

Vorlesung „Bildverarbeitung in der Medizin“

Teil 10: Klassifikation

Jürgen Braun, Dagmar Krefting – Institut für Medizinische Informatik



Klassifikation

- I. Klassifikation / Klassifikationsproblem
- II. Merkmalsraum, Merkmalsextraktion
- III. Klassifikation, Klassifikatoren
- IV. Merkmalsauswahl
- V. Klassifikationsverfahren

Einführung

2

Ziele der Vorlesung

Nach dieser Veranstaltung sollten Sie wissen:

- was ist Klassifikation und wie stellt sich das Klassifikationsproblem
- was bei der Signalauswertung zur Merkmalsextraktion zu beachten ist
- was Merkmal, Merkmalsextraktion und Merkmalsraum bedeuten
- was Klassifikatoren sind und wie sie erstellt werden
- wie Klassifikatoren trainiert werden können
- Problematik der Merkmalsauswahl und welche Verfahren dafür existieren
- welche wichtigen Klassifikationsverfahren existieren
- wie die Güte einer Klassifikation erfasst werden kann

Einführung

3

I. Einteilung von Klassifikationsverfahren

Bei einer Klassifikation werden in (Bild-) Daten enthaltene Sachverhalte, Erscheinungen und Ereignisse in einen Zusammenhang gestellt, indem ihre Beziehungen untereinander auf Regelmäßigkeiten untersucht werden. Die Einteilung der Werte einer meist kontinuierlichen Merkskala in Klassen erfolgt nach den Kriterien maximaler externer Separation und interner Homogenität (alle einer Klasse zugeordneten Werte sollen einander möglichst ähnlich sein, Unterschiede der Werte verschiedener Klassen sollen möglichst groß sein).

Klassifikationen sollten weiterhin vollständig (Abdeckung des gesamten Wertebereichs) und nichtüberlappend (eindeutig) sein.

Die Auswahl eines konkreten Klassifikationsverfahrens hängt letztendlich von der inhaltlichen Fragestellung und den Eigenschaften der Daten ab.

Klassifikation / Klassifikationsproblem

4

I. Auswertung von Signalen / Bildsegmenten

Bei der Auswertung von Signalen / Bildern / Bildsegmenten wird der Informationsgehalt auf wenige, für das jeweilige Material und Fragestellung aussagekräftige Merkmale reduziert und die relevanten Kenngrößen extrahiert.



Merkmale:
 unterer Schwellwert: a
 oberer Schwellwert: b
 Mittelwert: c

$$x_j \rightarrow \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 344 \\ 502 \\ 437 \end{pmatrix}$$

Klassifikation / Klassifikationsproblem

5

II. Merkmalsextraktion

Unter Merkmalsextraktion versteht man die Berechnung von n Merkmalen aus einem Signal oder einem Bild.

- Durch die Merkmalsextraktion wechselt man von der Betrachtung von Signalen oder Bildsegmenten zu Merkmalstupeln
- Dieser Vorgang ist mit einer starken Datenreduktion verbunden
- Das Originalsignal wird dadurch überflüssig

Merkmalsextraktion, Merkmalsraum

6

III. Überwachtes und unüberwachtes Training

Überwachtes Training: aufgrund von bekanntem Bildmaterial mit einer im voraus bekannten Zahl von Objekttypen, werden Musterklassen im Merkmalsraum ermittelt und festgehalten. Aus der statistischen Analyse dieser Trainingsgebiete werden spektrale Signaturen abgeleitet. Die Klassifikation des unbekanntem Bildmaterials beschränkt sich darauf, jeden Bildpunkt auf Ähnlichkeiten gegenüber diesen Signaturen hin zu untersuchen und zu entsprechend zu klassifizieren. Gebräuchliche Klassifikatoren dafür sind:

- maximum likelihood
- minimum distance

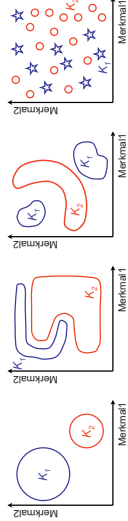
Unüberwachte Klassifikationen: zunächst sind weder die Anzahl der Musterklassen noch ihre Verteilung im Merkmalsraum bekannt. Dies ergibt sich durch eine Cluster-Analyse aus dem vorgelegten Bildmaterial während der eigentlichen Klassifikation.

