iPMR

Questions

Example

- In a large online gaming site, how do we match players with similar skill levels?
- Slightly harder. Who will beat who in a basketball league?

- See www.kaggle.com for the above challenge…
- Good answers to both of these use belief networks

Our Journey

Graphical Models	Decision Theory	Learning Probabilistic Models	Mixture and Factor Models	Markov Models	Approximate Inference

- Informally Introduce Belief Networks
- Formalise
  - Graph Theory
  - Probabilistic Graphical Models
    - Belief Networks (Bayesian Networks)
    - Markov Networks
    - Factor Graphs

Another Question

- Consider the variables below. By considering what items are directly dependent on other items can you build a dependency network?
- E.g.
- Here are the items.

Belief Networks

- **Belief Networks** represent the structure of probability distributions in ways that relate to the idea of a dependency network.
- **Starting Point:** The joint probability distribution.

$$P(\mathbf{x}) = P(\mathsf{x}_1)P(\mathsf{x}_2|\mathsf{x}_1)P(\mathsf{x}_3|\mathsf{x}_2,\mathsf{x}_1)P(\mathsf{x}_4|\mathsf{x}_3,\mathsf{x}_2,\mathsf{x}_1)\ldots$$

or more formally

$$P(\mathbf{x}) = P(\mathsf{x}_1)\prod_{i=2}^{M} P(\mathsf{x}_i|\mathsf{x}_{<i})$$

Notation:

- I use *M* rather than *D* (which Barber uses) for the dimensionality of a variable. (D means dataset).
- <*i* means the set of all the indices that are less than *i*.
- *x* with a set or vector subscript means the collection of *x* values with subscripts in that set/vector.

Conditional Probability

- Consider

$$P(\mathrm{x}_6|\mathrm{x}_5,\mathrm{x}_4,\mathrm{x}_3,\mathrm{x}_2,\mathrm{x}_1)$$

- Suppose all x_i were binary.
- How would you encode this probability?

Conditional Probability Tables

- Let us look at the simple case

For $P(\text{Toothache}, \text{Cavity})$ we can write

	Toothache = true	Toothache = false
Cavity = true	0.04	0.06
Cavity = false	0.01	0.89

For $P(\text{Cavity}|\text{Toothache})$ we can write

	Toothache = true	Toothache = false
Cavity = true	0.8	0.063
Cavity = false	0.2	0.937

- But what if conditioning on many items?
- Multidimensional table. Very costly.

Independence...

- Two random variables are independent if we can write

$$P(\mathrm{x}, \mathrm{y}) = P(\mathrm{x})P(\mathrm{y})$$

- We can extend this to sets of random variables

$$P(\mathbf{x}, \mathbf{y}) = P(\mathbf{x})P(\mathbf{y})$$

- We can use the rule of conditioning to get

$$P(\mathbf{x}|\mathbf{y}) = \frac{P(\mathbf{x})P(\mathbf{y})}{P(\mathbf{y})} = P(\mathbf{x})$$

…Independence

- Is Toothache **independent of** Cavity **below?**

$P(\text{Toothache}, \text{Cavity})$ is

	Toothache $=$ true	Toothache $=$ false
Cavity $=$ true	0.04	0.06
Cavity $=$ false	0.01	0.89

$P(\text{Cavity}|\text{Toothache})$ is

	Toothache $=$ true	Toothache $=$ false
Cavity $=$ true	0.8	0.063
Cavity $=$ false	0.2	0.937

Conditional Independence

- Rule of independence extends to conditional probabilities,

$$P(\mathbf{x}|\mathbf{y},\mathbf{z}) = \frac{P(\mathbf{x})P(\mathbf{y}|\mathbf{z})}{P(\mathbf{y}|\mathbf{z})} = P(\mathbf{x}|\mathbf{z})$$

- This is conditional independence and is notated by $I(\mathbf{x};\mathbf{y}|\mathbf{z})$
- Some comments on handling conditional probabilities…
  - Each variable must appear either on the right hand side or left hand side of | not both.
  - Conditional independence means you can drop a variable from the right side.

So what?

- What does this buy us?

