

Seminar Graphische Modelle SS 2005

1. April 2005

1 Administratives

1.1 Literatur

Im Seminar werden wir vor allem auf ein Buch von Edwards, teilweise aber auch auf ein paar andere Bücher oder Paper zurückgreifen. Im Folgenden ist eine Liste der verwendeten Werke. Der fettgedruckte Teil ist dabei jeweils ein Kürzel, mit dem später auf das Werk verwiesen wird. Die Bücher können meist bei mir (Markus Kalisch) eingesehen werden und teilweise aus der ETH-Bibliothek oder der Zentral Bibliothek ausgeliehen werden. Bei Bedarf können von meinem Exemplar Kopien angefertigt werden.

- **Edwards**: Introduction to Graphical Modelling, Second Edition (2000), David Edwards, Springer, ISBN 0-387-95054-0
- **Lauritzen**: Graphical Models, Steffen L. Lauritzen, Oxford University Press, ISBN 0-19-852219-3
- **Ripley**: Pattern Recognition and Neural Networks, Brian D. Ripley, Cambridge University Press, ISBN 0-521-46086-7

Zusätzliche Literaturhinweise werden in Einzelfällen separat gegeben.

1.2 Software

Zudem wird Computersoftware verwendet. Zu fast allen Vorträgen lassen sich Beispiele mit **MIM** berechnen und vorführen. Dies ist meist freiwillig und nur in Ausnahmen obligatorisch. In einem Vortrag (Expertensysteme) soll die Computersoftware **HUGIN** vorgestellt werden. Beide Programme sind im Internet kostenlos erhältlich und können auf dem Heimcomputer (unter Windows; Hugin auch unter Linux/Unix/Mac) installiert werden:

- **MIM**: <http://www.hypergraph.dk>, dann "Downloads", dann "MIM3202setup.exe".
- **HUGIN**: http://www.hugin.com/Products_Services/Products/Demo/Download/, dann ein paar Daten angeben (müssen ja nicht korrekt sein...), dann auf "Download" klicken.

Falls es Probleme gibt, kann eine CD mit der Software ausgeliehen werden.

2 Vortragsthemen

2.1 Grundlegende Konzepte

1. **Inhalt** Wiederholung von probabilistischen Grundkonzepten; Einführung in Terminologie, auf die in den anschliessenden Vorlesungen aufgebaut wird; Einführung in Graphen; Kurzer Überblick über Markoveigenschaften.
2. **Anregungen**

- Was ist Unabhängigkeit und bedingte Unabhängigkeit? Beispiele?
 - Was ist ein “ungerichteter” Graph? Was bedeutet “complete”, was ist eine “clique”?
 - Was sind Markoveigenschaften? Welche unterschiedlichen Markoveigenschaften werden besprochen?
 - Wie kann die Verwendung eines graphischen Modells anhand eines Beispiels erklärt werden?
 - Evtl. Überblick über Markoveigenschaften in verschiedenen Modellen
 - Evtl. Vorstellen der Software MIM
3. **Literatur** Edwards: Kapitel 1 (Preliminaries); Lauritzen: Kapitel 3.2.1 (Markov properties on undirected graphs)

2.2 Diskrete Modelle

1. **Inhalt** Anhand von Modellen mit diskreten Zufallsvariablen sollen wichtige Eigenschaften gezeigt werden. Es wird die Form solcher Modelle, das Schätzen von Parametern und der statistische Vergleich (asymptotischer χ^2 -Test) zwischen verschiedenen Modellen erläutert. Methoden der Modellwahl werden besprochen und anhand eines Beispiels wird die Software MIM präsentiert.
2. **Anregungen**
- Weshalb heissen die Modelle “diskret”?
 - Welche Form hat ein typisches diskretes Modell?
 - Wie findet man Modellparameter?
 - Wie vergleicht man die Anpassung verschiedener Modelle?
 - Was bedeutet “hierarchisches”, “graphisches” und “decomposabel” Modell?
 - Bereite eines der Beispiele in Edwards Kapitel 2 mit MIM vor. Präsentiere und interpretiere die einzelnen Schritte per Beamer (stellen wir zur Verfügung) oder Folie.
 - Wie funktioniert “forward-” und “backward-selection”?
3. **Literatur** Edwards: Kapitel 2 (diskrete Modelle); Edwards: Kapitel 5.2 (χ^2 -Test für den Vergleich von Modellen); Edwards: Kapitel 6.1.0, 6.1.1 (backward, forward selection)

