Docent Lecture

Independence in Probabilities and Undirected Graphs

Wray Buntine

Target audience: graduate students, have done a basic CS graph theory course, basic probabilities, and had an introductory motivational lecture on probabilistic graphical models.

Overview

- **Independence**
- Problem Decomposition
- Undirected Graphs

Background

- Independence is a way of

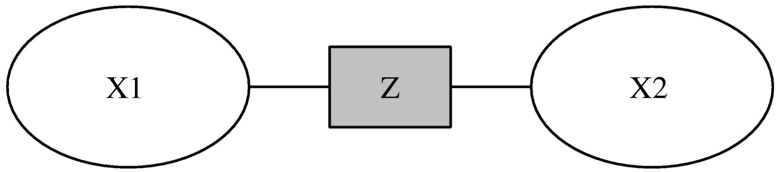
$$\left\{\begin{array}{c}\text{reducing}\\ \text{simplifying}\end{array}\right\}\left\{\begin{array}{c}\text{effort}\\ \text{complexity}\\ \text{cost}\end{array}\right\}\text{in}\left\{\begin{array}{c}\text{inference}\\ \text{learning}\\ \text{elicitation}\\ \text{optimization}\end{array}\right\}.$$

- Independence between variables arises naturally with *causal* and *generative* models.
- Major schools developed are:
  - in statistics (*e.g.*, Stephen Lauritizen, now at Cambridge), and
  - in artificial intelligence (*e.g.*, Judea Pearl, UCLA).

Definition of Independence

X_1 *independent* of X_2 *given* Z, defined by

$$p(X_1 \mid X_2, Z) = p(X_1 \mid Z) \text{ whenever } p(X_2, Z) > 0$$

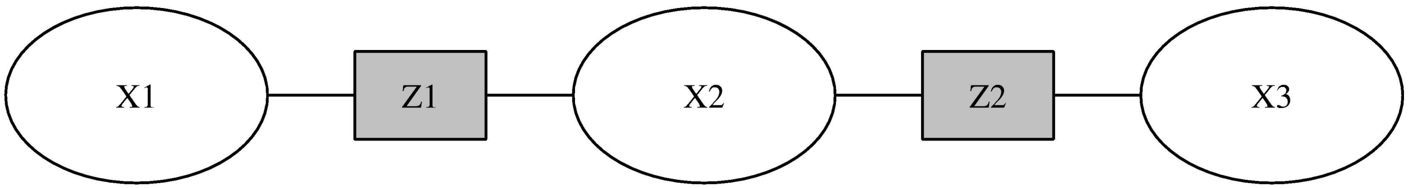


- Notationally, $X_1 \perp\!\!\!\perp X_2 \mid Z$.
- Z is the *conditioning* or *separating* set.
- Following are equivalent:

$$p(X_1, X_2 \mid Z) = p(X_1 \mid Z) p(X_2 \mid Z) \text{ whenever } p(Z) > 0$$
$$p(X_2 \mid X_1, Z) = p(X_2 \mid Z) \text{ whenever } p(X_1, Z) > 0$$
$$p(X_1, X_2, Z) = f(X_1, Z) g(X_2, Z) \text{ for non-negative functions } f(\cdot), g(\cdot)$$

- Is symmetric in X_1 and X_2.
- For consistency $X_1 \cap X_2 \subseteq Z$,
  e.g. $\{a\} \perp\!\!\!\perp \{a,b\} | \{c,d\}$ inconsistent with a on both sides

3-way Independence



Lemma. For all possible pair-wise decompositions on the graph above ($X_1 \perp\!\!\!\perp X_2 \,|\, Z_1$; $X_2 \perp\!\!\!\perp X_3 \,|\, Z_2$; $(X_1 \cup Z_1 \cup X_2) \perp\!\!\!\perp X_3 \,|\, Z_2$; $X_1 \perp\!\!\!\perp (X_2 \cup Z_2 \cup X_3) \,|\, Z_1$; $X_1 \perp\!\!\!\perp X_3 \,|\, Z_1 \cup X_2 \cup Z_2$) to be consistent independent statements, it is necessary and sufficient that:

$X_1 \cap X_2 \subseteq Z_1$; $X_2 \cap X_3 \subseteq Z_2$ (separating set property)

$X_1 \cap X_3 \subseteq X_2$ (running intersection property)

NB. This result generalises to independence with arbitrary undirected graphs (where nodes and arcs are both labelled by variable sets).

