174.7
E-Gitarre *verzerrt*	elecguitar 2	26	e2–e4	82.4– 329.6
EG., *'flanged harmonics'*	elecguitar 3	38	e2–e5	82.4– 659.3
EG., *'stereo chorus'*	elecguitar 4	42	e2–e5	82.4– 659.3
Englisch Horn	enghorn	30	e3–a5	164.8– 880.0
Flöte *'flutter-tongued'*	flute-flu	29	c4–e6	261.6–1318.5
Flöte *Vibrato*	flute-vib	37	c4–c7	261.6–2093.0

Tabelle A.1: Der Datensatz (Fortsetzung).

Instrument	**interner Name**	**Anzahl Dateien**	**Tonumfang** *Noten*	**Tonumfang** *Frequenz (Hz)*
Waldhorn	frehorn	37	d2–d5	73.4– 587.3
Waldhorn *gedämpft*	frehorn-m	37	d2–d5	73.4– 587.3
Glockenspiel	gks	30	g5–c8	784.0–4186.0
Marimba	marimba	56	f2–c7	87.3–2093.0
Oboe	oboe	32	a♯3–f6	233.1–1396.9
Klavier *laut*	piano-ld	88	a0–c8	27.5–4186.0
Klavier *gezupft*	piano-pl	88	a0–c8	27.5–4186.0
Klavier *weich*	piano-sft	88	a0–c8	27.5–4186.0
Piccoloflöte	picco	30	d5–g7	587.3–3136.0
Piccoloflöte *'flutter-tongued'*	picco-flu	24	d5–c♯7	587.3–2217.5
Alt-Saxophon	sax-alt	14	c♯4–d5	277.2– 587.3
Bariton-Saxophon	sax-bar	13	c2–c3	65.4– 130.8
Baß-Saxophon	sax-bass	8	g♯1–d♯2	51.9– 77.8
Sopran-Saxophon	sax-sop	15	c♯5–d♯6	554.4–1244.5
Tenor-Saxophon	sax-ten	14	c3–c♯4	130.8– 277.2
Sythesizerbass	syntbass	13	c1–c2	32.7– 65.4
Alt-Posaune	tromb-alt	13	f4–f5	349.2– 698.5
Baß-Posaune	tromb-bass	25	f1–f3	43.7– 174.6
Posaune *'pedal notes'*	tromb-pn	6	f1–a♯1	43.7– 58.3
Tenor-Posaune	tromb-ten	36	e2–d♯5	82.4– 622.3
Tenor-Posaune *gedämpft*	tromb-tenm	33	e–c5	82.4– 523.3
Bach-Trompete	trump-ba	32	b3–g6	246.9–1568.0
C-Trompete	trump-c	34	f♯3–d♯6	185.0–1244.5
C-Trompete *gedämpft*	trump-csto	31	f♯3–c6	185.0–1046.5
Tuba	tuba	32	e1–g4	41.2– 392.0
Vibraphon *gestrichen*	vibra-bow	37	f3–f6	174.6–1396.9
Vibraphon *geschlagen*	vibra-hm	37	f3–f6	174.6–1396.9
Viola *Vibrato*	viola-bv	42	c3–d6	130.8–1174.7
Viola *Vibrato, gedämpft*	viola-mv	39	c3–d6	130.8–1174.7
Violine *'artif. harmonics'*	violin-ah	13	g♯6–g♯7	1661.2–3322.4
Violine *Vibrato*	violin-bv	45	g3–c7	196.0–2093.0
Violine *Martellato*	violin-mar	37	g3–e6	196.0–1318.5
Violine *Vibrato, gedämpft*	violin-mv	45	g3–c7	196.0–2093.0
Violine *'nat. harmonics'*	violin-nh	12	g4–g6	196.0–3322.4
Violine *Pizzicato*	violin-piz	40	g3–g6	196.0–1586.0
Xylophon	xylo	44	f4–c8	349.2–4186.0

Tabelle A.2: Zusammenfassung der Instrumente zu Klassen.

Instrument	**Klasse**	**Instrument**	**Klasse**
aflute-vib	flute	piano-ld	piano
bassoon	bassoon	piano-pl	piano
bells	bells	piano-sft	piano
bflute-flu	flute	picco	flute
bflute-vib	flute	picco-flu	flute
cbassoon	bassoon	sax-alt	saxophone
cello-bv	cello	sax-bar	saxophone
clari-ba	clarinet	sax-bass	saxophone
clari-bfl	clarinet	sax-sop	saxophone
clari-cb	clarinet	sax-ten	saxophone
clari-efl	clarinet	synthbass	synthbass
crota	crota	tromb-alt	trombone
elecbass1	elecbass	tromb-bass	trombone
elecbass2	elecbass	tromb-pn	trombone
elecbass3	elecbass	tromb-ten	trombone
elecbass4	elecbass	tromb-tenm	trombone
elecbass5	elecbass	trump-ba	trumpet
elecbass6	elecbass	trump-c	trumpet
elecguitar1	elecguitar	trump-csto	trump-csto
elecguitar2	elecguitar-dist	tuba	tuba
elecguitar3	elecguitar-fh	vibra-bow	vibraphone
elecguitar4	elecguitar	vibra-hm	vibraphone
enghorn	oboe-enghorn	viola-bv	violin-viola
flute-flu	flute	viola-hm	violin-viola
flute-vib	flute	violin-ah	violin-viola
frehorn	frehorn	violin-bv	violin-viola
frehorn-m	frehorn	violin-mar	violin-viola
gks	gks	violin-mv	violin-viola
marimba	marimba	violin-nh	violin-viola
oboe	oboe-enghorn	violin-piz	violin-piz
		xylo	xylo

