

# Data-Mining

## Zusammenfassung

Aless Lasaruk

10. November 2005

### Zusammenfassung

Dieses Dokument enthält Fragen und Antworten zum Fach Data-Mining. Die Fragen sind durch systematisches Durchgehen des gleichnamigen Skriptums entstanden, decken aber den Stoff nicht vollständig. Man beachte auch, dass die Antworten nicht exakt formuliert sind. Das Ziel der Formulierung ist das Minimieren der schriftlichen Antwort und Maximieren der mündlichen.

## Inhaltsverzeichnis

1	Einleitung	1
2	Grundlagen	2
3	Datenvorverarbeitung	3
4	Merkmalss Selektion	3
5	Hauptkomponentenanalyse	5
6	Clusteringalgorithmen	6
7	Klassifikation	7
8	Entscheidungsbäume	8
9	Assoziationsregeln	9
10	Support-Vector-Machines	10
11	Bayes-Netze	11
12	Ensemble-Verfahren	12
13	Algorithmusliste	14

## 1 Einleitung

- **Was versteht man unter Data-Mining?**  
Extrahierung von wichtigen Daten aus großen Datenbanken.
- **Wann benötigt man Data-Mining?**  
Wissensbasierte Entscheidungen, Variable Umgebungsbedingungen, konventionelle Methoden führen nicht zu Erfolg, Daten sind verfügbar.
- **Was ist als Ergebnis des Data-mining Prozesses zu betrachten?**  
Nicht eine Teilmenge der Daten, sondern das Ergebnis der Analyse.

- **Welche zentralen Konzepte gibt es in Data-Mining?**  
Clustering, Klassifikation, Assoziationsregeln, Entscheidungsbäume, Bayes-Netze, Support-Vector-Machines, Ensembletechniken.
- **Was ist bei der Datenauswertung wichtig?**  
Zuerst Bereinigen und Auswählen (Vorverarbeitung) und dann Weiterverarbeiten.
- **Welche Anwendungen von Datamining kennen Sie?**  
Kundenverhalten, Intrusion Detection, Unterschriftenverifikation, Prozessidentifikation und Optimierung, Optische Qualitätskontrolle.

## 2 Grundlagen

- **Was ist ein *Datensatz*?**  
Matrix aus Beispieldaten.
- **Was ist ein *Attribut* oder *Feature*?**  
Einzelne Spalte des Datensatzes.
- **Was ist ein *Muster*?**  
Einzelne Zeile des Datensatzes.
- **Wann ist ein Attribut *nominal* (bzw. *ordinal*)?**  
Wenn er einen diskreten endlichen (geordneten) Wertebereich hat.
- **Was ist eine *Ratiogröße* (bzw. *Intervallgröße*)?**  
Feste Ordnung, aber (kein) spezifischer Nullpunkt. Verhältnisse ergeben (keinen) Sinn.
- **Ist jede kontinuierliche Größe eine Rationgröße?**  
Nein z.B. Winkel.
- **Wie behandelt man *missing Values*?**  
Nicht verwenden, Berücksichtigung durch das Verfahren, Schätzen.
- **Wie kann man Wissen repräsentieren?**  
Regeln, Bäume, Neuronale Netze, Assoziationsregeln.
- **Welche Kriterien soll das neu erworbene Wissen erfüllen?**  
Stichhaltig (Generalisierung), Neu, Nützlich, Verständlich.
- **Aus welchen Schritten besteht der Prozess des *KDD* (Knowledge Discovery in Databases)?**  
Vorbereitung, Datenselektion, Vorverarbeitung, Datentransformation, Datamining, Interpretation, Anwendung des Wissens.
- **In welche Teilbereiche kann man Datamining einordnen?**  
Deskriptiv, Explorativ, Prädiktiv.
- **Aus welchen Komponenten bestehen Data-Mining-Algorithmen?**  
Modell, Bewertungsfunktion, Suchalgorithmus.
- **Was versteht man unter *Bias*?**  
Genauigkeit vom Modell.
- **Was versteht man unter *Varianz*?**  
Streuung des Modells (Spezifität).
- **Welcher Zusammenhang besteht zwischen Bias und Varianz?**  
Es gibt ein Trade-Off. Mehr freie Parameter liefern geringen Bias aber hohe Varianz und umgekehrt.
- **Welche Methoden zum Testen von Modellen kennen Sie?**  
*Holdout*-Methode (Daten Aufteilen), *Kreuzvalidierung*, *Jackknife* (*Leave-One-Out*), *Bootstrap* (Ziehen mit Zurücklegen).
- **Wann ist eine Überanpassung möglich?**  
Hohe Modellkomplexität, wenige Trainingsmuster unterschiedliche Algorithmen und Parameteranpassung.