Overview

- Independence
- **Problem Decomposition**
- Undirected Graphs

Decomposition: Maximization

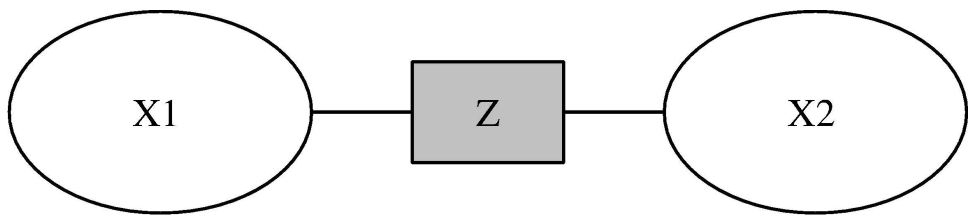
We have discrete boolean variable sets X_1, X_2, Z and assume $X_1 \cap X_2 = Z$. Suppose we wish to maximize a function of the form $f(X_1)g(X_2)$ for non-negative functions $f()$, $g()$.

$$\max_{X_1,X_2,Z} f(X_1,Z) \cdot g(X_2,Z) = \max_{Z} \left(\max_{X_1/Z} f(X_1,Z) \max_{X_2/Z} g(X_2,Z) \right)$$

- Time goes $O(2^{|X_1 \cup X_2 \cup Z|})$ to $O(2^{|X_1 \cup Z|} + 2^{|X_2 \cup Z|} + 2^{|Z|})$. *i.e.* good when $|X_1/Z| \gg 1$ and $|X_2/Z| \gg 1$.
- When computation is super-linear in number of variables, significant savings can be made.
- Applies to most optimization, satisfiability (replace $\times$ by $\wedge$ and $+$ by $\vee$), and probability problems (replace $\max$ by $\sum$).

Maximization Algorithm

1. Compute tables $f_{X_1}(Z) = \max_{X_1/Z} f(X_1, Z)$ and $g_{X_2}(Z) = \max_{X_2/Z} g(X_2, Z)$.
2. From these compute $\hat{Z} = \text{argmax}_Z\, f_{X_1}(Z) \cdot g_{X_2}(Z)$.
3. Compute $\widehat{X_1/Z} = \text{argmax}_{X_1/Z} f(X_1|_{Z=\hat{Z}}, \hat{Z})$ and $\widehat{X_2/Z} = \text{argmax}_{X_2/Z} g(X_2|_{Z=\hat{Z}}, \hat{Z})$.
4. Return $(\widehat{X_1}, \widehat{X_2}, \hat{Z})$.



- Reduces computation to local efforts on Z, X_1 and X_2 separately.
- Need to transfer summaries statistics of Z in both directions to make local tasks consistent with the global task.

Finding a Good Decomposition

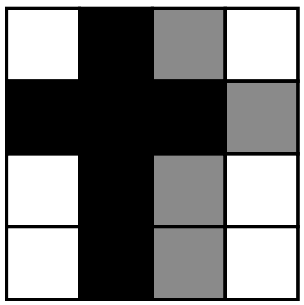
Finding a good separating set Z is like a Mincut problem, but in the dual space (swapping roles of nodes and edges).

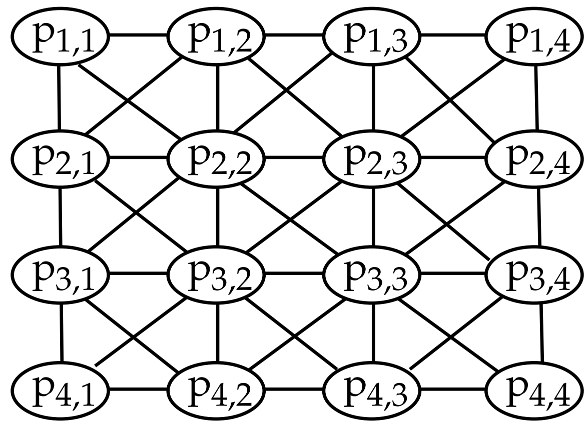
- *Graph partitioning* or *Hypergraph partitioning*, see Alpert and Kahng 1995, world's most studied optimization problem.
- Standard mincut in the dual space (which is cubic time) mostly finds trivial cuts where one side is almost empty, so is of little use.
- "Balanced" Mincut, forcing X_1, X_2 to be similar sizes is NP-complete.
- Local search works poorly on large graphs, e.g., $> 20,000$ nodes.
- Spectral methods (approximate task with maximum eigenvector computation) works quite well.