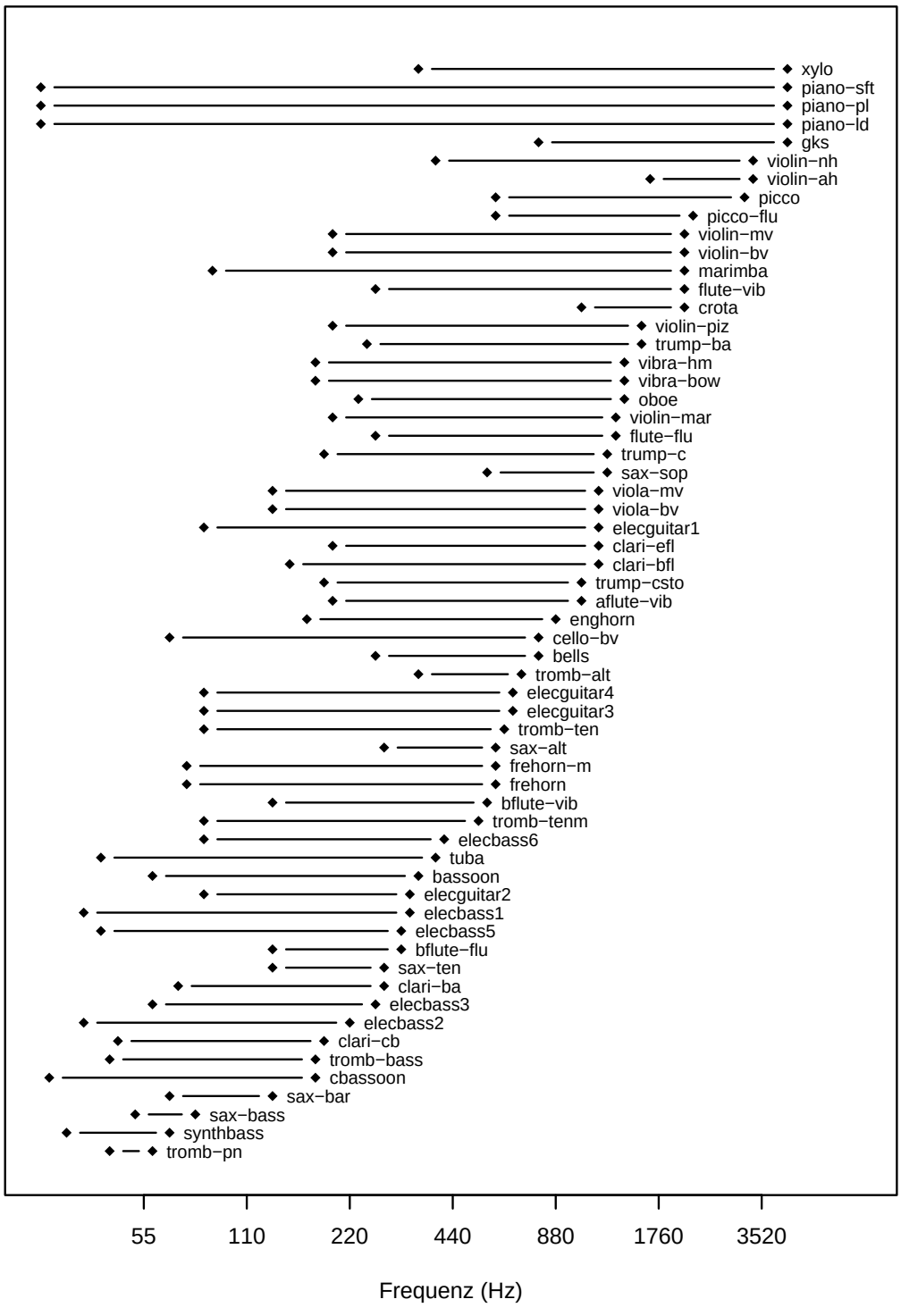

Abbildung A.1: Die Tonumfänge der Instrumente.