### 3 Datenvorverarbeitung

- **Wie rechnet man empirische Varianz aus?**  
 $\sigma^2 = \frac{1}{n-1} \sum_{k=1}^n (x_i(k) - \mu_i)^2$ .
- **Wie Berechnet man Kovarianz?**  
 $\sigma^2 = \frac{1}{n-1} \sum_{k=1}^n (x_i(k) - \mu_i)(x_j(k) - \mu_j)$ .
- **Kann Kovarianz negativ sein?**  
Ja.
- **Was drückt das Vorzeichen der Kovarianz aus?**  
Positiv bedeutet gleiche Tendenz und negativ entgegengesetzte Tendenzen.
- **Was ist ein Korrelationskoeffizient?**  
Normierung der Kovarianz durch Streuungsprodukt  $\sigma_i \sigma_j$ .
- **Was sagt der Korrelationskoeffizient aus?**  
Positiven oder negativen linearen Zusammenhang (je mehr desto mehr bei 1). Ist aber kein ursächlicher Zusammenhang.
- **Wenn der Korrelationskoeffizient bei 0 ist, was sagt es aus?**  
Es besteht kein linearer Zusammenhang (nicht gar kein Zusammenhang).
- **Welche Matrix liegt der Mahalanobis-Norm zugrunde?**  
Inverse Kovarianzmatrix.
- **Welche Abstände kann man sonst verwenden?**  
Kosinusdes Winkels, Hamming-Abstand.
- **Wann ist ein Größenvergleich sinnvoll?**  
Wenn die Größen normalisiert sind.
- **Wie normalisiert man Größen?**  
Man bildet den maximalen Wert auf  $b$  und minimalen auf  $a$  bei einem Intervall  $[a, b]$  durch eine lineare Transformation.
- **Welche Nachteile hat Normalisierung?**  
Neue Daten können aus dem maximalen Bereich herausfallen, das Intervall kann schlecht ausgenutzt sein.
- **Was ist Standardisierung?**  
Zentrieren (Abziehen vom Mittelwert) und Durchteilen durch die Streuung.
- **Wie ist die empirische Streuung definiert?**  
Wurzel aus der empirischen Varianz.
- **Wann wird ein Attribut als Ausreißer erkannt?**  
Wertebereich, dreifache Standardabweichung, Abweichung von einer Schätzung.
- **Wie erkennt man Ausreißer?**  
Filterung, Ergänzung, Interpolation.
- **Wie kodiert man nominale Merkmale?**  
Orthogonale Kodierung.

### 4 Merkmalsselektion

- **Welche Aufgaben hat Merkmalsselektion?**  
Auswählen einer "optimalen" Teilmenge.
- **Was nützt Merkmalsselektion?**  
Zeitersparnis, Bessere Datenqualität, Generalisierbarkeit, Verständnis der Daten, Kosten.
- **Wie kann man Merkmalsselektion formal definieren?**  
 $P(C|G) = P(C|F)$  oder so nahe, wie möglich.

- **Wieso liefert eine Merkmalsselektion oft bessere Ergebnisse in Datamining?**  
Geringere Anzahl von Merkmalen verhindert Überanpassung.
- **Wann ist ein Merkmal *stark relevant*?**  
 $P(C|F_i, S_i) \neq P(C|S_i)$ . Entfernen zerstört Verteilung.
- **Wann ist ein Merkmal *schwach relevant*?**  
 $P(C|F_i, S_i) = P(C|S_i)$  aber es existiert ein  $S'_i \subset S_i$  mit  $P(C|F_i, S'_i) \neq P(C|S'_i)$ . Entfernen zerstört Verteilung unter bestimmten Bedingungen.
- **Wann ist ein Merkmal *irrelevant*?**  
Für alle Teilmengen  $S'_i$  von  $S_i$  gilt  $P(C|F_i, S'_i) = P(C|S'_i)$ .
- **Wann sind Merkmale *redundant*?**  
Wenn sie vollständig korreliert sind.
- **Wann ist eine Menge von Merkmalen optimal?**  
Wenn sie alle stark relevanten, keine irrelevanten und eine Teilmenge der schwach relevanten Merkmale enthält.
- **Was ist die allgemeine Vorgehensweise?**  
Auswählen von Teilmengen solange ein Kriterium nicht erfüllt ist.
- **Was ist der Unterschied zwischen *Filtern* und *Wrappern*?**  
Filter bewerten die Güte von Merkmalen anhand einer Analyse der Daten im Merkmalsraum. Wrapper bewerten die Güte einer Modellinstanz.
- **Welche der beiden Techniken ist qualitativ besser anzusehen?**  
Wrapper, da die Aufgabenstellung des Datamining berücksichtigt wird.
- **Welche grundlegenden Tatsachen sind für ein Verfahren festzulegen?**  
Suchrichtung und Suchstrategie.
- **Welche Suchrichtungen kennen Sie?**  
Sequential Forward Generation (Merkmale Hinzunehmen), Sequential Backward Generation (Entfernen), Bidirectional Generation (beides konkurrierend), Random Generation.
- **Welche Suchstrategien kennen Sie?**  
Exhaustive/Complete (Monotonie), Heuristic, Nondeterministic.
- **Welche *Evaluationsmaße* kennen Sie?**  
Genauigkeitsmaß, Informations/Distanz, Abhängigkeits, Konsistenzmaße.
- **Wie ist das Entropiemaß von Shanon definiert?**  
 $E(D) = \sum_{i=1}^d p_i \log_2 p_i$ . Je vorhersagbarer, desto kleiner Informationsgehalt und Entropie.
- **Wie ist das *Information-Gain* Maß definiert?**  
 $Information-Gain(IG(X) = I(D) - \sum_{j=1}^p \frac{|D_j|}{|D|} I(D_j^X)$  mit  $I(D_j^X)$  im Wesentlichen Entropie, wobei die Wahrscheinlichkeit des Merkmals  $X$  in Klasse  $j$  ist.
- **Wie ist das Distanzmaß *Directed Divergence* definiert?**  
Nachschauen. Ich bin doch nicht blöde das auswendig zu lernen.
- **Wann heißen zwei Muster *inkonsistent*?**  
Wenn sie bis auf die Klassenzuordnung identisch sind.
- **Was ist die *Inkonsistenzzahl*?**  
Anzahl aller übereinstimmenden Muster minus der größten Anzahl der Muster mit unterschiedlicher Klassenzuordnung.
- **Was ist die *Inkonsistenzrate*?**  
Summe aller Inkonsistenzzahlen geteilt durch die Gesamtzahl der Muster.
- **Welche Eigenschaften hat die Konsistenzrate?**  
Linear berechenbar in der Anzahl der Muster, monotonen Maß, erlaubt vollständige Suche.