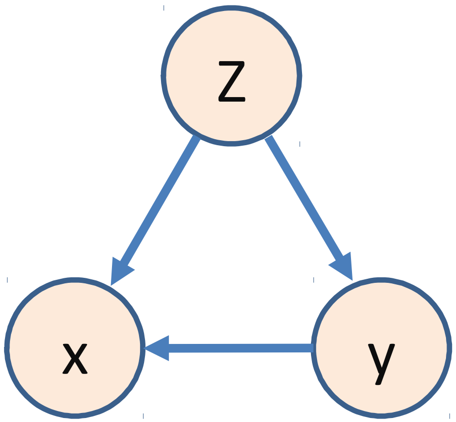
$$P(\mathbf{x}) = P(x_1)P(x_2|x_1)P(x_3|x_2,\cancel{x_1})P(x_4|x_3,\cancel{x_2},x_1)\dots$$

- Utilising conditional independence in the chain rule gives us a more compact representation.
- Think back to the dependency network you built earlier. What were you constructing?

Graphically

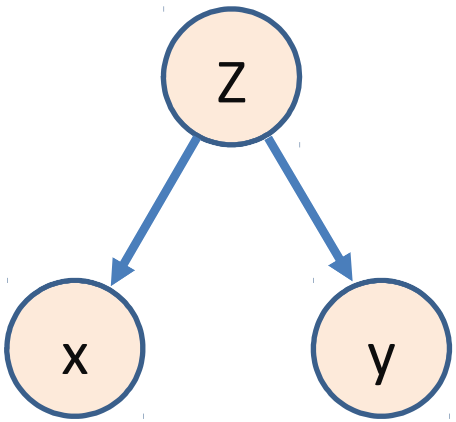
- No independence:

$$P(\mathrm{x},\mathrm{y},\mathrm{z}) = P(\mathrm{z})P(\mathrm{y}|\mathrm{z})P(\mathrm{x}|\mathrm{y},\mathrm{z})$$



- I(x,y|z):

$$P(\mathrm{x},\mathrm{y},\mathrm{z}) = P(\mathrm{z})P(\mathrm{y}|\mathrm{z})P(\mathrm{x}|\mathrm{z})$$



Stop Point

- Questions?

Belief Networks

- Graphical notation that represents various conditional independence assertions for a joint probability distribution.
- How?
  - A Directed Acyclic Graph (DAG) with one node per variable.
  - Look at chain rule expansion.
  - Include all edges except where a variable is dropped from the conditional probability.
  - If $P(\mathrm{r}|\mathrm{s,t,u,v})$ appears in chain rule, but

$$P(\mathsf{r}|\mathsf{s},\mathsf{t},\mathsf{u},\mathsf{v}) = P(\mathsf{r}|\mathsf{s},\mathsf{u})$$

  - then drop directed edge (arrow) from t to r and from v to r.
- Looks like we are going to need some graph theory…

Graphs

- See Barber Chapter 2. This is a quickfire summary.

Graphs

- Graph, Directed Graph, Undirected Graph

Graphs

- Parents, Children, Family, Path, Directed Path, Ancestor, Descendent

Graphs

- Cycle, Loop, Chord

Graphs

- Directed Acyclic Graphs
  - A Directed Acyclic Graph (DAG) is a graph with only directed edges between nodes, and where there are no directed cycles.
  - Can number the nodes so no edge can go from a node to a node with a lower number.
  - Relate this to the chain rule.

Graphs

- Neighbour, Clique, Maximal Clique

Graphs

- Connected, Singly Connected, Spanning Tree

Graphs

- Augmented Graph, Weighted Graph

Graphs

- Maximally Weighted Spanning Tree

Graphs

- Numerical encoding: edge list, adjacency matrix.

Probabilistic Graphical Models

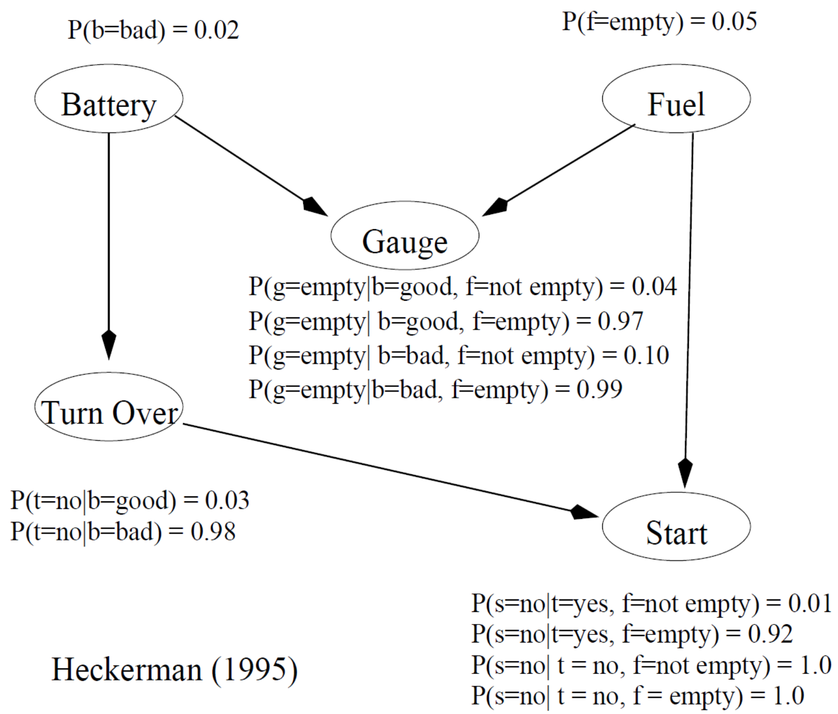
- When we use graph theoretic methods for representing the structure of probability distributions, we are implementing Probabilistic Graphical Models.
  - Why?
  - But it doesn't add anything does it?
- Undirected Graphical Models, Directed Graphical Models (Belief Networks, Bayesian Networks), Factor Graphs.

