Lecture 21. PGM Representation II

COMP90051 Statistical Machine Learning

Lecturer: Feng Liu

THE UNIVERSITY OF MELBOURNE

This lecture

- Undirected PGMs and conversion from D-PGMs
  - Undirected PGM formulation
  - Directed to undirected
  - Why U-PGM
- Example PGMs, applications
  - HMMs (and Kalman Filter)
  - Applications

Undirected PGMs

Undirected variant of PGM, parameterised by arbitrary positive valued functions of the variables, and global normalisation.

A.k.a. Markov Random Field.

Undirected vs directed

Undirected PGM

- Graph
  * Edges undirected
- Probability
  * Each node a r.v.
  * Each clique C has "factor" $\psi_C\left(X_j : j \in C\right) \geq 0$
  * Joint $\propto$ product of factors

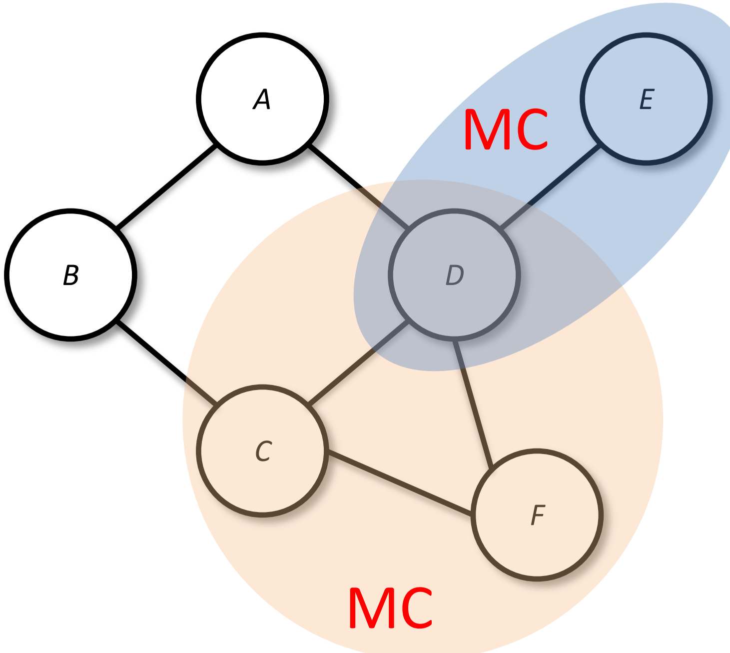
Directed PGM

- Graph
  * Edged directed
- Probability
  * Each node a r.v.
  * Each node has conditional $p\left(X_i | X_j \in parents(X_i)\right)$
  * Joint $=$ product of cond'ls

Key difference = normalisation

Undirected PGM formulation

- Based on notion of
  - *Clique: a set of fully connected nodes (e.g., A-D, C-D, C-D-F)*
  - *Maximal clique: largest cliques in graph (not C-D, due to C-D-F)*



- Joint probability defined as

$$P(a,b,c,d,e,f) = \frac{1}{Z}\psi_1(a,b)\psi_2(b,c)\psi_3(a,d)\psi_4(d,c,f)\psi_5(d,e)$$

  - where each ψ is a positive function and Z is the normalising **'partition' function**

$$Z = \sum_{a,b,c,d,e,f} \psi_1(a,b)\psi_2(b,c)\psi_3(a,d)\psi_4(d,c,f)\psi_5(d,e)$$