- **Wie wählt man auf der Basis von IR die Teilmengen?**  
Wenn die Teilmengen gleich groß mit gleichen Raten sind, dann sind die nicht unterscheidbar. Sonst wird lexikographische Rate, Kardinalität vorgezogen (kleinere wird vorgezogen).
- **Wann bricht man die Algorithmen ab?**  
Zeit, Anzahl Merkmale, Iterationen.
- **Welche Algorithmen kennen Sie?**  
*Focus, Automated Branch and Bound, Relief, EA-RBF.*
- **Was ist die Idee von *Focus*?**  
Betrachte aufsteigend alle  $k$ -elementigen Teilmengen bis die Inkonsistenzrate gleich 0 ist.
- **Welche Eigenschaften hat *Focus*?**  
Filter, Überwacht, SFG, Umfassende Suche, Konsistenzmaß.
- **Was ist die Idee von *ABB*?**  
Erzeuge Teilmengen durch Weglassen eines Merkmals. Wenn die Inkonsistenzrate steigt, markiere die Teilmenge als ungültig. Fahre mit Breitensuche fort.
- **Wovon hängt die Komplexität von *ABB* ab?**  
Betrachteten Teilmengen, Anzahl der Merkmale.
- **Welche Eigenschaften hat *ABB*?**  
Filter, Überwacht, SBG, vollständige Suche, Konsistenzmaß.
- **Was ist die Idee von *Relief*?**  
Es wird eine Rangfolge von Merkmalen durch Zuweisung von Gewichten erstellt.
- **Unter welchen Kriterien erfolgt die Zuweisung von Merkmalen?**  
Abstand von nearest hit und nearest miss.
- **Welche Eigenschaften hat *Relief*?**  
Filter, Überwacht, Suchrichtung und Strategie sind schwer einzuordnen, Maß ist Distanzmaß.
- **Was ist die Idee von *EA-RBF*?**  
Neuronale Netze zur Klassifikation (RBF) und ein evolutionärer Algorithmus zur Auswahl geeigneter Merkmale und zur Optimierung der Architektur des RBF-Netzes.
- **Welche Aspekte bringt *EA-RBF*?**  
Schnelle Evaluation, Nebenbedingungen und Adaptive Steuerung des Evolutionsprozesses.
- **Welche Eigenschaften hat *EA-RBF*?**  
Wrapper, Überwacht, Suchrichtung ist zufällig, Suchstrategie ist heuristisch, Maß : Klassifikationsraten.

## 5 Hauptkomponentenanalyse

- **Was ist das Ziel der *Hauptkomponentenanalyse*?**  
Finden eines gleichdimensionalen Datensatzes, bei dem der Informationsgehalt in ersten Komponenten abgespeichert ist. Datenreduktion.
- **Was ist *hoher Informationsgehalt*?**  
Hohe Varianz.
- **Welche erste Transformation erfolgt bei Mustern?**  
Zentrierung.
- **Was berechnet man dann?**  
Die Kovarianzmatrix.

- **Welche Eigenschaft von Eigenvektoren nützt man?**  
Der höchste Eigenvektor zeigt in die Richtung der größten Varianz.
- **Wie transformiert man die Vektoren?**  
Man erhält eine Basis aus Eigenvektoren. Die entsprechende Matrix rotiert die Eingabemuster. Oder man wählt die Eigenvektoren mit den größten Eigenwerten.
- **Ist eine Rekonstruktion der Daten möglich?**  
Wenn man nicht alle Eigenvektoren gewählt hat, nur mit Informationsverlust.
- **Wie wird die Anzahl der Eigenvektoren bestimmt?**  
Dimensionen werden weggelassen, wenn die Eigenwerte der entsprechenden Eigenvektoren geringer sind, als der Durchschnitt der Eigenwerte. Oder absteigende Wichtigkeit.

## 6 Clusteringalgorithmen

- **Was ist der Unterschied zwischen Klassifikation und *Clustering*?**  
Klassifikation ist überwachtes Lernen bei gegebener Klassenzuordnung.
- **Wie kann man Cluster charakterisieren?**  
Homogenität innerhalb der Cluster, Heterogenität zwischen Clustern.
- **Welche Probleme gibt es beim Clustering?**  
Ausreißer vs. eigene Klasse, Klassenzahl bei vermischten Clustern, Datenskalierung.
- **Wie kann man Clusteringverfahren gliedern?**  
Unvollständige Verfahren (Projektion, Hauptkomponentenanalyse), Deterministische Clusteringverfahren (jedes Muster wird genau einem Cluster zugeordnet), Überlappende Verfahren (jedes Muster wird mindestens einem Cluster zugeordnet), Probabilistische (Wahrscheinlichkeitsbasiert), Possibilistische Verfahren (reine Fuzzy mit Zugehörigkeiten), Hierarchische (Bottom-up oder Top-down), Partitionierende (Umordnung), Mit Zielfunktion (eine zu optimierende Zielfunktion).
- **Zu welcher Klasse gehören *single-linkage*, *complete-linkage* und *average-linkage*?**  
Agglomerativ (bottom-up), hierarchisch. Verschmelzen von Klassen solange, bis ein Kriterium erfüllt ist.
- **Welches Problem entsteht dabei?**  
Grobe Lösungen werden schnell gefunden und können nicht revidiert werden. Also Ergebnis oft ein lokales Optimum.
- **Wie wird Ähnlichkeit auf Clustern definiert?**  
Kleinster Nachbar, Weitester Nachbar, durchschnittlicher Nachbar.
- **Welche Vorteile/Nachteile weist *single-linkage* auf?**  
Erkennen von Ausreißern. Aber Ketteneffekt.
- **Welche Vorteile/Nachteile weist *complete-linkage* auf?**  
Keine Ketten aber auch kein Erkennen von Ausreißern.
- **Wie kann man *average-linkage* bewerten?**  
Guter Kompromiss zwischen den beiden.
- **Wie funktioniert der *k-means* Algorithmus?**  
Man wähle initiale Zentren. Bilde Zentren der Abstandcluster und starte mit diesen neu, solange ein Abbruchkriterium nicht erfüllt ist.
- **Welche Eigenschaften hat *k-means*?**  
Terminierendes Verfahren. Aber lokale Konvergenz möglich, Zahl der Cluster muss von hand bestimmt werden, empfindlich gegenüber Ausreißern und initialen Clusterzentren.
- **Welche Fehlerfunktion wird bei *k-means* verwendet?**  
Summe der Abstände der Muster zu den Zentren.
- **Warum terminiert *k-means*?**  
Die Anzahl der Möglichkeiten den Clustern Elemente zuzuordnen ist endlich.