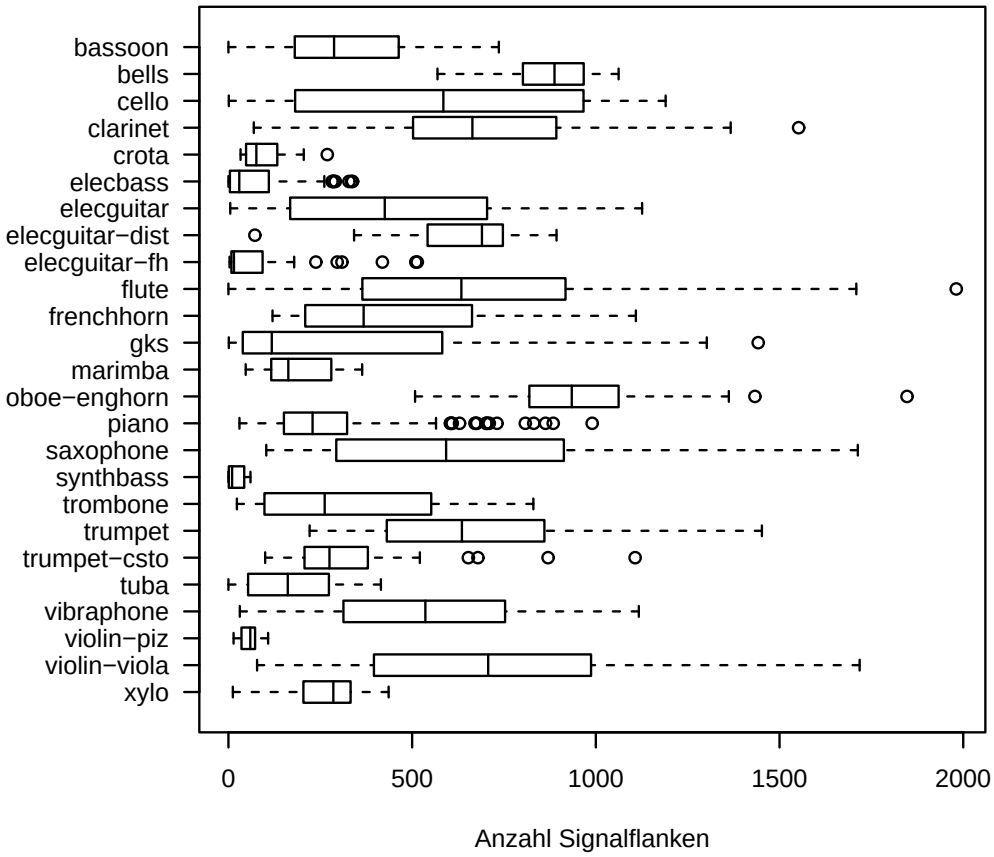

Abbildung A.2: Boxplot der Signalflanken nach Instrumenten.
Die „Box“ zeigt 1. bis 3. Quartil an, die Kreise sind Ausreißer.

Tabelle A.3: Aus den transformierten Daten abgeleitete Variablen.

Nr.	**Variable(n)**	**abgeleitete Maßzahl**	**Maß für...**
1	Tonfrequenz	(f)	Tonhöhe
2	”	Logarithmus $(\log_2(\frac{f}{440}))$	”
3	”	Signalflanken pro Periode	Rate
4	–	Signalflanken pro Sekunde	”
5	Amplitude	arithm. Mittel	Lage
6	”	Median	”
7	”	Modus	”
8	”	5%-Quantil	”
9, 10	”	1. und 3. Quartil	”
11	”	Standardabweichung	Streuung
12	”	mittlere Medianabweichung	”
13	”	Interquartilsabstand	”
14	”	Schiefemaß nach Pearson	Schiefe
15	”	Schiefemaß nach Yule-Pearson	”
16, 17	”	Kurtosis, (log-)	Wölbung
18, 19	”	Herfindahl-Index, (log-)	Konzentration
20, 21	”	Anteil der Amplituden bei Modus, (log-)	”
22	”	Kendall's τ	Autokorrelation
23	”	Mittelwertverschiebung	zeitl. Verlauf
24	”	Streuungsverschiebung	”
25	Frequenz	Mittel	Lage
26	”	Median	”
27, 28	”	5%- und 95%-Quantil	”
29, 30	”	1. und 3. Quartil	”
31	”	Standardabweichung	Streuung
32	”	Interquartilsabstand	”
33	”	Schiefemaß nach Yule-Pearson	Schiefe
34	”	Kendall's τ	Autokorrelation
35	”	Mittelwertverschiebung	zeitl. Verlauf
36	”	Streuungsverschiebung	”
37	Amplitude, Frequenz	Kendall's τ	Korrelation

Fortsetzung nächste Seite

Tabelle A.3: Aus den transformierten Daten abgeleitete Variablen (Fortsetzung).

Nr.	**Variable(n)**	**abgeleitete Maßzahl**	**Maß für...**
38	Quotient α	Mittel	Lage
39	”	Median	”
40, 41	”	1. und 3. Quartil	”
42, 43	”	5%- und 95%-Quantil	”
44	”	Standardabweichung	Streuung
45	”	Interquartilsabstand	”
46	”	Differenz von 5%- und 95%-Quantil	”
47	”	Schiefemaß nach Yule-Pearson	Schiefe
48	”	Kendall's τ	Autokorrelation
49	Quotient δ	Mittel	Lage
50	”	Median	”
51, 52	”	1. und 3. Quartil	”
53, 54	”	5%- und 95%-Quantil	”
55	”	Standardabweichung	Streuung
56	”	Interquartilsabstand	”
57	”	Differenz von 5%- und 95%-Quantil	”
58	”	Schiefemaß nach Yule-Pearson	Schiefe
59	”	Kendall's τ	Autokorrelation
60	α, δ	Kendall's τ	Korrelation
61	–	Zeitpunkt der ersten Signalflanke (log.)	Attackzeit
62	–	Zeitpunkt der letzten Signalflanke	Tondauer

Tabelle A.4: Fehlklassifikationsmatrix für das 1-Nearest-Neighbour auf den reinen Besetzungszahlen.

Die Zeilen geben jeweils an, wie oft das betreffende Instrument den verschiedenen Klassen zugeordnet wurde. Die letzte Spalte gibt die Fehlerrate für das Instrument an (jeweils in Prozent, gerundet).

