Probabilistic Graphical Models

Guide to AI and Data Courses · Part II: Branches of AI

Bayesian networks | Markov networks | Conditional independence | Inference | Parameter learning | Neighboring courses

Sercan KÜLCÜ · Giresun University · Computer Engineering

© Sercan Külcü. All rights reserved. · Interactive HTML version of these slides:
https://sercankulcu.github.io/files/ai_courses/slides/Chapter_12_Probabilistic_Graphical_Models.html

What Question Does It Ask?

> **How do we model and query many interdependent, uncertain variables?**

Probabilistic graphical models represent the joint distribution of many uncertain variables with a **graph**: nodes are variables and edges are direct dependencies. The conditional independences encoded by the graph break a huge distribution into small pieces, making it possible both to model and to query it.

At a glance

APPROACH	INPUT	OUTPUT
probability and graphs combined	variables, dependencies, data	conditional probability, best explanation

SUCCESS MEASURE	MATH LOAD	DATA NEEDS
likelihood, inference accuracy	high: probability	moderate

A Short History

1906
Markov studies chains in which the next state depends only on the current one.

1966
Baum and Petrie publish the statistical basis of the hidden Markov model (HMM).

1977
Dempster, Laird and Rubin publish the EM algorithm for learning with missing data.

1988
Pearl's "Probabilistic Reasoning in Intelligent Systems" brings Bayesian networks into AI.

1989
Rabiner's HMM tutorial becomes the standard reference in speech recognition.

2001
Lafferty, McCallum and Pereira propose conditional random fields (CRFs) for sequence labeling.

THE COURSE

Core Concepts

Joint distribution
Probability of every combination of values of all variables.

Conditional independence
Once one variable is known, the others no longer affect each other.

Bayesian network
Directed acyclic graph; a conditional probability table at each node.

Markov network
Undirected graph; relations given by potential functions.

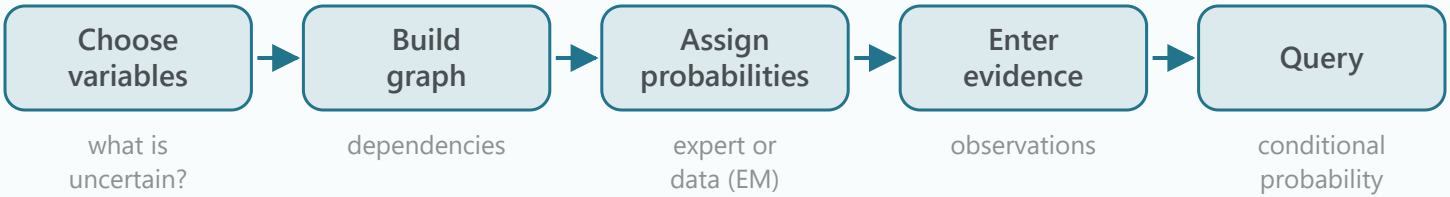
Hidden Markov model
An unobserved chain of states, each emitting an observation.

Exact inference
Variable elimination, junction tree.

Approximate inference
Sampling (MCMC), loopy belief propagation, variational.

Learning
Parameters (EM) and structure: finding the graph itself from data.

A Typical Workflow



Example: A diagnostic network has nodes for smoking, cancer and cough. A physician fills the tables from the literature. If a patient coughs and smokes, these two pieces of evidence are entered; the network recomputes the probability of cancer and shows which test would reduce uncertainty the most.

THE COURSE

What Does the Course Cover?