Directed to undirected

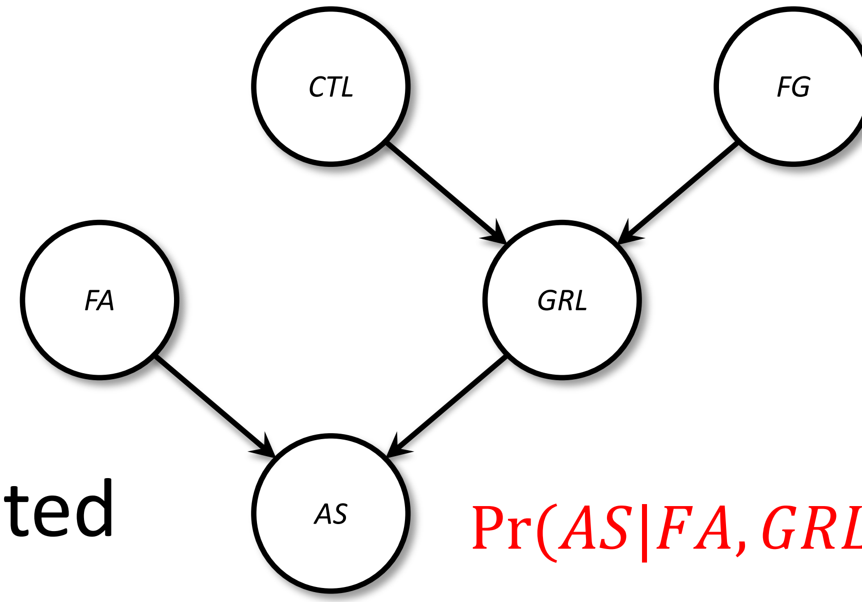
- Directed PGM formulated as

$$P(X_1, X_2, \ldots, X_k) = \prod_{i=1}^{k} Pr(X_i | X_{\pi_i})$$

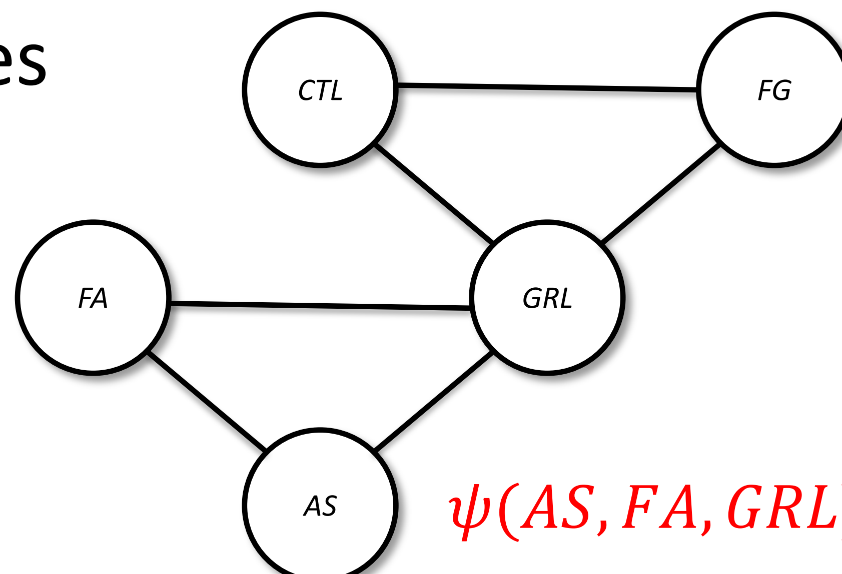
where **π** indexes parents.

- Equivalent to U-PGM with
  - each conditional probability term is included in one factor function, ψ_c
  - clique structure links *groups of variables,* i.e., $\{\{X_i\} \cup X_{\pi_i}, \forall i\}$
  - normalisation term trivial, Z = 1

1. copy nodes
2. copy edges, undirected
3. 'moralise' parent nodes



$\Pr(AS|FA, GRL)$



$\psi(AS, FA, GRL)$

Why U-PGM?

- Pros
  - generalisation of D-PGM
  - simpler means of modelling without the need for per-factor normalisation
  - general inference algorithms use U-PGM representation (supporting both types of PGM)
- Cons
  - (slightly) weaker independence
  - calculating global normalisation term (Z) intractable in general (but tractable for chains/trees, e.g., CRFs)

Mini Summary

Undirected probabilistic graphical models (U-PGMs)

- Definition
- Conversion to D-PGMs
- Pros/Cons over D-PGMs

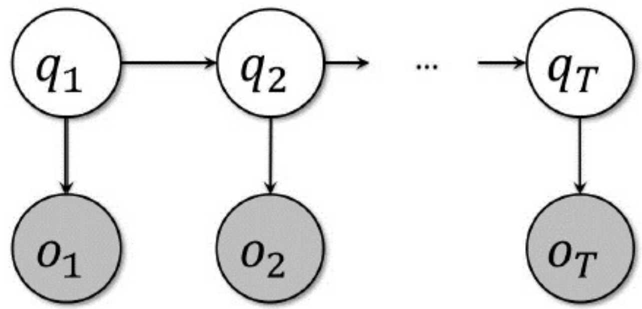
Next: Examples and applications of PGMs

Example PGMs

The hidden Markov model (HMM);
lattice Markov random field (MRF);
Conditional random field (CRF)