- **Wie kann man  $k$ -means modifizieren?**  
Hinzufügen von neuen Clustern, Wiederholung des Verfahrens mit unterschiedlichen Werten, Geschicktes wählen von Mustern: Am weitesten von zuvor gewählten.
- **Wie kann man  $k$ -means so modifizieren, dass Ausreißer toleriert werden?**  
Median statt Mittelwert.
- **Was ist ein *Fuzzy-k-means* Algorithmus?**  
Bestimmung der Zugehörigkeiten der Clustern. Multiplizieren des Zugehörigkeitsgrades hoch  $m$  bei der Addition.
- **Wie kann man die Zugehörigkeitsgrade modifizieren?**  
Folie 62.
- **Was ist die Idee bei *Nearest Neighbor Clustering*?**  
Hierarchisch agglomerativ, In allen Mustern nach einem Muster gesucht, welches am nächsten zu einem bestehenden Cluster liegt. Wenn der abstand kleiner als ein schwellwert ist, werden Cluster zusammengefaßt, sonst ein neues Cluster.
- **Was ist die Idee bei *Gaussian Mixture Models*?**  
Unter der Annahme, das die Cluster durch Gauss-Verteilungen erzeugt sind arbeitet man mit Wahrscheinlichkeiten statt Zugehörigkeitsgraden (*Expectation Maximization* zur Bestimmung der Parameter).
- **Wie evaluiert man Clustering Verfahren?**  
Bewertung durch Experten oder hinsichtlich einer Zielfunktion. Frage: Kann man aus Zugehörigkeit auf Werte schließen?

## 7 Klassifikation

- **Welche Klassifikaitonverfahren kennen Sie?**  
*1R*, *Naiver Bayes-Klassifikator*, *k-NN*, *Perzeptron*, *Lineares Ausgleichsproblem* und *Fischer-Kriterium*.
- **Wie funktioniert *1R*?**  
Für jedes Merkmal und für jeden Wert des Merkmals zähle, wie oft jede Klasse vorkommt. Weise der am häufigsten vorkommenden Klassen den Merkmalswert zu durch eine Regel. Wenn alles fertig wähle die Regeln mit kleinster Fehlerrate.
- **Welche Annahmen trifft man bei einem *naiven Bayes Klassifikator*?**  
Gleichwichtige Merkmale, Merkmale sind statistisch unabhängig.
- **Wie funktioniert ein *naiver Bayes Klassifikator*?**  
Naiver Bayes Klassifikator entspricht einem sternförmigen Bayes-Netz. Alle Eltern sind bekannt, daraus kann man die Mitte ausrechnen.
- **Wie funktioniert *k Nearest Neighbour*?**  
Man bestimmt die nächsten  $k$ -Nachbarn mit Klassenzugehörigkeiten. Man ordnet das Muster der Klasse zu, der die Mehrheit der Nachbarn zugeordnet ist.
- **Wie wählt man die Nachbarn aus?**  
Abstandsmaße.
- **Was wird gemacht, wenn Gleichstand gilt?**  
Zufällige Auswahl.
- **Was resultiert bei einem *1-NN-Netz*?**  
Voronoi Diagramm.
- **Was steht hinter linearen Modellen?**  
Eine Decicion Boundary. Lineare Trennung der Klassen.
- **Wie wird die Separierung vorgenommen?**  
Man beschreibt die Boundary mit Hilfe einer Normalform. Man adaptiert den Wert des Richtungs-Abstandsvektors solange, bis alle Muster richtig klassifiziert sind.

- **Welches Verfahren liefert das Modell?**  
Perzeptron lernen von Rosenblatt.
- **Wie kann man noch linear separieren?**  
Mit Hilfe eines Ausgleichsproblems  $Ya = t$ . Eine Lösung ist z.B.  $a = (Y^T Y)^{-1} Y^T t$ .
- **Welche Probleme gibt es?**  
Bei ungleichmässigen Klassenstärken kann die Lösung in eine der Klassen reinragen.
- **Welche Idee steckt hinter dem Fischer-Kriterium?**  
Man projiziert die Muster auf diejenige Gerade, bei der die Muster am besten separiert werden können. Das ergibt den Richtungsvektor (Normalenvektor). Danach muß man nur noch den Stützvektor wählen.
- **Welche Funktion maximiert man beim *Fischer-Kriterium*?**  
 $\frac{|\mu_1 - \mu_2|^2}{\sigma_1^2 + \sigma_2^2}$ . Dabei sind  $\mu$  und  $\sigma$  die üblichen Erwartungswerte und Varianz Bilder der  $\langle w, x(k) \rangle$ .
- **Wie wird das Problem gelöst?**  
Man kann die Fehlerfunktion explizit darstellen mit Hilfe der Intra/Interklassen-Streuungsmatrix (47-Folie). Dann gilt für die Fehlerfunktion  $\frac{w^T S_B w}{w^T S_W w}$ .
- **Was macht man, wenn die Klassen nicht linear separierbar sind?**  
Man untersucht andere Separierbarkeitsverfahren oder man transformiert Muster in einen höherdimensionalen Raum. (Z.B. mit Gauss-Funktionen).
- **Wie evaluiert man Klassifikationsverfahren?**  
Klassifikationsraten, Fehler erster/zweiter Art.