2.3 Kontinuierliche Modelle

1. **Inhalt** Ähnlich wie im Vortrag 2.2 werden hier Grundkonzepte von Modellen, diesmal jedoch mit normalverteilten Zufallsvariablen (also kontinuierlich und nicht mehr diskret), besprochen. Anhand eines Beispiels werden Modellwahl und Tests zwischen Modellen besprochen. Schliesslich werden “regression models” als kontinuierliche Modelle, bei denen einige Variablen fixiert werden, besprochen.
2. **Anregungen**
- Wie sieht ein kontinuierliches (N-Verteilung) Modell aus?
 - Erkläre folgende Begriffe (evtl. auch an Beispielen): precision matrix, decomposable, graphical, hierarchical.
 - Was sind die “likelihood equations”? Wofür sind sie gut?
 - Wie schätzt man die Parameter eines Modells?
 - Wie vergleicht man verschiedene Modelle miteinander?
 - Welchen Vorteil bringt eine Einschränkung auf decomposable models?
 - Erkläre den F-Test.
 - Was ist ein “regression model”?

- Erläutere das Beispiel “Mathematic Marks” und “Bone Mineral Content” und interpretiere die Ergebnisse. Idealerweise wird die Analyse mit MIM via Beamer oder Folien gezeigt.
3. **Literatur** Edwards: Kapitel 3 (continuous models); Edwards: Kapitel 6.1.0,6.1.1 (zur Wiederholung) und 6.1.2,6.1.3 (neu - Tests unter decomposable models und F-Test); Edwards: Kapitel 5.3 (F-Test)

2.4 Gemischte Modelle

1. **Inhalt** Ähnlich wie im Vortrag 2.2 und 2.3 werden hier Grundkonzepte von Modellen, diesmal jedoch mit normalverteilten **und** diskreten Zufallsvariablen, besprochen. Es werden Modellwahl und Tests zwischen Modellen besprochen. Schliesslich wird ein Beispiel erläutert. Zudem wird die Bedeutung von “decomposable models” erläutert. Schliesslich wird das Vorgehen bei fehlenden Beobachtungen skizziert.
2. **Anregungen**
 - Wie sieht ein gemischtes Modell typischerweise aus? Was ist die CG-Verteilung? Wann ist ein Modell “homogeneous/heterogeneous”?
 - Welche Modellformeln werden in MIM verwendet? Wie werden sie interpretiert? Was ist der Zusammenhang zwischen Modellformeln und Graphen?
 - Wie können Parameter in solchen Modellen geschätzt werden? Wie vergleicht man Modelle?
 - Zeige und interpretiere ein Beispiel deiner Wahl aus dem Edwards Kapitel 4.1. Idealerweise wird die Analyse per Beamer oder Folien gezeigt.
 - Erläutere die Bedeutung von “decomposable models”.
 - Wie geht man mit fehlenden Beobachtungen um? Was ist der EM-Algorithmus? Beispiel? Idealerweise wird das Beispiel “Mathematics Marks, Revisited” mit MIM analysiert und mit Hilfe des Beamers oder Folien präsentiert.
3. **Literatur** Edwards: Kapitel 4.1 (Hierarchical Interaction Models); Edwards: Kapitel 4.4 (Decomposable Models); Edwards: 4.6 (Incomplete Data); Edwards: Appendix D2 (EM-Algorithmus)

2.5 Directed Acyclic Graphs (DAGs)

1. **Inhalt** Bisher wurden Graphen mit ungerichteten Kanten besprochen (dabei konnten die Knoten diskreten und/oder kontinuierlichen Zufallsvariablen entsprechen). In diesem Vortrag werden nun Graphen mit gerichteten Kanten (Pfeilen) und ohne gerichtete Zyklen besprochen (DAGs). Nach der Einführung in die Terminologie werden Form und Eigenschaften solcher Modelle erklärt.
2. **Anregungen**
 - Erkläre die Begriffe path, cycle, parent, child, ancestor, descendent im Zusammenhang mit DAGs.
 - Welche Form hat ein Modell, das mit DAGs beschrieben werden kann (Zusammenhang zwischen Knoten und seinen Eltern)?
 - Für diskrete Modelle: Welche Auswirkung hat der Zusammenhang zwischen Kindern und Eltern für die Grösse einer Tabelle, in der alle möglichen Kombinationen der Variablen vorkommen sollen? Zeige dies anhand eines kleinen Beispiels.
 - Beschreibe die Markoveigenschaften für DAGs. Was ist ein “moral” Graph?
 - Wie testet man Unabhängigkeit bzw. bedingte Unabhängigkeit in einem DAG (zwei Arten)?
 - Analysiere und präsentiere ein Beispiel (z.B. Edwards Kapitel 7.1.3). Idealerweise wird die Analyse und das Ergebnis via Beamer oder Folien gezeigt.