Overview

- Independence
- Problem Decomposition
- **Undirected Graphs**

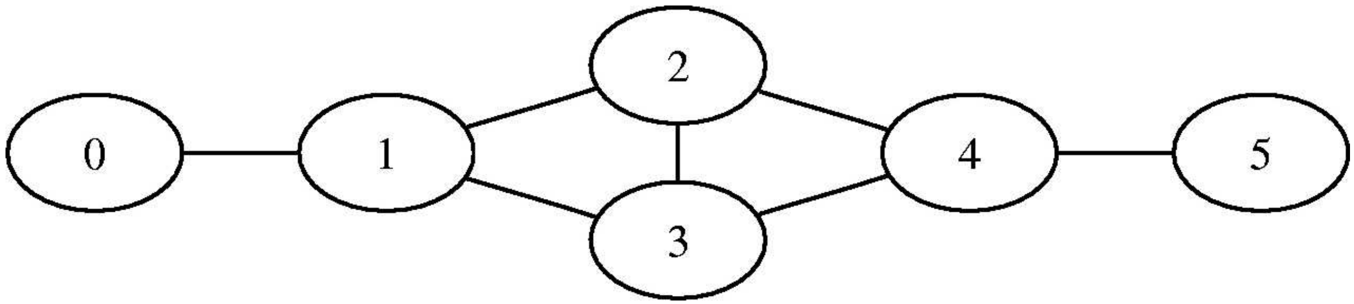
Image Models





Simple 4×4 image. Graph says all pixels influenced only by their neighbour's values. Has checkered history in image analysis, but becoming more successful. Now used in HTML page analysis as well.

Undirected Graph, example



Local Markov Property :

independence when X_1 is a singleton:

$1 \perp\!\!\!\perp 4,5 | 0,2,3$; $2 \perp\!\!\!\perp 0,5 | 1,3,4$; $5 \perp\!\!\!\perp 0,1,2,3 | 4$; etc.

Global Markov Property :

independence on general sets:

$0,1 \perp\!\!\!\perp 4,5 | 2,3$;

Functional form :

$$e(0,1)f(1,2,3)g(2,3,4)h(4,5)$$

Undirected Graph, cont.

Intepretation:

Local Markov Property: the minimal independences required to uniquely determine undirected graph.

Global Markov Property: how to test for independence on general sets using the undirected graph.

Functional Form: form used for purposes of real analysis and computation.

Undirected Graph

Theorem. For an undirected graph on variables X, the following are equivalent when $p(X) > 0$ for all values of X:

Local Markov Property: for all $x \in X$,

$$\{x\} \perp\!\!\!\perp (X - \mathrm{nbrs}(x) - \{x\}) \,|\, \mathrm{nbrs}(x)$$

Global Markov Property: for all $X_1, X_2, Z \subseteq X$, $X_1 \perp\!\!\!\perp X_2 | Z$ iff X_1/Z is separated from X_2/Z in the graph by Z.

Functional Form: for $\mathscr{C}$ the set of cliques in the graph, X_C the restriction of X to the set C, functions $f_C(\cdot)$ exist so that

$$p(X) \;=\; \prod_{C \in \mathscr{C}} f_C(X_C) \;.$$

Undirected Graph, cont.

Proof.

- Clearly global Markov implies local.
- Functional form implies global Markov by functional definition of independence.
- That local Markov property implies the functional (for finite discrete variables) is called the Hammersley-Clifford Theorem.

Optional exercise: find a short half page proof of the Hammersley-Clifford Theorem.

> *Warning:* Don't look up text books, they will give you 10 page proofs using complex results from discrete math.

Next Lecture

- Directed Graphs
- Tree of Cliques
- Cataloguing Problems and Graphical Forms