The HMM (and Kalman Filter)

- Sequential observed outputs from hidden state



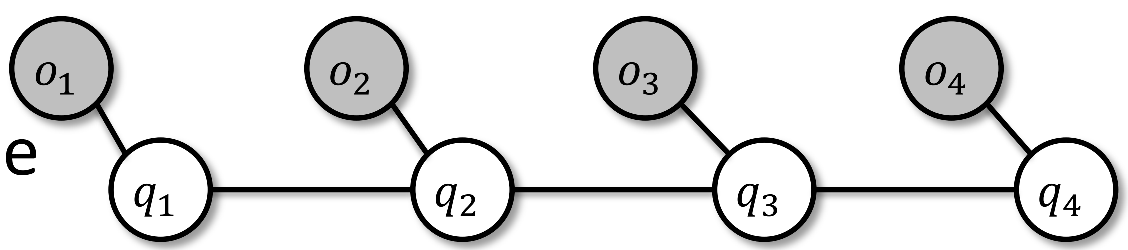
$A = \{a_{ij}\}$ transition probability matrix; $\forall i : \sum_j a_{ij} = 1$

$B = \{b_i(o_k)\}$ output probability matrix; $\forall i : \sum_k b_i(o_k) = 1$

$\Pi = \{\pi_i\}$ the initial state distribution; $\sum_i \pi_i = 1$

- The Kalman filter same with continuous Gaussian r.v.'s

- A CRF is the undirected analogue

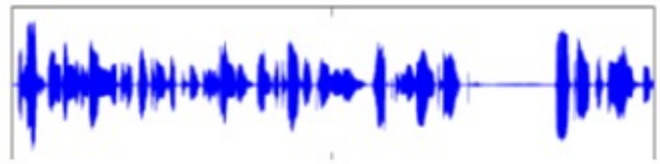


HMM Applications

- NLP – part of speech tagging: given words in sentence, infer hidden parts of speech

  "I love Machine Learning" → noun, verb, noun, noun

- Speech recognition: given waveform, determine phonemes



- Biological sequences: classification, search, alignment
- Computer vision: identify who's walking in video, tracking

Fundamental HMM Tasks

HMM Task	PGM Task
Evaluation. Given an HMM μ and observation sequence O, determine likelihood $\Pr(O\|\mu)$	Probabilistic inference
Decoding. Given an HMM μ and observation sequence O, determine most probable hidden state sequence Q	MAP point estimate
Learning. Given an observation sequence O and set of states, learn parameters A, B, Π	Statistical inference

Pixel labelling tasks in Computer Vision

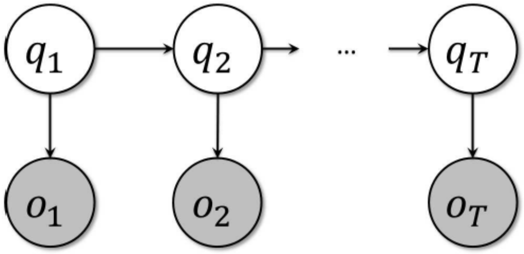
Semantic labelling (Gould et al. 09)

Interactive figure-ground segmentation (Boykov & Jolly 2011)

Denoising (Felzenszwalb & Huttenlocher 04)

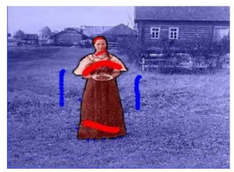
What these tasks have in common

- Hidden state representing semantics of image
  - Semantic labelling: Cow vs. tree vs. grass vs. sky vs. house
  - Fore-back segment: Figure vs. ground
  - Denoising: Clean pixels
- Pixels of image
  - What we observe of hidden state
- Remind you of HMMs?