%	ba	be	ce	cl	cr	eb	eg	ed	ef	fl	fr	gk	ma	ob	pi	sx	sy	tb	tp	tp	tu	vb	vp	vi	xy	Σ
bassoon	18	0	5	4	0	5	6	1	1	9	2	0	0	0	20	1	2	6	0	0	4	7	0	6	0	82
bells	0	74	1	0	0	0	4	9	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	9	0	26
cello	2	6	41	0	0	2	8	5	0	11	0	0	0	0	17	0	0	0	0	0	1	1	0	7	0	59
clarinet	1	0	4	14	0	2	3	5	0	16	2	0	0	5	11	7	0	2	8	0	0	4	0	16	0	86
crota	0	0	0	0	93	0	0	0	0	0	0	6	0	0	1	0	0	0	0	0	0	0	0	0	0	7
elecbass	0	0	0	0	0	58	9	0	4	4	0	0	6	0	13	0	2	0	0	0	1	1	2	0	0	42
elecguitar	2	5	5	2	0	13	19	4	0	11	1	0	1	0	17	0	0	2	1	1	1	2	0	12	0	81
elecguitar-dist	1	6	10	9	0	2	6	36	0	7	0	0	0	0	8	0	0	0	0	0	0	1	0	13	0	64
elecguitar-fh	0	0	0	0	0	31	3	0	34	11	0	0	0	0	6	0	0	0	0	0	0	3	8	4	0	65
flute	2	2	6	3	0	3	5	2	0	43	0	0	0	1	11	1	1	3	1	0	1	5	0	10	0	57
frenchhorn	0	0	0	0	0	0	0	0	0	1	92	0	0	0	0	0	0	2	2	0	0	1	0	1	0	8
gks	0	0	0	2	1	0	0	0	0	6	0	56	1	0	16	2	0	0	0	0	0	5	4	7	1	44
marimba	0	0	0	0	0	18	0	0	0	2	0	0	36	0	24	0	0	0	0	0	0	0	3	9	7	64
oboe/enghorn	0	3	7	5	0	0	4	3	0	17	1	0	0	14	5	2	0	1	11	0	0	10	0	17	0	86
piano	0	0	2	1	0	5	3	1	0	4	0	0	4	0	61	0	0	1	0	0	0	4	1	4	7	39
saxophone	6	0	2	7	0	4	3	0	1	11	5	0	0	3	9	7	0	9	11	0	2	3	0	16	0	93
synthbass	9	0	0	0	0	17	0	0	0	0	0	0	0	0	0	0	74	0	0	0	0	0	0	0	0	26
trombone	2	0	2	3	0	0	4	0	0	6	0	0	0	1	5	3	0	58	5	0	0	2	0	8	0	42
trumpet	0	0	0	8	0	0	9	0	0	9	5	0	0	6	2	8	0	15	27	2	0	3	0	6	0	73
trumpet-csto	5	0	3	0	2	7	2	0	0	3	1	0	0	0	6	0	0	3	0	60	0	0	0	7	0	40
tuba	7	0	15	0	0	11	8	0	0	17	0	0	0	0	16	0	1	0	0	0	21	4	0	0	0	79
vibraphone	1	2	4	6	0	6	6	0	1	20	1	0	1	6	16	0	0	3	1	0	0	13	0	12	1	87
violin-piz	0	0	0	0	0	12	0	0	9	2	0	0	4	0	0	0	0	0	0	0	0	0	71	2	0	28
violin/viola	1	5	8	2	0	3	4	4	0	14	0	0	2	2	9	0	0	1	0	0	0	2	1	41	0	59
xylo	0	0	0	0	0	1	0	0	0	0	0	0	21	0	35	0	0	0	0	0	0	0	0	0	44	56

Gesamtfehlerrate: 55.8%

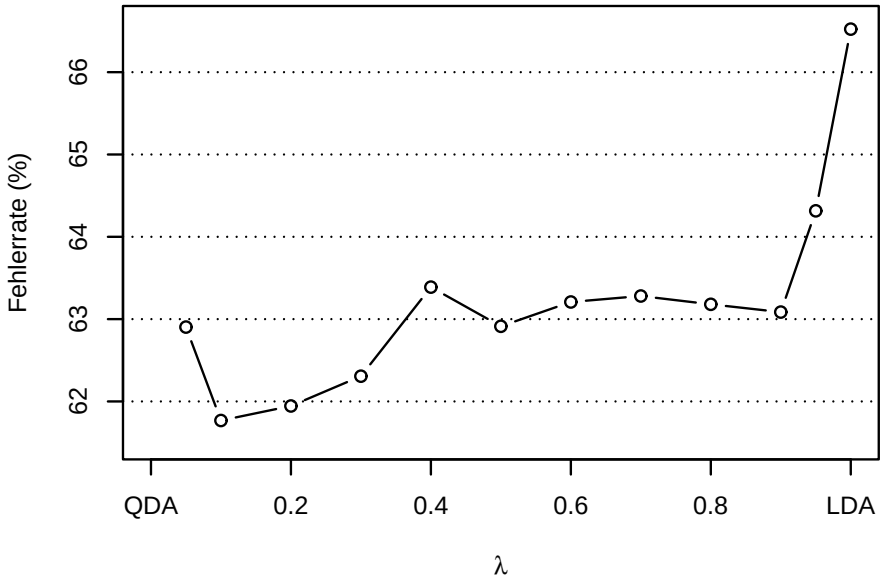

Abbildung A.3: Fehlerrate bei RDA auf Besetzungszahlen in Abhängigkeit von λ. γ ist $= 0$ und λ rangiert von 0.05 bis 1; das Optimum liegt bei 0.1.

Tabelle A.5: Ergebnis der Variablenselektion.

Es sind jeweils die Variablen-Nummern nach Tabelle A.3 sowie die resultierenden Fehlerraten in Prozent angegeben (siehe auch Abbildung 4.1 auf Seite 50).

Man beachte, daß z.B. die Variablen aus Schritt 13 und folgenden bei der RDA keine Bedeutung haben, da diese die Fehlerrate nicht mehr nennenswert verbessern.