III. Klassifikation - Grundlagen

Eigenschaften eines Merkmalsraums:

- n-dimensionaler Vektorraum
- Metrik (Abstandsmaß)
- Basisvektoren
- Basistransformationen

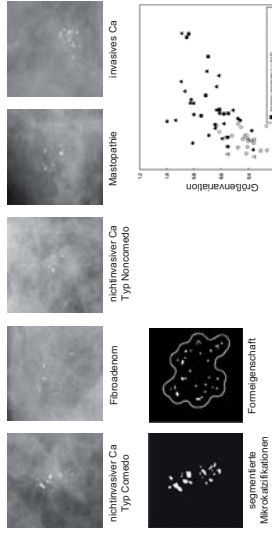
Cluster entsprechen Objekten oder Segmenten



Die Wahl des Klassifikationsverfahrens hängt auch von der Clusterform ab.

III. Klassifikation

Beispiel: Klassifikation von Brusttumoren in Mammographien



IV. Merkmalsauswahl

Problem

Reduktion der Merkmalsanzahl bei gleichbleibend guter Klassifikation (Reduktion Dimensionalität des Merkmalsraums), d.h. es werden möglichst wenige starke Merkmale für eine gute Klassifikation gesucht. → Rechenaufwand, begrenzter Lernsichprobenumfang

Lösungsansatz 1

Modellierung einer weniger, aber klinisch erprobter diagnostischer Kriterien (Merkmalsauswahl geringes Problem)

Lösungsansatz 2

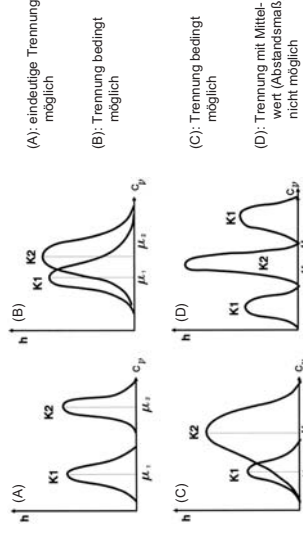
Modellierung einer Vielzahl von elementaren Merkmalen (Merkmalsauswahl von großer Bedeutung)

IV. Merkmalsauswahl

Gütekriterien zur Beurteilung und Auswahl geeigneter Untermengen von Merkmalen sind je nach Eigenschaften des Merkmalsraumes:

- Mittelwert und der Streuung
- quadratisches Abstandsmaß bei starken Abweichungen von der Normalverteilung
- Ambiguity-Funktion als Informationsmaß, das die Trennkraft eines Merkmals beurteilt
- ROC-Methode (Beobachtungskurve) zeigt wie gut ein System zwei Verteilungen unterscheiden kann

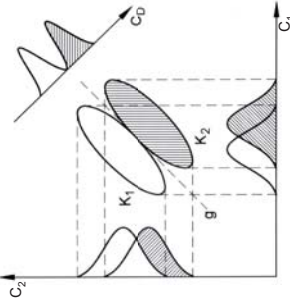
IV. Merkmalsauswahl



4 Beispiele von Merkmalsverteilungen der Klassen K1 und K2 (Mittelwerte μ_1 und μ_2 , h = Häufigkeit, c_p = beurteiltes Merkmal)

IV. Merkmalsauswahl

Koordinatentransformation (Linearkombination der Merkmale C_1 und C_2) zur Extraktion des Merkmals $C_3 = f(C_1, C_2)$, mit dem in diesem Beispiel das Klassifikationsproblem (Klassen K_1, K_2) gelöst werden kann.



Klassen müssen jedoch nicht immer so verteilt sein, daß aus einer Linearkombination aller Merkmale ein neues Merkmal berechnet wird und damit das Problem mit einem linearen Klassifikator gelöst werden kann.

19

Merkmalsauswahl

IV. Merkmalsauswahl

Weitere Methoden zur Merkmalsauswahl sind:

- Brute force: Alle möglichen Merkmalskombinationen testen
- CART (Classification And Regression Trees)
- Regressionsanalyse mit Likelihood-Ratio-Kriterium (Vorwärts- oder Rückwärtsselektion)
- Hauptkomponentenanalyse
 - Ausrichtung der Achsen in Richtungen stärkster Streuung (Übergang zu kombinierten Merkmalen)
- Verwendung der stärksten kombinierten Merkmale (Eigenvektoren der Kovarianzmatrix mit den größten Eigenwerten)

20

Merkmalsauswahl

IV. Hauptkomponentenanalyse

- Die Hauptkomponentenanalyse (Principal Component Analysis, PCA), Pearson, 1901) ist eine Methode der multivariaten Verfahren in der Statistik.
- grundsätzlich nicht an statistische Modellvoraussetzungen gebunden
 - rein numerisches Verfahren
 - Einführung in den 1930er Jahren von Harold Hotelling
 - erst seit den 1970ern aufgrund der Verfügbarkeit leistungsstärkerer Computer häufiger verwendet.