- Bei Interesse können die Markoveigenschaften anhand von Lauritzen Kapitel 3.2.2 vertieft werden.
3. **Literatur** Edwards: Kapitel 7.1 (Directed Acyclic Graphs); Lauritzen: Kapitel 3.2.2 (Markov Properties on directed acyclic graphs)

2.6 Schätzen von DAGs

1. **Inhalt** Nachdem im letzten Vortrag DAGs eingeführt wurden, wird in diesem Teil erläutert, wie man bei gegebenen Daten DAGs findet. Zunächst wird dabei skizziert, wie man bei gegebener graphischer Struktur die Parameter des Modells findet (Parameter Learning). Da die graphische Struktur des Modells allerdings in den meisten Fällen nicht bekannt sein dürfte, muss sie auch geschätzt werden (Structure Learning). Methoden, die die Struktur schätzen können in “search-and-score” und “constrained-based” Verfahren aufgeteilt werden. Aus jedem Gebiet wird ein wichtiger Vertreter vorgestellt.

2. Anregungen

- Wie können bei gegebener Struktur die Parameter des Modells bestimmt werden?
 - Wie funktioniert die search-and-score Methode? Probleme?
 - Wie funktioniert die constrained-based Methode? Beispiel?
3. **Literatur** Parameter Learning wird in Lauritzen Kapitel 4.5.1 (Recursive graphical models; ein rekursives Modell ist für unsere Zwecke nichts anderes als ein DAG) beschrieben. Dabei sind Theorem 4.36 und der Abschnitt über “Deviance Tests” zentral. Evtl. hilft es, ein Beispiel zu erläutern.

Zu Structure Learning: Search-and-Score kann man im folgenden Paper gute Hinweise finden: **Optimal Structure Identification with greedy search**, David M. Chickering, *Journal of Machine Learning Research* 3 (2002). Das Paper kann bei mir kopiert werden. Zentral ist vor allem Kapitel 4.1. Beweise können weggelassen werden, der Rest vom Paper muss nur kurz überflogen werden, um die Terminologie zu verstehen. Bei Interesse kann der Algorithmus im Kapitel 4.2 und die Diskussion im Kapitel 4.3 behandelt werden.

Zu Structure Learning: Constrained-based findet man in folgendem Buch eine gute Übersicht: **Causation, Prediction, and Search**; P. Spirtes, C. Glymour, R. Scheines; **Second Edition**; The MIT Press; ISBN 0-262-19440-6 Die Relevanten Teile können bei mir kopiert werden. Es soll Kapitel 5.4.2 vorgestellt werden (5.4.2.1 usw. müssen nicht behandelt werden). Am besten veranschaulicht man den Algorithmus anhand eines Beispiels (wie im Buch).

2.7 Expertensysteme

1. **Inhalt** In den beiden vorhergehenden Vorträgen wurden DAGs eingeführt. Zunächst wurde erklärt, was sie bedeuten. Anschliessend wurde gezeigt, wie man DAGs in der Praxis aus Daten schätzen kann. In diesem Vortrag geht es um eine beliebte und verbreitete Anwendung der DAGs: Expertensysteme. In Expertensystemen stellen Experten (z.B. Ärzte) eine Reihe von Zusammenhängen zwischen Variablen (z.B. Blutdruck, Herzinfarkt, Brustschmerzen, ...) zusammen und Beobachtungen werden gemacht (“knowledge base”). Diese “Wissensdatenbank” wird anschliessend verwendet, um in einem neuen Fall anhand von einer beobachteten Variable (z.B. Brustschmerz und Blutdruck) auf unbeobachtete Variablen zu schliessen (z.B. Herzinfarkt). Die Technik, die diese Aussagen ermöglicht wird als “inference engine” bezeichnet. Die Konzepte werden erläutert und anhand eines Beispiels mit Hilfe von Software veranschaulicht.

2. Anregungen

- Wozu ist ein Expertensystem gut?
- Was versteht man unter “Knowledge Base” und “Inference Engine”?