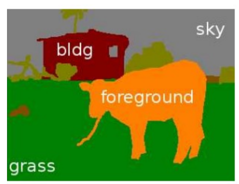


A hidden square-lattice Markov random field

- Hidden states: square-lattice model
  - Boolean for two-class states

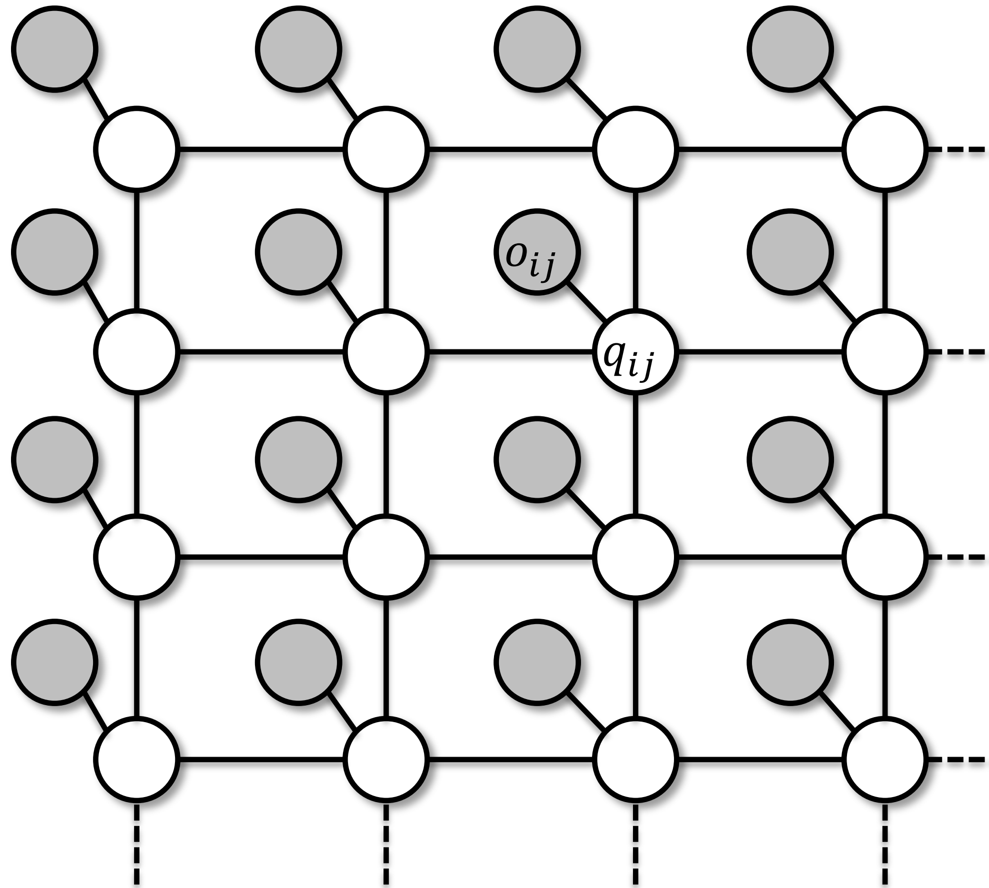


  - Discrete for multi-class



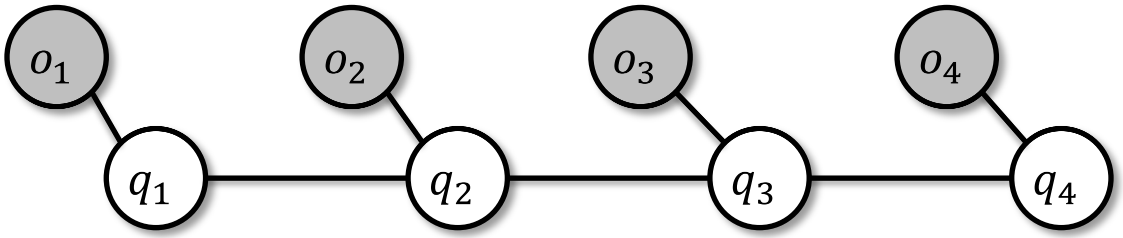
  - Continuous for denoising

- Pixels: observed outputs
  - Continuous e.g. Normal



Application to sequences: CRFs

- Conditional Random Field: Same model applied to sequences
  - observed outputs are words, speech, amino acids etc
  - states are tags: part-of-speech, phone, alignment…
- CRFs are discriminative, model *P(Q|O)*
  - versus HMM's which are generative, *P(Q,O)*
  - undirected PGM more general and expressive



Summary

- Probabilistic graphical models
  - Motivation: applications, unifies algorithms
  - Motivation: ideal tool for Bayesians
  - Independence lowers computational/model complexity
  - PGMs: compact representation of factorised joints
  - U-PGMs
- Example PGMs and applications

Next time: elimination for probabilistic inference