In der Bildverarbeitung: Karhunen-Loève-Transformation

21

Merkmalsauswahl

IV. Konzept der PCA

Es wird versucht, aus Variablen mit vielen Eigenschaften einige wenige signifikante Faktoren zu extrahieren, die für diese Eigenschaften bestimmend sind.

Mathematisch wird eine Hauptachsentransformation durchgeführt:

- Minimierung der Korrelation mehrdimensionaler Merkmale durch Überführung in einen Vektorraum mit neuer Basis
- die Hauptachsentransformation läßt sich durch eine Matrix angeben, die aus den Eigenvektoren der Kovarianzmatrix gebildet wird
- die Hauptkomponentenanalyse ist damit problemabhängig, für jeden Datensatz muß eine eigene Transformationsmatrix berechnet werden

22

Merkmalsauswahl

IV. Methodik der PCA

Voraussetzung: Varianz von Daten ist ein Maß für deren Informationsgehalt.

Varianz: Streuungsmaß in der Statistik, d.h. ein Maß für die Abweichung einer Zufallsvariable X von ihrem Erwartungswert E (Verallgemeinerung des Konzeptes der Summe der quadrierten Abweichungen vom Mittelwert in einer Beobachtungsreihe).

Daten: Punktwolke in n -dimensionalem kartesischen Koordinatensystem

Prinzipielles Vorgehen:

- es wird ein neues Koordinatensystem in die Punktwolke gelegt und rotiert
- die erste Achse wird so durch die Punktwolke gelegt, daß die Varianz der Daten in dieser Richtung maximal wird
- die zweite Achse steht senkrecht auf der ersten, in ihrer Richtung ist die Varianz am zweitgrößten (n -dimensionale Daten: n orthogonale Achsen)
- Decken die ersten Achsen (Gesamtvarianz Daten = Σ Achsenvarianzen) den größten Prozentsatz der Gesamtvarianz ab, sind die durch die neuen Achsen repräsentierten Faktoren ausreichend zur Beschreibung des Informationsgehalts der Daten.

23

Merkmalsauswahl

IV. PCA - statistisches Modell

- p Zufallsvariablen X_j , bezüglich ihrer Erwartungswerte zentriert (ihre Erwartungswerte wurden von der Zufallsvariablen subtrahiert)
- Zufallsvariablen werden in einem $(p \times 1)$ -Zufallsvektor x zusammengefaßt
- Erwartungswertvektor von x : Nullvektor und $(p \times p)$ -Kovarianzmatrix Σ , wobei Σ symmetrisch und positiv definit ist
- Eigenwerte λ_j ($j = 1, \dots, p$) der Matrix Σ sind absteigend der Größe nach geordnet. Sie sind Diagonalelemente in der Diagonalmatrix
- Die zugehörigen Eigenvektoren γ_j bilden die orthogonale Matrix Γ , es gilt: $\Lambda = \Gamma^{-1} \Sigma \Gamma$
- Der Zufallsvektor x wird linear transformiert: $x \rightarrow y = \Gamma^{-1} x$

Beispiel: 3D-Zufallsvektor

$$x = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \quad \Lambda = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} \quad \text{Matrix der Eigenwerte} \\ \text{mit } \lambda_1 > \lambda_2 > \lambda_3$$

24

Merkmalsauswahl

IV. PCA - statistisches Modell

Die Eigenvektoren γ_j lassen sich in der Matrix Γ zusammenfassen:

$$\Gamma = \begin{pmatrix} \gamma_{1A} & \gamma_{1B} & \gamma_{1C} \\ \gamma_{2A} & \gamma_{2B} & \gamma_{2C} \\ \gamma_{3A} & \gamma_{3B} & \gamma_{3C} \end{pmatrix}$$

Die Transformation $x \rightarrow y = \Gamma^T x$ führt zu den Gleichungen:

$$Y_A = \gamma_{1A}X_1 + \gamma_{2A}X_2 + \gamma_{3A}X_3$$

$$Y_B = \gamma_{1B}X_1 + \gamma_{2B}X_2 + \gamma_{3B}X_3$$

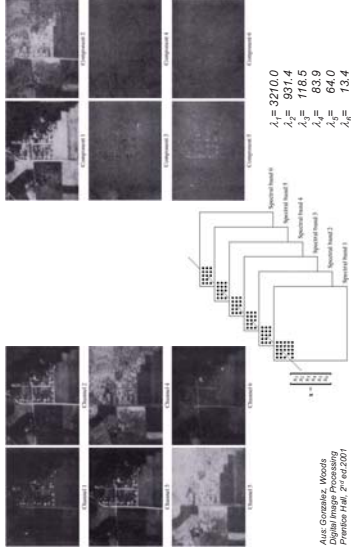
$$Y_C = \gamma_{1C}X_1 + \gamma_{2C}X_2 + \gamma_{3C}X_3$$

Die Varianz von $Y_A = \lambda_{\gamma_A}$. Die Hauptkomponente Y_A hat den größten Anteil an der Gesamtvarianz der Daten, Y_B den zweitgrößten Anteil, ...