## 8 Entscheidungsbäume

- **Wie kann man Entscheidungsbäume grundsätzlich erzeugen?**  
Expertenwissen oder Automatisch.
- **Wie kann man Entscheidungsbäume induktiv erzeugen?**  
Enumerativ (bilde alle Bäume und wähle den Baum mit kleinster Anzahl von Knoten) oder Heuristisch (ergänze einen Baum um neue Entscheidungsknoten).
- **Was liefert eine andere Reihenfolge?**  
Einen anderen Baum.
- **Welche Reihenfolge sollte man wählen?**  
Informationsgain (Attribute mit größten Ausprägungen werden bevorzugt, Überanpassung), Gain-Ratio, Gini-Index ( $1 - \sum_{i=1}^d (D_i/D)^2$ ), Wähle das Atribut mit dem kleinsten  $GI_{split} = \sum_{i=1}^p D_j/D * GI(D_j)$ .
- **Wie geht man mit numerischen Attributen um?**  
Intervallbildung und Splitting-Strategie. Man berechnet den Informationsgain für alle möglichen Grenzen und wählt die Grenze mit maximaln IG.
- **Wie behandelt man Missing Values?**  
Einführung eines neuen Attributwertes, Verwendung des häufigsten Wertes (möglicherweise unter Berücksichtigung der Klassenzugehörigkeit).
- **Was versteht man unter *Pruning*?**  
Zuschneiden eines Entscheidungsbaums mit dem Ziel der Vermeidung der Überanpassung und Verständlichkeit.
- **Welche Arten von Pruning kennen Sie?**  
Post (Nach der Konstruktion) und Pre-Pruning.
- **Was versteht man unter *Windowing*?**  
Konstruktion eines Entscheidungsbaumes auf einer Teilmenge (Window) der Trainingsmuster.



- **Was kann man durch Windowing erreichen?**  
Geschwindigkeit, Vermeidung von Überanpassung, kleinere Bäume.
- **Wie verfährt man bei Windowing?**  
Aufteilung der Trainingsmenge in zwei disjunkte Teilmengen A und B. Bildung des Baums mit A und Prüfung mit B. Verschieben der falsch klassifizierten Muster von B nach A.
- **Welche Techniken kann man sonst anwenden?**  
*Regelextraktion*, Attribute mit Kosten, Multivariante Entscheidungsbäume.
- **Was kann man mit Regeln vornehmen?**  
Regel-Pruning (Prämissen streichen, solange innerhalb einer gegebenen Fehlertoleranz).
- **Wie kann man Attribute mit Kosten nutzen?**  
Kosten statt Informationsgain.
- **Was versteht man unter *multivariaten Bäumen*?**  
Linearkombinationen aus Merkmalen.

## 9 Assoziationsregeln

- **Was ist ein Problem beim Finden von Assoziationsregeln?**  
Das Dimensionsproblem.
- **Wie beschreibt man Muster?**  
Durch Itemsets.
- **Was ist *Support* eines Itemsets?**  
Anzahl der Obermengenmuster im Datensatz.
- **Was ist eine Assoziationsregel?**  
Implikation  $X \rightarrow Y$ , wobei  $X$  und  $Y$  disjunkte Itemsets sind.
- **Was ist *Support einer Assoziationsregel*?**  
Support des Itemsets  $X \cup Y$ .
- **Was ist die *Konfidenz* einer Assoziationsregel?**  
Verhältnis des Supports von  $X \rightarrow Y$  zu dem von  $X$ .
- **Warum sagt die Konfidenz einer Regel wenig über den Nutzen der Regel aus?**  
Für selten Auftretende Muster (z.B. A wurde nur einmal verkauft mit B)
- **Welche Regeln sucht man?**  
Solche, die genügend Support und große Konfidenz aufweisen.
- **Was ist ein *häufiges Itemset*?**  
Itemset, dessen Support über einem vorgegebenen Schwellwert liegt.
- **Was ist die *A-Priori-Eigenschaft*?**  
Teilmengen häufiger Itemsets sind häufig. Obermengen nicht häufiger Itemsets sind nicht häufig.
- **Wie funktioniert der *Apriori-Algorithmus*?**  
Man wähle häufige einelementige Itemsets. Generiere mehrelementige Itemsets bereits bestimmten. Man führe das aus solange, bis keine neuen Itemsets gefunden wurden.
- **Wann können *k*-elementige Itemsets häufig sein?**  
Wenn alle *k*-Teilitemsets mit  $k - 1$  Elementen häufig sind.
- **Wie berechnet man Assoziationsregel?**  
Man erzeuge Eine Regel  $Y \rightarrow X \setminus Y$ , wenn das Verhältnis  $sup(X)/sup(Y) > \alpha$  ist. Dabei geht Y über echte Teilmengen von X.
- **Bekommt man mit dem Apriori-Algorithmus interessante Regeln?**  
Nein. Man bekommt auch Trivialitäten raus.

- **Welche Erweiterungen des a-priori-Algorithmus kennen Sie?**  
Sampling, Partitioning, Count-Distribution-Algorithms, Data-Distribution-Algorithm.
- **Was ist die Idee von *Sampling*?**  
Aus einer mit Apriori bestimmten Menge der häufigen Itemsets (zufälligen Teilmenge) wird der negative Border gewählt. D.h. alle Itemsets, die nicht in PL sind, aber deren Teilmengen es sind. Nach der Vereinigung mit Border wird gesucht.
- **Welche Idee steckt hinter *Partitionierung*?**  
Teilmengen werden mit Apriori behandelt. Itemsets, die mindestens in einer Partition häufig sind, werden für die gesamtmenge Durchsucht.
- **Was ist die Idee hinter Count-Distribution Algorithms *CDA*?**  
Paralleles Verfahren zur Bestimmung von global häufigen Itemsets.