1. Probability review; joint distributions, factorization
2. Bayesian networks and d-separation
3. Markov networks and factor graphs
4. Exact inference: variable elimination
5. Belief propagation and the junction tree
6. Inference by sampling: Gibbs, MCMC
7. Variational inference
8. Hidden Markov models: forward-backward, Viterbi
9. Conditional random fields and sequence labelling
10. Parameter learning: maximum likelihood, Bayesian
11. Learning with missing data: the EM algorithm
12. Structure learning and a first look at causality

Skills you will gain

building Bayesian networks with pgmpy

sequence modelling with hmmlearn

Bayesian models in PyMC or Stan

reading conditional independence off a graph

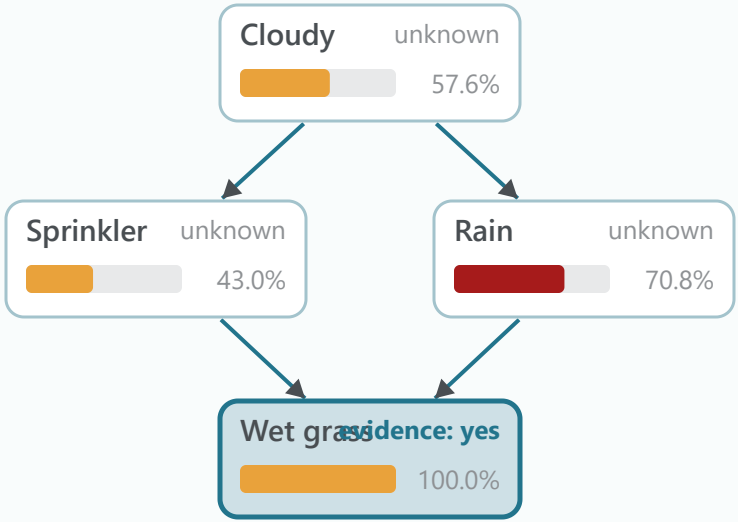
interpreting inference results

A typical semester; the order and coverage vary between universities.

TRY IT

Why Is the Grass Wet?

The grass is wet in the morning: did it rain, or did the sprinkler run? Click a node in the four-variable Bayesian network to set evidence (yes / no / unknown); the other nodes are updated by exact probability calculation.



Clear evidence | Grass wet | Grass wet + sprinkler on

Evidence: Wet grass = yes
P(Rain | evidence) = **70.8%**
P(Sprinkler | evidence) = 43.0%
Summed 8 of 16 states consistent with it · P(evidence) = 0.647

Arrows point along causes, but inference flows both ways: back from wet grass to rain and clouds. First make the grass wet, then turn the sprinkler on.

The 16 joint states of the four variables are not stored as a table; the network keeps only **local conditional probabilities**. When evidence arrives, enumeration computes the requested posterior from these small tables.

CONTEXT

Prerequisites

Courses worth knowing before this one, and why they are needed.

Course	Why it is needed
Probability and Statistics	conditional independence, joint distribution
Algorithms	graph traversal, dynamic programming (Viterbi, forward-backward)
Data Structures	graph representation, trees and priority queues

CONTEXT

Are You Ready for This Course?