Liegen konkret erhobene Daten mit p Merkmalen vor, wird aus den Merkmalswerten die Stichprobenkorrelationsmatrix errechnet. Aus ihr werden dann die Eigenwerte und Eigenvektoren für die Hauptkomponentenanalyse berechnet.

Merkmalauswahl

IV. PCA - Datenreduktion in Multispektralbild



Merkmalauswahl

V. Klassifikationsverfahren: KNN Klassifikator

Beim KNN Klassifikator (statistisches, nicht-parametrisches Verfahren) erfolgt die Klassifikation eines Merkmalsvektors m nach folgendem Ablauf (siehe auch Segmentierung I, Klassifikatoren):

- In der Stichprobe S werden Merkmalsvektoren \tilde{m} mit den geringsten Distanzen zu m bestimmt
- Zu den selektierten Merkmalsvektoren werden die zugehörigen Klassen K ermittelt und deren Auftretshäufigkeiten K_j berechnet
- Der Merkmalsvektor m wird der Klasse Ω_{max} mit maximaler Auftretshäufigkeit unter seinen K nächsten Nachbarn wie folgt zugeordnet:

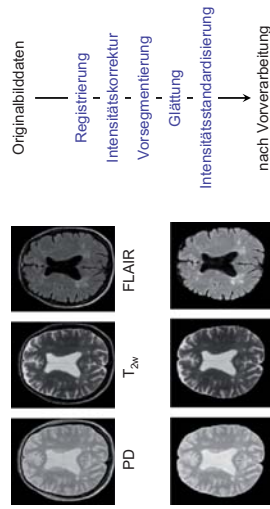
$$\Omega_{max} = \max_{k \in \{1, \dots, K\}} K_j$$

KNN-Verfahren werden mit geeigneten Daten trainiert. Sie sind rechenintensiv (Distanzberechnung) und besitzen bei mehrdimensionalen Merkmalsräumen hohe Speicheranforderungen).

Klassifikationsverfahren

V. KNN-Klassifikation von MR-Bildern

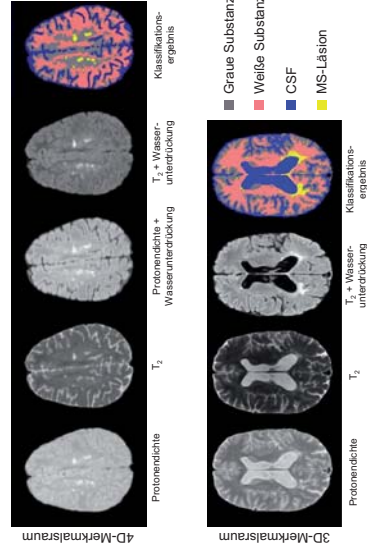
1. Schritt: Signalvorverarbeitung



aus: Ann Biomed Eng. 33, 1438-1448 (2005)

Klassifikationsverfahren

V. KNN-Klassifikation von MR-Bildern



Klassifikationsverfahren

V. Klassifikationsverfahren: Neuronale Netze

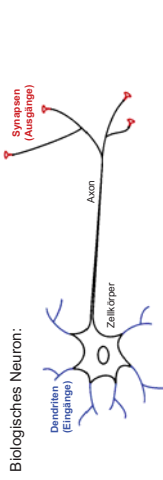
Künstliche neuronale Netze (KNN, engl. artificial neural network - ANN) sind Netze aus künstlichen Neuronen. Der Ursprung der künstlichen neuronalen Netze liegt in der Biologie. Man stellt sie natürlichen neuronalen Netzen gegenüber, welche Nervenzellverbindungen im Gehirn und im Rückenmark bilden. Ziel dabei ist die Abstraktion von Informationsverarbeitung, und nicht die Nachbildung biologischer neuronaler Netze.

KNN's basieren meist auf der Vernetzung künstlicher Neuronen. Die Topologie eines Netzes muß abhängig von seiner Aufgabe gut durchdacht sein. Nach der Konstruktion eines Netzes folgt die Trainingsphase, in der das Netz "lernt".

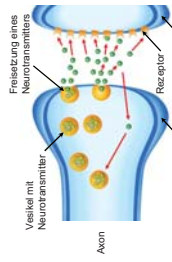
KNN's sind bei allen Anwendungen interessant, bei denen kein bzw. nur geringes explizites Wissen über das zu lösende Problem vorliegt.