## 10 Support-Vector-Machines

- **Was ist die *VC-Dimension*?**  
Maximale Anzahl von Punkten, die von einer Menge von Klassifikatoren zerschlagen werden kann.
- **Was ist die *VC-Konfidenz*?**  
Minimale Abweichung zwischen dem empirischen und realen Risiko.
- **Wie ist die *VC-Dimension* von gerichteten Hyperebenen?**  
Dimension des Raumes plus 1.
- **Was sind die *Support-Vektoren*?**  
Punkte, die minimalen Abstand zur kanonischen Hyperebene haben.
- **Wie ist der kleinste Abstand zur Ebene?**  
Norm des Normalenvektors.
- **Wie lautet das *Optimierungsproblem*?**  
Minimiere die Norm des Normalvektors unter der Nebenbedingung  $y_i(\langle w, x_i \rangle + b) \geq 1$ .
- **Welche *Folgerung* kann man aus dem *Lagrange-Problem* ziehen?**  
Koeffizienten, mit denen die Nebenbedingungen multipliziert werden sind echt positiv genau für die Supportvektoren.
- **Was macht man, wenn der Eingaberaum nicht linear separierbar ist?**  
Man geht in einen höherdimensionalen Raum.
- **Was ist die *Kernel-Funktion*?**  
Ersatz für den Skalarprodukt im Optimierungsproblem.
- **Was muß für die *Kernelmatrix* gelten damit das *Problem* global lösbar ist?**  
Positiv definit.
- **Wie kann man neue *Kernel* aus alten konstruieren?**  
Addition, Multiplikation, Einsetzen in ein Polynom oder Exponentialfunktion.
- **Welche *Kernel-Funktionen* kennen Sie?**  
Gauss, Sigmoide, Tangens-Hyperbolicus.
- **Was versteht man unter *soft margin*?**  
Man addiert Unsicherheitsterm an die Zielfunktion (Strafterm), die aus Summen von Skalaren besteht um die die Klassenzugehörigkeit abgeschwächt wird (Schlupfvariablen).
- **Welches *Ziel* erfüllt ein *1-Class SVM*?**  
Eine Funktion finden, die positive Werte dort liefert, die die meisten Vektoren der Klasse liegen und negative sonst.
- **Wie ist die *Idee* hinter einer *1-Class SVM*?**  
Man finde die kleinste Sphäre, die alle Punkte enthält.

- **Welche Methoden kennen Sie für Mehrklassenklassifikation?**  
One against all (binäre Probleme mit Auswahl durch maximale Trennung),
- **Wie funktioniert die DAG-Mehrklassen SVM?**  
Gerichteter Graph mit Kanten, die eine Zuordnung von Klassen ermöglichen.
- **Wie kann man SVM zur Regression nutzen?**  
Man führt eine Fehlerfunktion, die Abweichungen innerhalb eines Bandes nicht bestraft.
- **Wie funktioniert *Chunking*?**  
Man wählt eine Teilmenge aus und bestimmt die Support Vektoren. Danach fügt man solche Muster hinzu, die die Klassifikation am stärksten Verletzen.
- **Was ist die *Idee* hinter Decomposition?**  
Man entfernt eine bestimmte Anzahl von Mustern fügt aber neue hinzu.

## 11 Bayes-Netze

- **Was ist ein *Bayes-Netz*?**  
Graphische Modelle für Zusammenhänge zwischen Ereignissen unter Berücksichtigung von Wahrscheinlichkeiten.
- **Was besagt die *Produktregel* über Bedingte Wahrscheinlichkeiten?**  
 $P(X, Y) = P(X|Y)P(Y) = P(Y|X)P(X)$ .
- **Was ist der Unterschied zwischen der Tabelle der Bedingtenverteilung und der gemeinsamen Verteilung?**  
In der ersten sind Zeilensummen 1. In der zweiten Gesamtsumme 1.
- **Was versteht man unter *Marginalisation*?**  
Ausrechnen von Einzelwahrscheinlichkeiten aus einer gemeinsamen Wahrscheinlichkeit.
- **Wann heißen zwei Variablen unabhängig?**  
Wenn gilt  $P(X, Y) = P(X)P(Y)$ .
- **Was besagt die Formel von Bayes?**  
 $P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)}$ .
- **Aus welchen Teilen besteht ein Bayes-Netz?**  
Aus einem gerichteten azyklischen Graphen und Tabellen mit bedingten Wahrscheinlichkeiten (Knoten sind Variablen, Kanten sind Zusammenhänge).
- **Was ist in den Knoten eines Bayes Netzes abgespeichert?**  
Wahrscheinlichkeiten des Knotens unter der Bedingung seiner Eltern.
- **Welchen Zusammenhang beschreibt bedingte Unabhängigkeit?**  
Abhängigkeit von zwei Variablen für bestimmte Zustände.  $P(X|Y, Z) = P(X|Z)$ .
- **Wie können bedingte Abhängigkeiten entstehen oder blockiert werden?**  
Durch serielle Verbindungen (mittlerer Zustand bekannt, Blockade), divergierende Verbindungen (Blockade, Elternknoten bekannt) und konvergierende Verbindungen (Entstehen, Kinderknoten bekannt).
- **Was versteht man unter dem Begriff *d-separated* bzw. *d-connected*?**  
Ab-/Unabhängigkeit zweier Knoten gegeben einer Menge von anderen Knoten.
- **Wie erhält man d-separierbarkeit?**  
Durch den moralischen Graphen. Der *moralische Graph* entsteht, wenn man alle Elternknoten eines Kindes paarweise verbindet und alle Richtungen vernachlässigt.
- **Wie erkennt man am moralischen Graphen ob zwei Knoten d-separated sind gegeben einer Knotenmenge S?**  
Wenn alle Wege zwischen den Knoten durch S führen.