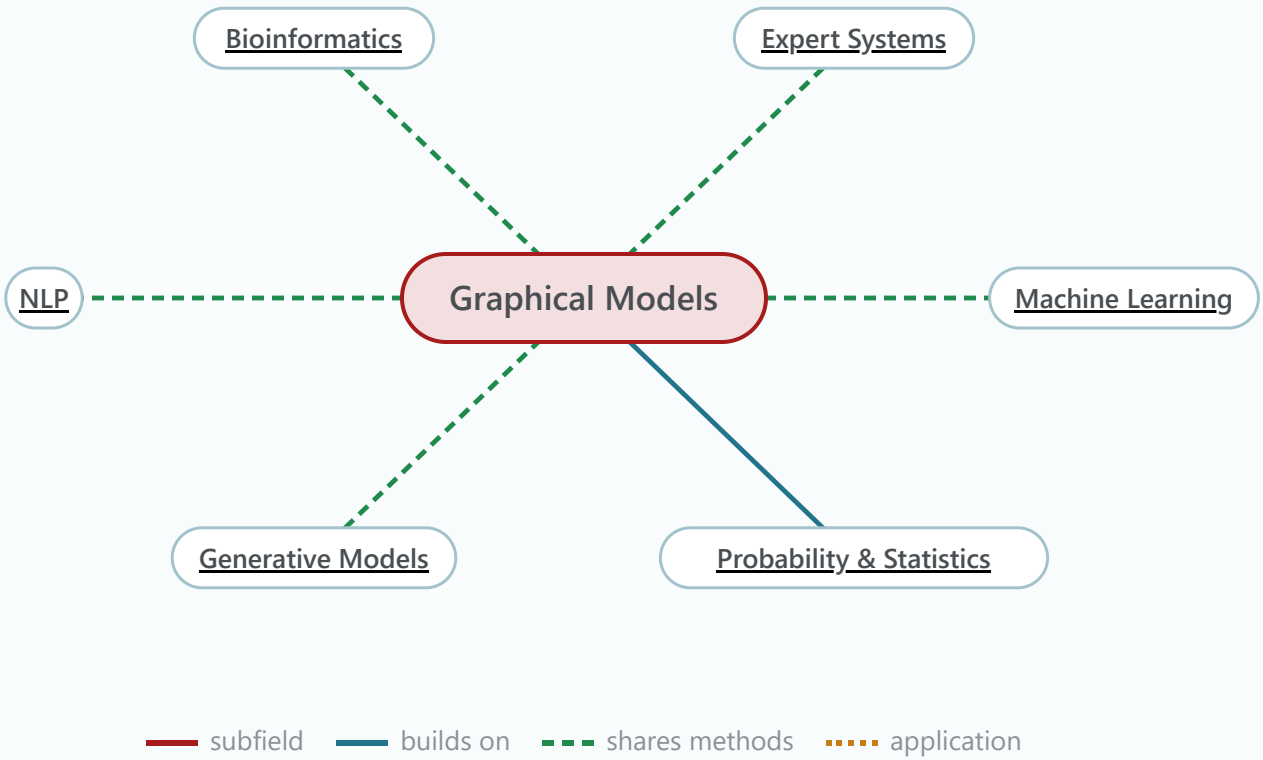
Ask yourself; for every "no", the right-hand column shows where to start.

Ask yourself	If not, start with
Can you compute a conditional probability with Bayes' rule?	Probability and Statistics
Can you store a graph as an adjacency list and traverse it?	Graph representations from a data structures course
Can you solve a problem by splitting it into subproblems with dynamic programming?	Dynamic programming from an algorithms course
Can you explain maximum likelihood estimation with an example?	Machine Learning

If you can say "yes" to most of these, you are ready; one or two gaps can be closed during the semester.

CONTEXT

Neighbouring Courses



Overlapping Courses

Relation	Course	Shared ground
builds on	This course builds on Probability and Statistics	conditional independence, joint distribution
shares methods	Shares methods with Machine Learning	probabilistic classifiers, EM algorithm
shares methods	Shares methods with Natural Language Processing and Text Mining	hidden Markov model, conditional random field
shares methods	Shares methods with Expert Systems and Knowledge Representation	inference under uncertainty
shares methods	Shares methods with Generative Models	latent variable models, variational inference
shares methods	Shares methods with Bioinformatics	hidden Markov models for sequence alignment

DIFFERENCES

Where It Differs (1/2)

Criterion	Graphical Models	Probability & Statistics	Machine Learning
Core question	How do we model and query many interdependent, uncertain variables?	How is uncertainty measured, and how do we draw reliable conclusions from limited data?	How does a program generalize from examples?
Approach	probability and graphs combined	probability models and inference	statistical learning
Input	variables, dependencies, data	sample data	labeled or unlabeled examples
Output	conditional probability, best explanation	estimate, confidence interval, test result	a model that makes predictions
Success measure	likelihood, inference accuracy	unbiasedness, power, significance	error on unseen data
Math load	high: probability	the theory itself	high: linear algebra, probability, optimization
Data needs	moderate	a sample	moderate to large

- **Probability & Statistics:** Probability and statistics studies distributions of a few variables; this course teaches modeling hundreds of variables by factoring them with a graph.
- **Machine Learning:** ML usually predicts a single target; a PGM builds the joint distribution of all variables and lets you query any of them.