Schritt	**LDA**		**NB**		**RDA**		**SVM**		***k*-NN**	
Nr.	Var.	Rate	Var.	Rate	Var.	Rate	Var.	Rate	Var.	Rate
2	62	84.35	3	76.76	57	80.26	61	71.86	61	78.62
3	4	76.18	23	62.72	61	67.28	57	58.25	3	57.07
4	19	67.69	32	54.91	17	50.88	17	48.57	51	46.79
5	61	59.73	62	50.38	12	42.98	46	42.46	62	41.57
6	57	53.10	61	45.02	23	37.81	23	40.19	30	38.23
7	23	49.11	8	41.44	25	34.08	21	38.52	23	34.75
8	1	47.18	37	39.94	62	31.00	47	36.96	9	32.64
9	30	45.06	47	38.88	6	28.89	11	37.16	57	30.74
10	47	43.02	35	37.20	4	26.99	51	36.18	10	30.40
11	5	41.40	21	36.41	37	25.52	19	36.03	49	29.76
12	60	39.92	54	36.07	32	23.71	60	36.26	1	29.68
13	12	38.95	41	36.14	41	23.80	54	36.23	38	29.72
14	29	38.69	24	35.42	27	23.77	37	36.45	31	29.39
15	34	38.36	25	35.63	36	23.58	14	36.62	43	28.70
16	26	37.72	34	35.50	26	23.54	41	36.46	17	28.78
17	53	37.33	28	35.35	1	23.61	58	36.85	29	28.45
18	14	36.99	33	35.82	38	23.47	62	36.94	16	28.49
19	10	36.69	13	36.09	5	23.41	24	37.11	4	28.54
20	37	36.30	58	36.36	34	23.55	40	37.23	54	28.48

B Mathematischer Anhang

B.1 Empirische Maßzahlen

Gegeben: *Stichprobe* $x_1, \ldots, x_n$, bzw. *geordnete Stichprobe* $x_{(1)}, \ldots, x_{(n)}$ mit $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$ und die *Häufigkeit* $h(x) = \sum_{i=1}^{n} \mathrm{I}_{\{x\}}(x_i)$ der Ausprägung x (nur sinnvoll bei diskretem Wertebereich).

- **Lagemaße**
  - arithmetisches Mittel: $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$
  - Modus: $x_D = x : h(x) = \max_{x_i} h(x_i)$
  - Median: $\tilde{x} = \begin{cases} x_{(\frac{n+1}{2})} & \text{falls} \quad n \text{ ungerade} \\ \frac{1}{2}(x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}) & \text{falls} \quad n \text{ gerade} \end{cases}$
  - p-Quantil: $x_p = x_{([n \cdot p + 0.5])}$ (für $0 < p < 1$)
- **Streuungsmaße**
  - Streuung (empirische Varianz): $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$
  - Standardabweichung: $s = \sqrt{s^2}$
  - mittlere quadratische Abweichung: $d^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$
  - Interquartilsabstand: $Q = x_{0.75} - x_{0.25}$
- **weitere Maßzahlen**
  - Schiefemaß nach Pearson: $g_1 = \frac{\bar{x} - x_D}{d}$

- Schiefemaß nach Yule-Pearson: $g_2 = \frac{3 \cdot (\bar{x} - \tilde{x})}{d}$
- Wölbung (auch *Kurtosis* oder *Exzeß*): $W = \frac{\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^4}{d^4} - 3$
- Herfindahl-Index (Konzentrationsmaß): $H = \sum_{j=1}^{m} p_j^2$, wobei $p_j = \frac{h(x_j)}{n}$ der Anteil der Stichprobe mit der j-ten Ausprägung (x_j) ist $(j = 1, \ldots, m)$.
- Kendall's tau (Korrelationsmaß):
  $\tau = \frac{2}{n(n-1)} \sum_{i<j} (\text{sign}(x_j - x_i) \cdot \text{sign}(y_j - y_i))$,
  wobei $\text{sign}(x) = \begin{cases} 1 & \text{falls} \quad x > 0 \\ 0 & '' \quad x = 0 \\ -1 & '' \quad x < 0 \end{cases}$
  Bestimmt man Kendall's τ für alle aufeinanderfolgenden Beobachtungspaare einer Zeitreihe, so mißt es die *Autokorrelation*; man betrachtet also die bivariate Stichprobe $\{(x_2, x_1), (x_3, x_2), \ldots, (x_n, x_{n-1})\}$.

- **spezielle Maßzahlen**
  - Signalflanken pro Sekunde: $\frac{n-1}{\varphi_n - \varphi_1}$
  - Signalflanken pro Periode: $\frac{n-1}{f \cdot (\varphi_n - \varphi_1)}$ (f ist dabei die Tonfrequenz)
  - Mittelwertverschiebung, Streuungsverschiebung:
    hierfür werden die Daten anhand der „zeitlichen Mitte“ $\frac{\varphi_n - \varphi_1}{2}$ in zwei Teile geteilt und anschließend für beide Hälften Mittel ($\bar{x}_1$ und $\bar{x}_2$) bzw. Standardabweichung (s_1 und s_2) der betreffenden Variable bestimmt. Die Verschiebung ist dann $\bar{x}_2 - \bar{x}_1$ bzw. $s_2 - s_1$.

B.2 Äquivalenz der Modelle

Vergleich der beiden (äquivalenten) Modelle aus Abschnitt 3.8 (Seite 41).
Hier sei nun o.B.d.A $t := 1$, so daß sich der Term λt jeweils zu λ vereinfacht.
$n_1, \ldots, n_k$ seien hier die Besetzungszahlen und N deren Gesamtsumme.
Gegebene Parameter sind: $\lambda \in \mathbb{R}^+$ und $p_1, \ldots, p_k \in [0,1]$ mit $\sum_{i=1}^{k} p_i = 1$.
Modell 1:

$$N \sim \text{Poisson}(\lambda), \tag{B.1}$$

$$n_1, \ldots, n_k | N \sim \text{Multinomial}(N, p_1, \ldots, p_k) \tag{B.2}$$

Demnach ist die k-dimensionale Dichte f_1 gegeben durch:

$$\begin{aligned} f_1(x_1, \ldots, x_k) &= \mathrm{P}_\lambda(N = \sum_{i=1}^{k} x_i) \cdot \mathrm{P}_{p_1,\ldots,p_k}(n_1 = x_1, \ldots, n_k = x_k | N = \sum_{i=1}^{k} x_i) \\ &= \frac{1}{(\sum x_i)!} \lambda^{(\sum x_i)} \exp(-\lambda) \cdot \frac{(\sum x_i)!}{x_1! \cdot \ldots \cdot x_k!} \prod_{i=1}^{k} p_i^{x_i} \end{aligned} \tag{B.3}$$

für alle $x_1, \ldots, x_k \in \mathbb{N}_0$.