Klassifikationsverfahren

V. Klassifikationsverfahren: Neuronale Netze

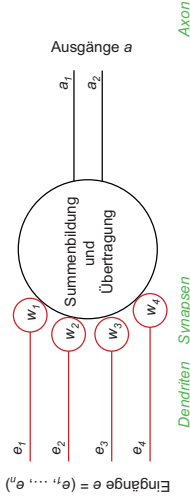


Informationstransfer:



V. Klassifikationsverfahren: Neuronale Netze

Künstliches Neuron:



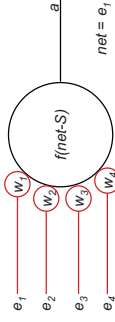
- **Eingang:** Die am Neuron eintreffenden Impulse verstärken oder schwächen dessen Aktivität (Summenbildung).
- **Zustand:** Die Übertragungsfunktion spiegelt den Ausgangswert des Neurons in Abhängigkeit von seiner Aktivierung wider.
- **Ausgang:** Wird ein Schwellenwert überschritten, so „feuert“ das Neuron: es sendet einen Impuls über das Axon aus.

V. Klassifikationsverfahren: Neuronale Netze

Komponenten von Neuronalen Netzen:

- Neuronen (sichtbar/unsichtbar)
- Ein- und Ausgänge
- Kopplungen (Übertragungsfunktionen, vorwärts/rückwärts)
- Schichten (Neuronenebene, Eingangs-/Ausgangs-/verborgene Schichten)

Übertragungsfunktion:



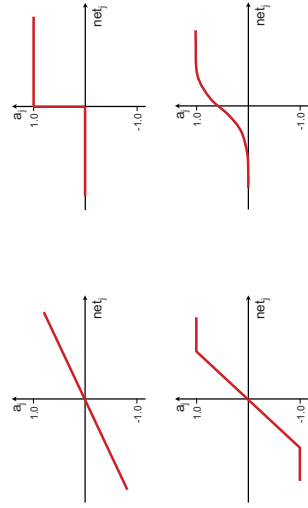
$$net = e_1 \cdot w_1 + e_2 \cdot w_2 + e_3 \cdot w_3$$

S: Schwelle

f(net - S): Übertragungsfunktion

V. Klassifikationsverfahren: Neuronale Netze

Beispiele für Übertragungsfunktionen



V. Klassifikationsverfahren: Neuronale Netze

Zum Lernen von Neuronalen Netzen werden die Gewichte so verändert, daß bei definierten Werten an den Eingängen des Netzes die entsprechend erwarteten Werte an den Ausgängen auftreten.
→ das „Wissen“ des Netzes verbirgt sich in den Gewichten

Ist bekannt, welchen Ausgang ein Netz bei gegebenen Eingängen produzieren soll (überwachtes Lernen), werden bei jedem Lernschritt Soll- und Ist-Ausgang des Netzes verglichen. Wenn Soll und Ist verschieden sind, werden die Gewichte mit Hilfe einer Lernregel systematisch verändert. Ist eine Klasseneinteilung nicht bekannt, muß das Netz Ähnlichkeiten in den Mustern selbst finden und Klassen festlegen (nicht-überwachtes Lernen).

V. Klassifikationsverfahren: Neuronale Netze

Wichtige Lernregeln sind die Hebb'sche und die Delta-Lernregel.

Hebb'sche Lernregel: Die synaptische Eigenschaft (Verstärken oder Hemmen) ändert sich proportional zum Produkt von prä- und postsynaptischer Aktivität (Hypothese, 1949).

$$\Delta w_{ij} = k \cdot e_i \cdot a_j$$

$k > 0$ Konstante (Lernrate),

e_i Eingang (Aktivierung von Neuron i)

a_j Ausgang (Aktivierung von Neuron j)

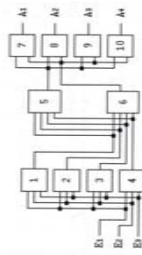
Delta Lernregel: sorgt dafür, daß die Gewichte nicht mehr korrigiert werden, wenn Soll und Ist übereinstimmen

$$\Delta w_{ij} = k \cdot e_i \cdot \Delta a_j \quad \Delta a_j \text{ Differenz zwischen dem Soll- und dem Ist-Output}$$

V. Klassifikationsverfahren: Neuronale Netze

Backpropagation (Rückpropagierung) ist ein verbreitetes Verfahren für das Einlernen von künstlichen neuronalen Netzen (1974, Paul Werbos, Durchbruch ab 1986, → „Renaissance“ in der Erforschung von KNN's).

