'08 PGMS (L1) Intro 23/04/2020

· You already have notes; use this space to record instructor assistance/exposition

- Koller + Friedman → intense
- Jordan - An Introduction - easier

(R) Supplement with materials e.g. papers and tutorials

· List grading on GitHub → this will give you a sense of how much you are missing out
· 4 HWs are not trivial!

· IID data e.g. $X_1, X_2, \dots, X_n \sim P$ (usual ML setting)
· Ex: How graphs can uniquely and consistently specify a model M_G, under what conditions we can estimate model parametrization and topology

e.g. NN as a graphical model

· Ex: PGM - define probability distribution over data with complex structure
  - in distinction against rule-based inference; reasoning under uncertainty,
  - ~~ness~~ noise
  - PGMs give systematic methods for reasoning under uncertainty

Ex: Research questions
  - See slides

Representation - Articulation in common language (mathematics, algorithms)

Inference - Prediction/estimation
  - may be able to ask valid questions that are intractable to answer (NP-hard)

Learning - Combine model + data in some kind of (function ⇒) score that encodes optimality over all possible models specifications

· Example: Multiple representations of same data (trees)

Ex: Can we mathematically express/quantify what makes a 'good' representation → e.g. via some 'distance metric'

- There may exist hidden nodes; allow for placeholders
- PGM: Think through lens representation - inference - learning
  - observation data - hidden states - structures - probabilities

---

- Difference between "models" and "graphical models" representation
- (1): we can write down a joint distribution of a collection of random variables (assuming independence) (on binary r.v.s)
  - via a probability table (remember Wasserman)
  - $2^8 - 1$ state configurations for 8 binary r.v.s.
- (2): memory issues as no. of r.v.s ↑; computer scientist would balk at using a large joint distrib. table

EX: In any case, there may be r.v.s we cannot observe if too many

Inference: e.g. $p(H|A)$

EX: All questions (using enumeration) is NP-hard
- Probability distributions with no structure other than an enumerative table is likely not going to be helpful.
- e.g. 1000 stocks on NASDAQ for a portfolio
  - Structure: → correlation via sector? and dependencies
- EX: How can we make use of domain knowledge/structure to make models more economical than enumerative probability tables.

Graphical Models

- Molecular biology
- PGM: Structure simplifies representation
  e.g. via physical location/communicative pathways (dependencies amongst variables)

PGM: instead of enumeration; think about traversal (factorisation law)
- Given a graph, traverse it, whenever you run into a node; write down ~~conditionally~~ ~~ind~~ distribution of r.v.s given their parents;

— no ~~pa~~ parents → marginal probability

· multiply together

· currently we assume this is feasible (prove this later)

· Rewriting joint as factorisation; more parsimonious representation of probability + dependencies

(W): (A) - check calculation - Benefits of PGM

1) Handle large multivariate distri using graph structure to factorise the distribution (representation cost)

- Formally; using conditional independence (next)

= data integ

· Each term is self-contained, local conditional distri

· In context of biology → allows for parallelism/data integration over biological labs; each lab only works with relevant LCD.

· use PGM to combine LCD at each "modality"

(W): possibilities for combining diverse, heterogeneous data sources in a modular fashion

=

statistical inference

- use priors to confine search for distribution of earth surface temperature (domain knowledge)

(e.g. not −273°C)

- via Bayes Theorem: allows inference; placeholder for injection of prior knowledge

- PGM - Hidden parameters, observed data

prior knowledge over hidden parameters/r.v.s.

- universal way of representing structure of knowledge/mathematical algorithms ~~for~~

Ex: Also lots of downsides; PGM

- PGM is a particular mode of inference (not really ~~probab~~ 'model')

EX: - simplify exponentially-large probability distri without associated costs
- and endow with structured semantics

Formal description: A family of distri on a set of r.v.s. compatible with all probabilistic independence propositions encoded with a graph that connects variables

- emphasis on allowing/enabling scientific communication

(E1) - 2 GMs:

1) directed edges: causality rel. (Bayesian networks / Directed Graphical Models)

2) undirected edges: correlations (Markov Random Field / ..)

- Bayesian networks: $\boxed{53:20}$
- conditional independence of yellow x of red, conditional on green.
- social network interpretation
parents, children, co-parents

(P2) Be clear on terminology

(P1): $P(X|Y, \ldots) = P(X|Y)$

(P3): Be clear on distinction of c.i in BN/MRFs

MRFs

- conditional independence

- Given graph; use topology to extract conditional independence relations
- some formalism is required mathematically between conditional independence relations - topological representation
- EX: 2 ways of specifying distri:-
i) Identify independence exhaustively via graph traversal algorithm; write down distri that satisfies via testing proce.

ii) use factorisation; superimpose graphs on top of r.vs.; use graph factorisation rules and multiply.

EX: Are i) and ii) the same? (there are proofs in Koller + Friedman)

- (W) (A4): Equivalence theorem → get to the point

- (OT): EX: Formalises ML/stats in terms of graphs

EX: Allows contextualisation of many algorithms; PGM allows explicit consideration of topology

---

- DNA of PGMs:-
  - 1920s - Wright
  - 1980s - Spiegelhalter (stats), Lauritzen, Judea Pearl (CS)

- many slides are in Appendix; a lot of slides, don't cover all; understand, distill key principles