Modell 2:

$$n_i \sim \text{Poisson}(\lambda \cdot p_i) \quad \forall\, i \in \{1, \ldots, k\} \tag{B.4}$$

Nach diesem Modell ergibt sich die Dichte f_2 als:

$$\begin{aligned} f_2(x_1, \ldots, x_k) &= \mathrm{P}_{\lambda,p_1,\ldots,p_k}(n_1 = x_1, \ldots, n_k = x_k) \\ &= \prod_{i=1}^{k} \mathrm{P}_{\lambda,p_i}(n_i = x_i) \\ &= \prod_{i=1}^{k} \left(\frac{1}{x_i!} (\lambda \cdot p_i)^{x_i} \cdot \exp(-\lambda \cdot p_i) \right) \end{aligned} \tag{B.5}$$

wiederum für alle $x_1, \ldots, x_k \in \mathbb{N}_0$.

Die Dichte f_1 nach Modell 1 läßt sich umformen...

$$\begin{aligned} f_1(x_1,\ldots,x_k) &= \tfrac{1}{(\sum x_i)!}\lambda^{(\sum x_i)}\exp(-\lambda)\cdot\tfrac{(\sum x_i)!}{x_1!\cdot\ldots\cdot x_k!}\prod_{i=1}^{k}p_i^{x_i} \\ &= \lambda^{(\sum x_i)}\exp(-\lambda)\cdot\prod_{i=1}^{k}\left(\tfrac{1}{x_i!}p_i^{x_i}\right) \\ &= \lambda^{(\sum x_i)}\exp\Big(-\lambda\underbrace{\sum_{i=1}^{k}p_i}_{=1}\Big)\cdot\prod_{i=1}^{k}\left(\tfrac{1}{x_i!}p_i^{x_i}\right) \qquad \text{(B.6)} \\ &= \left(\prod_{i=1}^{k}\lambda^{x_i}\right)\left(\prod_{i=1}^{k}\exp(-\lambda\cdot p_i)\right)\cdot\prod_{i=1}^{k}\left(\tfrac{1}{x_i!}p_i^{x_i}\right) \\ &= \prod_{i=1}^{k}\left(\tfrac{1}{x_i!}(\lambda\cdot p_i)^{x_i}\cdot\exp(-\lambda\cdot p_i)\right) \qquad = f_2(x_1,\ldots,x_k) \end{aligned}$$

...zur Dichte f_2 nach Modell 2. Sind beide Dichten gleich, so folgt daraus, daß auch die Verteilungen identisch sind. Da sie dieselbe Verteilung implizieren, sind beide Modelle äquivalent.

Modell 2 scheint zunächst wesentlich einfacher als Modell 1; z.B. wären für ein *festes* N in Modell 1 die Besetzungszahlen $n_1,\ldots,n_k$ paarweise (negativ) korreliert, wobei die Korrelation

$$\varrho(n_i,n_j)=\frac{\mathrm{Cov}(n_i,n_j)}{\sqrt{\mathrm{Var}(n_i)}\cdot\sqrt{\mathrm{Var}(n_j)}}=\frac{-p_ip_j}{\sqrt{p_i(1-p_i)}\cdot\sqrt{p_j(1-p_j)}}<0$$

beträgt, während bei Modell 2 die Besetzungszahlen von vornherein als unabhängig voneinander angenommen werden. Dadurch, daß N allerdings eben nicht fest, sondern eine poissonverteilte Zufallsvariable ist, geht die Abhängigkeit aber offensichtlich verloren.

Wenn $n_1,\ldots,n_k$ unabhängig poissonverteilt sind mit Parametern $\lambda_1,\ldots,\lambda_k$ (wie in Modell 2), so ist die Summe $N=\sum_{i=1}^{k}n_i$ wiederum Poisson$(\sum_{i=1}^{k}\lambda_i)$-verteilt (Gelman u. a., 1997). Bemerkenswert ist allerdings, daß der Vorgang offenbar eben auch *umgekehrt* funktioniert, sich eine poissonverteilte Zufallsvariable also mithilfe einer Multinomialverteilung wiederum in unabhängig poissonverteilte Summanden zerlegen läßt.

B.3 Vergleich Poisson- und χ^2-Modell

Das Modell aus Abschnitt 3.8 (Seite 41) soll dazu dienen, zwei Sätze von Besetzungszahlen (Histogramme) zu vergleichen. Ein ähnliches Problem behandelt der χ^2-Test (oder auch χ^2-Anpassungstest): Es werden die Häufigkeiten von Beobachtungen in bestimmten Klassen mit deren erwarteten Häufigkeiten verglichen. Genauer:

- $X_1, \ldots, X_n$ sind unabhängige Zufallsvariablen
- die Beobachtungen $x_1, \ldots, x_n$ werden in k (disjunkte) Klassen eingeteilt, es ergeben sich die *beobachteten Häufigkeiten* $n_1, \ldots, n_k$
- unter der Nullhypothese fallen die Beobachtungen mit Wahrscheinlichkeit p_i in Klasse i ($i = 1, \ldots, k$; $\sum p_i = 1$); die *erwartete Häufigkeit* in Klasse i ist damit $\tilde{n}_i = np_i$

Dann wird anhand der erwarteten und beobachteten Werte die χ^2-Teststatistik berechnet:

$$X^2 := \sum_{i=1}^{k} \frac{(n_i - \tilde{n}_i)^2}{\tilde{n}_i} \tag{B.7}$$

Die Teststatistik wird als χ^2_{k-1}-verteilt angenommen, entsprechend wird die Nullhypothese für „große“ X^2 abgelehnt, also wenn X^2 größer ist als das $(1-\alpha)$-Quantil der χ^2-Verteilung mit $k-1$ Freiheitsgraden, bei vorher festgelegtem Testniveau α (Büning und Trenkler, 1994).