Where It Differs (2/2)

Criterion	Graphical Models	Generative Models	Expert Systems
Core question	How do we model and query many interdependent, uncertain variables?	How does a model generate new samples that look like real data?	How is an expert's knowledge written down and used for inference?
Approach	probability and graphs combined	learning the data distribution	symbolic logic and rules
Input	variables, dependencies, data	huge, mostly unlabeled data	facts and expert rules
Output	conditional probability, best explanation	new samples: images, text, audio	conclusion with justification
Success measure	likelihood, inference accuracy	realism and diversity (FID etc.)	expert agreement, explainability
Math load	high: probability	high	low: propositional and predicate logic
Data needs	moderate	very large	little; a knowledge base

- **Generative Models:** Generative models use large neural networks to produce samples; this course focuses on explicit, interpretable structure and exact inference.
- **Expert Systems:** Expert systems write knowledge as crisp rules; a PGM writes the same knowledge with probabilities and combines evidence consistently.

PRACTICE

Applications

Medical diagnosis
Disease probability from symptoms and tests.

Speech recognition
Word sequences from audio with HMMs.

Sequence labeling
Part-of-speech and named entity tagging (CRF).

Bioinformatics
Gene finding and sequence alignment.

Fault diagnosis
Most likely fault from observed symptoms.

Image segmentation
Label agreement of neighboring pixels (MRF).

PRACTICE

Risks and Ethics

Tables filled with guesses

Without data, probability tables are filled with expert guesses; a precise-looking 83 per cent hides how uncertain those guesses are.

False independence assumptions

Treating two dependent pieces of evidence as independent counts the same information twice and makes the model overconfident.

Probability errors in court

Mistaking the chance of the evidence for the chance of guilt (the prosecutor's fallacy) and ignoring base rates can convict the innocent.

Unconverged sampling

Stop an MCMC chain too early and the posterior is wrong; results cannot be trusted without reported convergence diagnostics.

In depth: Explainable AI and AI Ethics.

PRACTICE

Tools and Datasets

Tools and libraries

pgmpy | PyMC | Stan | hmmlearn | Pyro | GeNIe and SMILE | bnlearn (R)

Classic networks and datasets

Asia (chest clinic) network | ALARM network | Rain-sprinkler network | Insurance network

CoNLL-2003 | TIMIT

PRACTICE

Common Misconceptions

Common belief	In fact
An edge means cause and effect.	In a Bayesian network an edge is only a probabilistic dependency; causality needs extra assumptions or experiments.
Unconnected variables are independent.	Independence is conditional; two nodes can become dependent once a common child is observed.
Inference is always easy to compute.	Exact inference on general graphs is NP-hard; large models need approximate methods.
Deep learning made this field obsolete.	PGMs remain strong where data is scarce and explainability or uncertainty estimates matter; the two are also combined.

NEXT STEPS

After This Course

Courses that build on or extend this one

This course is mostly an end point: it builds on the others.

Related careers

Machine learning researcher | Data scientist | Biostatistician | Speech and language technology engineer

Probabilistic graphical models bring probability and structure together. Natural next steps are generative models with latent variables, natural language processing and bioinformatics with sequence models, or causal inference.

NEXT STEPS

Further Reading

- Koller and Friedman, *Probabilistic Graphical Models: Principles and Techniques*, MIT Press, 2009.
- Pearl, *Probabilistic Reasoning in Intelligent Systems*, Morgan Kaufmann, 1988.
- Murphy, *Machine Learning: A Probabilistic Perspective*, MIT Press, 2012.
- Darwiche, *Modeling and Reasoning with Bayesian Networks*, Cambridge University Press, 2009.
- Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", *Proceedings of the IEEE* 77(2), 1989.

The End

Chapter 12: Probabilistic Graphical Models

← Previous chapter · Map of the field · Next chapter →