- Warum ist es sinnvoll, in dem Anwendungsgebiet DAGs zu verwenden?
 - Was ist ein “join tree”? Wie bildet man ihn aus einem DAG?
 - Was ist die “potential representation”?
 - Wofür ist “message passing” gut? Wie funktioniert es?
 - Illustriere ein Beispiel (z.B. das medizinische Beispiel) mit Hilfe von HUGIN.
 - Veranschauliche verschiedene typische Fragestellungen, die mit HUGIN und einem gegebenen Expertensystem einfach beantwortet werden können
 - Am besten wird wohl entweder das medizinische Beispiel aus dem Buch oder aber ein vorgegebenes im HUGIN Tutorial besprochen.
3. **Literatur** Eine gründliche Einführung in das Thema kann in Ripley: Kapitel 8.1 und 8.2 gefunden werden. Für das konkrete Verständnis ist wohl allerdings der Teil “Calculations on moral graphs - An example” in Kapitel 8.2 am sinnvollsten. In jedem Fall gebe ich eine vereinfachte und für unsere Zwecke völlig ausreichende Zusammenfassung von diesen Kapiteln aus. In dieser Zusammenfassung ist nur Kapitel 1 und Kapitel 4-6 von Bedeutung.

Zur Verwendung von HUGIN: HUGIN sollte ohne Probleme auf einem Windows Computer installiert werden können (s.o.). Einführungen in verschiedene Themen findet man unter <http://developer.hugin-project.org/>. Besonders hilfreich (und ausreichend) ist dabei das Tutorial “Building Bayesian Network”. Verwende für die Vorführung entweder das dort gezeigte Beispiel oder das medizinische Beispiel aus dem Skript.

2.8 Chain Graphs

1. **Inhalt** Nachdem in den bisherigen Vorträgen zunächst ungerichtete Graphen und dann gerichtete Graphen (DAGs) besprochen wurden, wenden wir uns nun Graphen zu, die sowohl gerichtete als auch ungerichtete Kanten enthalten können. Dies wird besonders dann nützlich, wenn den einzelnen Variablen keine sinnvolle Reihenfolge zugeordnet werden kann, wohl aber Gruppen von Variablen.
2. **Anregungen**
 - Was ist der Unterschied zwischen DAGs und Chain Graphs?
 - Welcher Zusammenhang besteht zwischen Chain Graphs auf der einen Seite und DAGs bzw. ungerichtete Graphen auf der anderen Seite?
 - Was ist eine “Chain Component”?
 - Was sind Markov Eigenschaften von Chain Graphs?
 - Erläutere ein Beispiel (Edwards: Kapitel 7.2.3).
3. **Literatur** Edwards: Kapitel 7.2; Lauritzen: Kapitel 3.2.3

2.9 Kausalität

1. **Inhalt** Durch Beobachtung einer Assoziation kann i.A. noch kein kausaler Zusammenhang hergestellt werden. Z.B. kann mitunter (tragischer Weise) beobachtet werden, dass exzessives Rauchen zu Lungenkrebs führt. Allerdings wäre es falsch zu behaupten, dass diese Beobachtung einen Kausalzusammenhang beweisen würde. Es könnte z.B. auch so sein, dass eine genetische Disposition sowohl die Lust am Rauchen als auch die Neigung zur Entwicklung von Lungenkrebs fördert und dann wäre Rauchen keine Ursache von Krebs mehr. In diesem Vortrag geht es darum, mit Hilfe von DAGs zu analysieren, ob, und wenn ja wann, es möglich ist, nicht nur auf Assoziationen sondern tatsächlich auf kausale Zusammenhänge schließen zu können. Dies wird anhand von Pearl’s Causal Graphs erklärt. Es wird mit Hilfe des back-door- und front-door-Kriteriums erläutert, wann ein Kausalzusammenhang identifizierbar ist.

2. Anregungen

- Was versteht man unter “intervention”?
- Was ist ein “confounding effect”?
- Inwiefern ist “completeness” für kausale Modelle wichtig?
- Was ist die “causal graph assumption”?
- Wie ändert sich eine gemeinsame Verteilung durch Intervention?
- Unter welchen Umständen kann ein Kausalitätszusammenhang gefunden werden?
- Was ist das “back-door criterion”? Wofür ist es nützlich?
- Was ist das “front-door criterion”? Wofür ist es nützlich?

3. **Literatur** Edwards: Kapitel 8.3 (Pearl’s Causal Graphs); evtl. Edwards: Kapitel 8.4.2; bei Interesse auch S. L. Lauritzen, *Causal inference from graphical models*, In Barndorff-Nielsen, O.E., Cox, D. R. and Klüppelberg, C. (eds.) *Complex Stochastic Systems*. Chapman and Hall/CRC. London, Boca Raton. pp. 63-107, 2001.