- **Wie lässt sich eine gemeinsame Wahrscheinlichkeit zweier Knoten ausrechnen?**  
Über die *Kettenregel*  $P(X|pa(X_i))P(Y|pa(Y_i))$ , sodass  $pa(A)$  die Menge alle Elternknoten von  $A$  ist.
- **Was versteht man unter dem Begriff *Inferenz* in Bayschen Netzen?**  
Bestimmung von Wahrscheinlichkeiten von Knoten unter der Bedingung dass Zustände anderer Knoten bekannt sind.
- **Was versteht man unter *Evidenz*?**  
Wenn sich ein Knoten  $A$  im Zustand  $a$  befindet.
- **Was sucht man bei der Inferenz genau?**  
 $P(Q = q|E = e) = \frac{P(Q=q, E=e)}{P(E=e)}$ .
- **Welche Inferenztypen kennen Sie?**  
Diagnostische (Ursache), Vorhersagende (Wirkung), Interkausale (gemeinsames Kind), Kombinerende (Mehrere Inferenztypen)
- **Wie berechnet man die Inferenz naiv?**  
Durch die Kettenregel, indem man die Wahrscheinlichkeit aller Knotenkombinationen ausrechnet.
- **Gibt es effizientere Algorithmen?**  
Ja.
- **Wie kann man Strukturen von Bayes-Netzen lernen?**  
Indem man über mögliche Graphen geht.
- **Wie kann man Wahrscheinlichkeiten aus Daten erhalten?**  
Maximum Likelihood estimation.

## 12 Ensemble-Verfahren

- **Was ist ein *Ensamble*?**  
Kombination von redundanten Klassifikatoren mit gleicher Aufgabe.
- **Was ist ein *Modul*?**  
Aufgabenstellung wird zerlegt in Teilaufgaben. Jeder Klassifikator bearbeitet die gleiche Teilaufgabe und der Beitrag aller wird notwendig.
- **Sind die Begriffe *Ensamble* und *Modul* dual zueinander?**  
Nein. Man kann sogar beide Ansätze kombinieren.
- **Wie kann man sonst Klassifikatoren kombinieren?**  
Seriell und Parallel oder Hierarchisch.
- **Was sind Unterschiede zwischen *Non-Generative* und *Generative*?**  
Beim ersten werden Klassifikatoren nicht speziell für das Ensemble erstellt.
- **Wie kann man Entscheidungsfindung bilden?**  
Besten Klassifikator wählen, Mehrheitsentscheidung mit gleichem Gewicht oder individuelle Entscheidung.
- **Was versteht man unter *Bagging*?**  
Für jeden Klassifikator wird eine eigene Menge erzeugt. Die Entscheidung wird durch Mehrheitsentscheidung gefällt.
- **Wie wird die Menge bei *Bootstrap* ausgewählt**  
Zufällig mit Zurücklegen mit konstanter Wahrscheinlichkeit.
- **Wie geht man bei *Boosting* vor?**  
Nicht konstante Wahrscheinlichkeiten bei der Auswahl der Trainingsmenge. Schlechte Mengen bekommen häufige Wahrscheinlichkeit.

- **Wie wird beim Boosting die Auswahl getroffen?**  
Mehrheitsentscheidung. Man bezieht die Güte der Klassifikatoren mit ein.
- **Warum funktioniert Boosting?**  
Nach dem Prinzip *Weak Learner* kann man durch die Gewichteänderung ein guter Klassifikator erzeugt werden.
- **Was ist entscheidend beim Boosting?**  
Gewichtung der Trainingsmuster.
- **Was versteht man unter *Mixture of Experts*?**  
Jeder Klassifikator ist auf einer Äquivalenzklasse von Mustern ein Experte. Es wird ausgewählt durch eine gating Funktion.
- **Wie kann man das Konzept *Mixture of Experts* verallgemeinern?**  
Hierarchische Komposition.
- **Was versteht man unter *Stacked Generalization*?**  
Mehrschichtiger Aufbau. Nachfolger erhalten Eingaben der Vorgänger.
- **Warum funktionieren Ensemble-Techniken?**  
Man versucht bei gleichem Bias die Varianz zu verkleinern.