Der Teststatistik wird (unter der Nullhypothese) eine χ^2-Verteilung unterstellt. Eine χ^2_k-Verteilung liegt vor, wenn die Teststatistik die Summe von k unabhängigen quadrierten standardnormalverteilten Zufallsvariablen ist.
Vor dem Quadrieren und Summieren werden die beobachteten Häufigkeit normiert: es ist $X^2 := \sum_{i=1}^{k} z_i^2$ mit $z_i := \frac{n_i - \tilde{n}_i}{\sqrt{\tilde{n}_i}}$. Damit die z_i (approximativ) standardnormalverteilt sind, muß also $\mathrm{E}[n_i] = \mathrm{Var}(n_i) = \tilde{n}_i$ sein.

In beiden Fällen (Poisson-Modell und χ^2-Modell) wird also eine Verteilung der Häufigkeiten n_i mit $\mathrm{E}[n_i] = \mathrm{Var}(n_i)$ unterstellt. Für große Stichprobenumfänge N

(und damit große Parameterwerte λ) kovergiert die Poissonverteilung gegen die Normalverteilung, also die Verteilung des χ^2-Modells.
Abbildung B.1 stellt einmal die Likelihoodfunktionen nach beiden Modellen für analoge Parameter gegenüber (entsprechend Abbildung 3.15, Seite 43). Die jeweils resultierende Diskriminanzfunktionen (entlang derer die Dichten gleich groß sind) sind nicht wesentlich verschieden. Man beachte, daß die linke Dichte im Gegensatz zur rechten diskret ist.

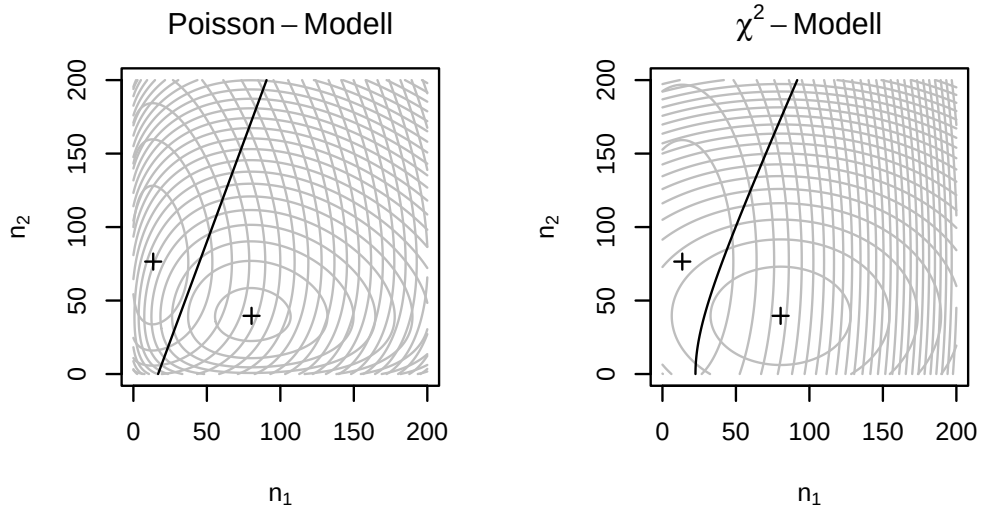

Abbildung B.1: Likelihoods bei Poisson- und χ^2-Modell.

C Literaturverzeichnis

[Backes und Gerlach 2000] Backes, G. ; Gerlach, M.: *Implementierung eines bestehenden Programms zur Simulation eines neuartigen Chips in eine zu entwickelnde Software-Umgebung*, FH Gießen-Friedberg, Diplomarbeit, Juli 2000

[Ballard 1981] Ballard, D. H.: Generalizing the Hough transform to detect arbitrary shapes. In: *Pattern Recognition* 13 (1981), Nr. 2, S. 111–122

[Bruderer 2003] Bruderer, M. J.: *Automatic recognition of musical instruments*, Ecole Polytechnique Fédérale de Lausanne, Diplomarbeit, Februar 2003

[Büning und Trenkler 1994] Büning, H. ; Trenkler, G.: *Nichtparametrische statistische Methoden*. 2. Auflage. Berlin : de Gruyter, 1994

[Epstein u. a. 2001] Epstein, A. ; Paul, G. U. ; Vetterman, B. ; Boulin, C. ; Klefenz, F.: A parallel systolic array ASIC for real time execution of the Hough-transform. In: *Proceedings of the 12th IEEE International Congress on Real Time for Nuclear and Plasma Sciences*. Valencia, Juni 2001, S. 68–72

[Fahrmeir u. a. 1996] Fahrmeir, L. ; Häussler, W. ; Tutz, G.: *Multivariate statistische Verfahren*. Kap. 8: Diskriminanzanalyse, S. 357–385. Berlin : de Gruyter, 1996