Überwachtes Lernverfahren, wird als Verallgemeinerung der Delta-Regel auf mehrschichtige Netze angewandt. Dazu muß ein externer Lehrer existieren, der zu jedem Zeitpunkt der Eingabe den Zielwert, kennt. Die Backpropagation ist ein Spezialfall eines allgemeinen Gradientenabstiegsverfahrens, basierend auf dem mittleren quadratischen Fehler.



- 1...4 Eingangsschicht
- 5,6 verborgene Neuronen
- 7...10 Ausgangsschicht

37

Klassifikationsverfahren

V. Klassifikationsverfahren: Neuronale Netze

Beim Lernproblem wird für beliebige Netze eine möglichst genaue Abbildung von gegebenen Eingabevektoren auf gegebene Ausgabevektoren angestrebt. Dazu wird die Qualität der Abbildung durch eine Fehlerfunktion beschrieben, die hier durch den quadratischen Fehler R definiert wird:

$$R = \frac{1}{2} \sum_{i=1}^n (a_i - o_i)^2$$

n Anzahl der dem Netz vorgestellten Muster
 a_i gewünschte Soll-Ausgabe oder Zielwert
 o_i errechnete Ist-Ausgabe

Ziel ist nun die Minimierung der Fehlerfunktion, im Allgemeinen wird jedoch oft ein lokales Minimum gefunden. Das Einlernen des ANN's erfolgt beim Backpropagation-Verfahren durch die Änderung der Gewichte, da die Ausgabe des Netzes direkt von ihnen abhängig ist.

Bei der Fehlerminimierung erfolgt eine Unterscheidung der Neuronen. Liegt das Neuron in einer verdeckten Schicht, wird seine Gewichtung abhängig von dem Fehler geändert, den die nachfolgenden Neuronen erzeugen, welche wiederum ihre Eingaben aus dem betrachteten Neuron beziehen.

38

Klassifikationsverfahren

V. Klassifikationsverfahren: Neuronale Netze

Anwendung von künstlichen Neuronalen Netzen (artificial neural networks ANN) in der Medizin: **Risikoabschätzung für Prostatakrebs**
 Mit folgenden Merkmalen soll die Wahrscheinlichkeit für das Vorliegen eines Prostatakrebses unter Vermeidung unnötiger Prostatabiopsien ermittelt werden:



Alter: Patientalter
 PSA: Prostata Spezifisches Antigen, wird mit einer Blutuntersuchung bestimmt, ist ein Indiz für Veränderungen in der Prostata
 Ratio: Anteil von freiem PSA (f-PSA) zu gebundenem PSA. Bei Männern mit Prostatakrebs ist der f-PSA-Wert erniedrigt
 TRUS: Prostatavolumen bestimmt mit transrektalem Ultraschall
 DRU: Tastbefund der Prostata

39

Klassifikationsverfahren

V. Klassifikationsverfahren: Neuronale Netze

Netzeigenschaften:

- Neuronen:
 - 5 Eingangsschicht
 - 3 verborgene Neuronen
 - 1 Ausgangsneuron
- 3 Neuronenebenen:
 - Eingangsebene, verborgene Ebene, Ausgangsebene
- Übertragungsfunktion: tanh
- Backpropagation-Netzwerk
 - ~ 1000 Trainingsdatensätze
 - 10 Trainingsschritte / Datensatz
- Trainingsalgorithmus: Bayes'sche Regularisierung
- Eingangsmerkmale normiert
- Mittelwert = 0
- Standardabweichung = 1

40

Klassifikationsverfahren

V. Klassifikationsverfahren: Neuronale Netze



Alter: 56, PSA: 3, %fPSA: 11, Prostatavolumen: 25, DRU: neg → geringes Risiko

41

Klassifikationsverfahren

V. Klassifikationsverfahren: Neuronale Netze

Nachteile von KNN's:

- Das Trainieren von KNN's führt oft zu hochdimensionalen, nichtlinearen Optimierungsproblemen. Schwierigkeit: wurde das globale Optimum oder ein Nebenminimum gefunden.
- Die Sammlung oder Erzeugung von Trainingsdaten kann schwierig sein. Es muß verhindert werden, daß das Netz Eigenschaften der Muster lernt, die nur auf dem Trainingsset mit dem Ergebnis korreliert sind.
- Bei Anwendung „heuristischer“, nicht statistischer Vorgehensweisen neigen KNN's dazu, die Trainingsdaten auswendig zu lernen (Überanpassung). Es ist dann ist keine Verallgemeinerung auf neue Daten möglich.
- Die Kodierung der Trainingsdaten muß problemangepaßt und möglichst redundanzfrei sein. Die Form in der die zu lernenden Daten dem KNN präsentiert werden, hat großen Einfluß auf die Lerngeschwindigkeit und darauf, ob das Problem überhaupt vom Netz gelernt werden kann.