## 13 Algorithmusliste

Algorithmus	Typ/Stichwörter	Beschreibung
Focus	Merkmalsselektion, Überwacher Filter, Forwardssuche, Umfassend, Konsistenzrate	Untersuche nacheinander 1,2,3,... elementige Teilmengen, bis Inkonsistenzrate 0 erreicht ist.
Automated Branch&Bound	Merkmalsselektion, Überwacher Filter, Rückwärtssuche, Vollständige Breitensuche, Konsistenzrate	Erzeuge Teilmengen durch Weglassen jeweils eines Elements. Markiere Teilmengen mit steigender Inkonsistenzrate.
Relief	Merkmalsselektion, Überwacher Filter, Distanzmaß	Auswerten von nearest hit und nearest miss auf zufälligen Instanzen.
Single-, Complete-, Average-Linkage	Clustering, Agglomerativ	Zuweisung jedem Element einer Klasse. Zusammenfügen nach Kriterien entferntester-, nächster, mittlerer Nachbar.
c-means	Clustering, Abstände	Wähle zufällig Clusterzentren. Bestimme für jedes Muster das nächste Zentrum. Bewege Zentren zu den Clusterzentren, wenn der Fehler sinkt.
Nearest Neighbour Clustering	Clustering, Agglomerativ	Erzeugen eines neuen Clusters, falls kein Cluster bisher zuständig nach nearest neighbour, sonst hinzufügen zum Cluster.
1R	Klassifikation, Nominale Merkmale, Regel mit einem Merkmal	Zähle für jeden Merkmalswert, wie oft jede Klasse vorkommt. Wähle eine Regel die dem Merkmal die Klasse zuordnet.
Naiver Bayes Klassifikator	Klassifikation, Bayes Regel	Zuordnung der Klasse mit der höchsten Wahrscheinlichkeit
k-NN	Klassifikation	Bestimmt die Zugehörigkeit der k nächsten Nachbarn und ordnet sich die Klasse der Mehrheit
Lineares Modell	Klassifikation	Bestimmung einer Decision-Boundary. Man berechne $\langle a, y(k) \rangle$ für ein $y(k) = (1, x(k))$ und $a = (w_0, w)$ . Falls Muster falsch bestimmt wird (falsches Vorzeichen), dann addiere $\nu y(k)$ zu $a$ .
Lineares Ausgleichsproblem	Klassifikation	Lösung der Trennung mit einem Linearen Ausgleichsproblem mit Pseudoinversen
Fischer Kriterium	Klassifikation	Suche nach einer solchen Gerade durch den Ursprung an der die Muster am besten linear separierbar sind.
Pruning	Entscheidungsbäume, Post/Prepruning	Entfernen von Teilbäumen unter Vernachlässigung von Baumeigenschaften
Subtree Raising	Entscheidungsbäume	Hochheben von Teilbäumen

Apriori	Assoziationsregeln	Von einelementigen häufigen Itemsets ausgehend konstruiere grössere Itemsets durch Hinzufügen von Elementen. Solange keine neuen dazukommen. Konstruiere Regeln mit $X \rightarrow X \setminus Y$ für jede Teilmenge $Y$ für jedes häufige Itemset mit geeigneter Konfidenz.
Sampling	Assoziationsregeln	Auswahl potentiell häufiger Itemsets und Negative Border Algorithmus: Teilmengen deren alle Teilmengen im Sampleset sind.
Partitionierung	Assoziationsregeln	Auswahl einer Teilmenge im ersten Durchlauf. Im zweiten Arbeit mit gefundenen häufigen Sets in der ganzen Datenbank.
Chunking	SVM-Lernen	Wähle eine Teilmenge des Sets, Neue Teilmenge sind Support Vektoren und stark verletzende Muster.
Decomposition	SVM-Lernen	Trainieren einer Datenmenge. Löschen einr bestimmten Anzahl von Mustern und Hinzufügen von neuen.
SMO	SVM-Lernen	Auswahl kleiner Mengen und Training mit wenigen Lagrange-Multiplikatoren

# Index

- 1-Class SVM, 10
- 1R, 7
  
- A-Priori-Eigenschaft?, 9
- ABB, 5
- Apriori-Algorithmus, 9
- Attribut, 2
- Automated Branch and Bound, 5
- average-linkage, 6
  
- Bayes-Netz, 11
- Bias, 2
- Boosting, 12
- Bootstrap, 2
  
- CDA, 10
- Chunking, 11
- Clustering, 6
- complete-linkage, 6
  
- d-connectoed, 11
- d-separated, 11
- DAG, 11
- Datensatz, 2
- Directed Divergence, 4
  
- EA-RBF, 5
- empirische Streuung, 3
- empirische Varianz, 3
- Ensamble, 12
- Evaluationsmaße, 4
- Evidenz, 12
- Expectation Maximization, 7
  
- Feature, 2
- Filter, 4
- Fischer-Kriterium, 7, 8
- Focus, 5
- Fuzzy-*k*-means, 7
  
- Gaussian Mixture Models?, 7
- Generative, 12
  
- häufiges Itemset, 9
- Hauptkomponentenanalyse, 5
- hoher Informationsgehalt, 5
- Holdout, 2
  
- Idee, 11
- Inferenz, 12
- Information-Gain, 4
- inkonsistent, 4
- Inkonsistenzrate, 4
- Inkonsistenzzahl, 4
- Intervallgröße, 2
- irrelevant, 4
  
- Jackknife (Leave-One-Out), 2
  
- k* Nearest Neighbour, 7
- k*-NN, 7
- KDD, 2
- Kettenregel, 12
- Konfidenz, 9
- Korrelationskoeffizient, 3
- Kovarianz, 3
- Kreuzvalidierung, 2
  
- Lineares Ausgleichsproblem, 7
  
- Mahalanobis-Norm, 3
- Marginalisation, 11
- missing Values, 2
- Mixture of Experts, 13
- Modul, 12
- moralische Graph, 11
- multivariaten Bäumen, 9
- Muster, 2
  
- naiven Bayes Klassifikator, 7
- naiver Bayes Klassifikator, 7
- Naiver Bayes-Klassifikator, 7
- Nearest Neighbor Clustering, 7
- nominal, 2
- Non-Generative, 12
  
- ordinal, 2
  
- Partitionierung, 10
- Perzeptron, 7
- Produktregel, 11
- Pruning, 8
  
- Ratiogröße, 2
- rednundnt, 4
- Regelextraktion, 9
- Relief, 5
  
- Sampling, 10
- schwach relevant, 4
- sigle-linkage, 6
- soft margin, 10
- Stacked Generalization, 13
- Standardisierung, 3
- stark relevant, 4
- Support, 9
- Support einer Assoziationsregel, 9
  
- Varianz, 2
- VC-Dimension, 10
- VC-Konfidenz, 10
  
- Weak Learner, 13
- Windowing, 8
- Wrapper, 4