[Friedman 1989] Friedman, J. H.: Regularized Discriminant Analysis. In: *Journal of the American Statistical Association* 84 (1989), Nr. 405, S. 165–175

[Gelman u. a. 1997] Gelman, A. ; Carlin, J. B. ; Stern, H. ; Rubin, D. B.: *Bayesian data analysis*. New York : Chapman & Hall / CRC, 1997

[Goldenshluger und Zeevi 2002] Goldenshluger, A. ; Zeevi, A.: *The Hough transform estimator*. Mai 2002. – URL `http://www.gsb.columbia.edu/faculty/azeevi/PAPERS/Hough.pdf`. – Zugriffsdatum: Juni 2003

[Hastie u. a. 2001] Hastie, T. ; Tibshirani, R. ; Friedman, J.: *The elements of statistical learning; data mining, inference, and prediction*. New York : Springer-Verlag, 2001

[Hopfield und Brody 2000] Hopfield, J. J. ; Brody, C. D.: What is a moment? "Cortical" sensory integration over a brief interval. In: *Proc. Natl. Acad. Sci. USA* 97 (2000), Dezember, Nr. 25, S. 13919–13924

[Hopfield und Brody 2001] Hopfield, J. J. ; Brody, C. D.: What is a moment? Transient synchrony as a collective mechanism for spatiotemporal integration. In: *Proc. Natl. Acad. Sci. USA* 98 (2001), Januar, Nr. 3, S. 1282–1287

[Hough 1959] Hough, P. V. C.: Machine analysis of bubble chamber pictures. In: *International conference on high-energy accelerators and instrumentation*. Genève, September 1959, S. 554–556

[Ihaka und Gentleman 1996] Ihaka, R. ; Gentleman, R.: R: A language for Data Analysis and Graphics. In: *Journal of Computational and Graphical Statistics* 5 (1996), Nr. 3, S. 299–314

[Klefenz 1999] Klefenz, F.: Offenlegungsschrift DE 197 50 835 A 1 (Verfahren und Einrichtung zur Laufzeitdifferenzenbestimmung von akustischen Signalen), Deutsches Patent- und Markenamt, Juni 1999

[Klefenz und Brandenburg 2003] Klefenz, F. ; Brandenburg, K.: Offenlegungsschrift DE 101 57 454 A 1 (Verfahren und Vorrichtung zum Erzeugen einer Kennung für ein Audiosignal, Verfahren und Vorrichtung zum Aufbauen einer Instrumentendatenbank und Verfahren und Vorrichtung zum Bestimmen der Art eines Instruments), Deutsches Patent- und Markenamt, Juni 2003

[Mardia u. a. 1979] Mardia, K. V. ; Kent, J. T. ; Bibby, J. M.: *Multivariate Analysis*. London : Academic Press, 1979

[McGill 1987] Opolko, F. ; Wapnick, J.: *McGill University Master Samples*. 1987. – URL `http://www.music.mcgill.ca/resources/mums/html/`. – (CD-Set)

[Meyer 2001] Meyer, D.: Support Vector Machines. In: *R News* 1 (2001), September, Nr. 3, S. 23–26. – URL `http://CRAN.R-project.org/doc/Rnews/`

[Mood u. a. 1974] Mood, A. M. ; Graybill, F. A. ; Boes, D. C.: *Introduction to the theory of statistics*. 3rd edition. Singapore : McGraw-Hill, 1974

[Press u. a. 1992] Press, W. H. u. a.: *Numerical recipes in C*. 2nd edition. Cambridge : Cambridge University Press, 1992

[Shapiro 1978] Shapiro, S. D.: Feature space transforms for curve detection. In: *Pattern Recognition* 10 (1978), S. 129–143

[Theis u. a. 2002] Theis, W. ; Röver, C. ; Pouwels, B.: Implementing a new method for discriminant analysis when group covariance matrices are nearly singular / Sonderforschungsbereich 475, Universität Dortmund. URL `http://www.sfb475.uni-dortmund.de/dienst/de/content/veroeff-d/tr02-d.html`, 2002 (62/2002). – Technical Report

[Venables und Ripley 2002] Venables, W. N. ; Ripley, B. D.: *Modern Applied Statistics with S*. 4th edition. New York : Springer-Verlag, 2002

Index

Abtastrate, 6
Amplitude (A), 13
a-priori-Verteilung, 30
ASIC, 3
Auflösung, 6

Besetzungszahlen, 21
Bildpunkte, 9
Bildraum, 9

Center-Frequency, 13
Clusteranalyse, 24

Diskriminanzanalyse, 29
Diskriminanzfunktion, 31

Fehlerrate, 47

Hough-Charakteristika, 22
Hough-Histogramm, 11
Hough-Transformation, 8

Klang, 5
Klassifikationsbaum, 38
Klassifikation, 19
k-Nearest-Neighbour, 40
Kreuzvalidierung, 45

Lineare Diskriminanzanalyse (LDA), 29

Maximum-Likelihood-Entscheidungsregel, 30

Naive Bayes, 33

Parameterraum, 9
Phasenverschiebung (φ), 13
Pruning, 39

Quadratische Diskriminanzanalyse (QDA), 32

R, 46
Regularisierte Diskriminanzanalyse (RDA), 34

Sampling, 6
Signalflanke, 12
Stepwise Selection, 44
Support Vector Machine, 37

Überanpassung (Overfitting), 39, 44

Variablenselektion, 44
Verlust (-funktion), 30

Hiermit erkläre ich an Eides statt, daß ich die vorliegende Arbeit selbständig verfaßt und keine anderen als die im Literaturverzeichnis angegebenen Quellen verwendet habe.

Dortmund im Juli 2003 (Christian Röver)