42

Klassifikationsverfahren

V. Geometrische Klassifikatoren - Minimum Distance

Der Minimum-Distance-Klassifikator ist ein geometrischer Ansatz zur Klassifikation der Pixel eines Bildes.

- Charakterisierung von Musterklassen durch ihr jeweiliges Zentrum (Mittelwert einer spektralen Signatur)
- Die Pixel eines Bildes werden derjenigen Klasse zugeordnet, zu deren Zentrum sie den geringsten Abstand haben

"Abstand" ist hier im geometrischen Sinn als euklidische Distanz von Pixeln zum Zentrum einer Musterklasse gemeint.

Dieser Ansatz hat den Vorteil, daß die Zuordnung vollständig und meist eindeutig ist. Die Minimum Distance Klassifikation stellt oft einen günstigen Kompromiß zwischen rechen-technischem Aufwand und Klassifikationsqualität dar.

V. Geometrische Klassifikatoren - Maximum Likelihood

Bei diesem Verfahren wird die multivariate Streuungsinformation im Merkmalsraum zur Richtungs-differenzierung berücksichtigt. Im Gegensatz zu deterministisch - arithmetischer Distanzberechnung wird ein statistisches Ähnlichkeitsmaß berechnet, das von der mehrdimensionalen Verteilung der "Punktwolke" jedes Trainingsgebietes abhängig ist.

Dem Nachteil des hohen Rechenaufwandes dieses Verfahrens steht der Vorteil guter Klassifikationsergebnisse entgegen, die in der Regel denen alternativer Ansätze überlegen sind. Zusätzlich besteht die Möglichkeit statistischer Qualitätsaussagen hinsichtlich der Klassifikation jedes Pixels aufgrund der bekannten Verteilungseigenschaften, die aus der mehrdimensionalen Stichprobe der Trainingsgebiete geschätzt werden.

Ebenso können auch a-priori Wahrscheinlichkeiten auf Grund des vorab bekannten Anteiles einzelner Klassen innerhalb des Untersuchungsgebietes festgelegt werden (Bayesian Classifier, Bayesian decision rule).

V. Klassifikationsverfahren: Logistische Regression

Die logistische Regression (Hosmer, Lemeshow 1988) ist ein Verfahren zur multivariaten Analyse binärer abhängiger Variablen, die nicht mit linearen Regressionsanalysen untersucht werden können.

Bei der Klassifikation durch logistische Regression wird berechnet, ob ein Element des Merkmalsraums zu einer von 2 möglichen Klassen gehört. Die Werte für diese Vorhersage (Zielvariable oder abhängige Variable Y) sind 'Ja' oder 'Nein' (bzw. 1 oder 0). Statt der Zielvariablen selbst wird die Wahrscheinlichkeit des Auftretens einer ihrer beiden Werte abgeschätzt (Schätzung der Chance, daß einer der beiden Zustände eintritt).

Für den Logarithmus des Verhältnisses von Wahrscheinlichkeit und Gegenwahrscheinlichkeit (Logit), mit der Y den Wert 1 oder 0 annimmt

$$\text{Logit}(Y_{ij}) = \ln \frac{P(Y=1)}{1-P(Y=1)} = \ln \frac{P(Y=1)}{P(Y=0)}$$

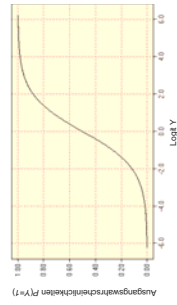
wird dann folgende Regressionsgleichung geschätzt:

$$\text{Logit}(Y_{ij}) = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_n X_n$$

V. Klassifikationsverfahren: Logistische Regression

Es werden Regressionsgewichte bestimmt, mit denen die Logits für eine gegebene Matrix von unabhängigen Variablen X berechnet werden können (Transformation von Wahrscheinlichkeiten).

Bestimmung der Regressionsparameter: Maximum Likelihood-Methode.



Zusammenhang Logits (x-Achse) mit den Ausgangswahrscheinlichkeiten $P(Y=1)$ (y-Achse).

Zusammenhang logistische Regression und Wahrscheinlichkeiten $P(Y=1)$:

$$P(Y=1) = \frac{\exp(b_0 + b_1 X_1 + b_2 X_2 + \dots + b_n X_n)}{1 + \exp(b_0 + b_1 X_1 + b_2 X_2 + \dots + b_n X_n)}$$

V. Klassifikationsverfahren: Bayes-Klassifikator

Es handelt sich um einen statistischen Klassifikator. Das zugrundeliegende Bayestheorem (Satz von Bayes) stammt aus der Wahrscheinlichkeitstheorie und beschäftigt sich mit der Berechnung bedingter Wahrscheinlichkeiten. Der Satz von Bayes erlaubt in gewissem Sinn die Umkehr von Ereignis-Ursache zusammenhängen.

$$\text{Satz von Bayes: } P(A|B) = \frac{P(\beta, A) \cdot P(A)}{P(\beta)}$$

Hierbei sind:

- $P(A)$ die a-priori-Wahrscheinlichkeit für ein Ereignis A und
- $P(\beta, A)$ die Wahrscheinlichkeit für ein Ereignis B unter der Bedingung, daß A auftritt und
- $P(B)$ die a-priori-Wahrscheinlichkeit für ein Ereignis B

V. Klassifikationsverfahren: Bayes-Klassifikator

Der Bayes-Klassifikator ordnet jedes Objekt der Klasse zu, zu der es mit der größten Wahrscheinlichkeit gehört (vollständige Klassifikation des Merkmalsraumes, d.h. jedem Punkt des Merkmalsraums wird eine Klasse zugeordnet).

Zur Definition eines Bayes-Klassifikators wird ein Kostenmaß benötigt, das jeder Klassifizierung Kosten zuweist. Der Bayes-Klassifikator minimiert die durch alle Klassifizierungen entstehenden Kosten.

Bei einem Kostenmaß, das ausschließlich bei Fehlentscheidungen Kosten verursacht, minimiert der Bayes-Klassifikator die Wahrscheinlichkeit einer Fehlentscheidung.

Voraussetzung: die Wahrscheinlichkeit, mit der ein Punkt des Merkmalsraums zu einer bestimmten Klasse gehört, ist bekannt (jede Klasse ist durch eine Wahrscheinlichkeitsdichtefunktion beschrieben).

Da die Dichtefunktionen unbekannt sind, erfolgt eine Abschätzung unter der Annahme einer Normalverteilung der Wahrscheinlichkeiten für jede Klasse. Deren Parameter werden anhand der vorhandenen Daten abgeschätzt.

V. Klassifikationsverfahren: Bayes-Klassifikator

Für einen Bayes-Klassifikator liest sich der Satz von Bayes wie folgt:

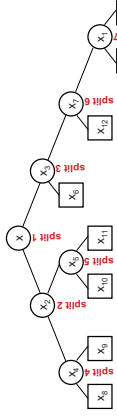
$$P(K_i | x) = \frac{P(x|K_i) \cdot P(K_i)}{P(x)}$$

- $x = (x_1, x_2, \dots, x_M)$ M -dimensionaler Merkmalsvektor
- $P(K_i)$ Wahrscheinlichkeit für das Auftreten der Klasse K_i , $i = 1, \dots, M$
(a-priori Wahrscheinlichkeit vom K_i)
- $P(x|K_i)$ Wahrscheinlichkeit für das Auftreten des Vektors x in Klasse K_i
(klassenbedingte Wahrscheinlichkeit)
- $P(x)$ Wahrscheinlichkeit den Vektor x zu messen (ohne Vorgabe einer bestimmten Klasse, a-priori Wahrscheinlichkeit vom x)
- $P(K_i|x)$ Wahrscheinlichkeit, daß ein gegebener Vektor x aus der Klasse K_i stammt (**gesucht**)

V. Klassifikationsverfahren: CART

CART (Classification and Regression Trees, 1984 Leo Breiman), statistisches Verfahren zur Entscheidungsfindung mit Hilfe von Entscheidungsbäumen.

Charakteristisches Merkmal des CART-Algorithmus: es werden nur Binärbäume erzeugt (jede Verzweigung besitzt immer zwei Äste). Das zentrale Element dieses Algorithmus ist das Finden einer optimalen binären Trennung.



Beginnend von einem stark diskriminierenden Merkmal erfolgt die fortgesetzte Unterteilung der Objektmenge durch „splits“. Komplexere Bäume ergeben eine bessere Anpassung an die Lerndaten, bei einem Übertraining ergeben sich Probleme mit der Generalisierungsfähigkeit.

Zusammenfassung

- Ziel der Klassifikation
 - Zuordnung unbekannter Bildobjekte zu Klassen
- Voraussetzungen
 - Segmentierung und Merkmalsextraktion
- Merkmalsauswahl
 - Wahl signifikanter, gegebenenfalls normierter Merkmale ist entscheidend, da das Klassifikationsergebnis auf der Güte gewählter Merkmale beruht
- Klassenzuordnung
 - Zuordnung über verschiedene Verfahren
- Statistische Klassifikatoren
 - Bayes, Cart, logistisches Modell
- Nichtstatistische Klassifikatoren
 - Neuronale Netze, geometrische Klassifikatoren