Stop Point

- Any Questions?

Working with Joint Probabilities

- Consider



- Assume for now we are given the probabilities.
- Then we can build the full probability distribution.
- Then we can make queries about questions we might wa
  - E.g. What is

$$P(f|g = \text{empty}, s = \text{no})$$

Inference

Belief Network

Definition 1 *Belief Network. A Belief Network is a distribution of the form*

$$P(\mathbf{x}) = \prod_i P(\mathsf{x}_i | \mathsf{x}_{Pa(i)})$$

along with a corresponding directed graph with arrows pointing to i from $Pa(i)$. Here, $Pa(i)$ denotes the parents of the node corresponding to x_i in the graph.

- Some questions.
  - Does a particular distribution correspond to one belief network?
  - Given a set of independence relationships encoded in a network, is that network representation unique?
  - Can we always encode all conditional independencies using a belief network?
  - Is it right to interpret a belief network causally?

No, No, No, No (not in general), but useful to construct belief nets causally…

Constructing Belief Networks

- Choose a set of variables (those relevant to the domain).
- Choose an order to those variables.
- For each variable in turn
  - Add a node to the graph for that variable.
  - Add directed edges *from* all the existing nodes the variable *directly* depends on *to* the new node.
  - Add the corresponding conditional probability to the chain rule.
- Note the sensitivity to the order.
- If we choose a less good order, we may end up encoding fewer conditional independencies (and so have costly encoding).
- Hint: Choose order causally, from cause to effect, to naturally capture the most independence relationships in the graph.
- But always remember that a belief network does not necessarily encode a causal order.
  - A belief network that does encode causal order is called a *causal graph.*
- Some examples.

Belief Network

- A belief network decouples
  - Structural aspects of the distribution of variables in the system

from

  - Quantification of the probabilities for the variables in the system.
- This is good because it is possible to develop operators on the structure that apply regardless of the precise probabilities.
- This is also good as structural elements can be easier to elicit than the actual probabilities.
- This is the point of Belief Networks.

Where do the probabilities come from?

- Indeed…
- We just pulled them out of thin air didn’t we?
- From experts, from local models, learning. More later.

Independence Relationships

So a Belief Network encodes conditional independence relationships, right?

Yes...

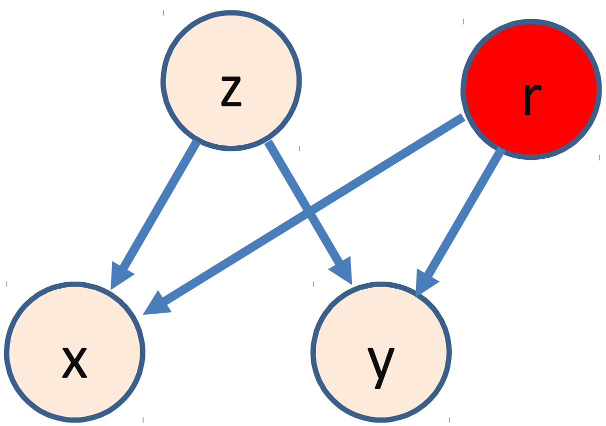
So it must be trivial to read off the conditional independence relationships from a belief net, right?

Um, Er, Um, Er, well kind of...

- How can find out if A is conditionally independent of B given C?
- Not good enough to just look locally.

Independence Relationships

- How can find out if A is conditionally independent of B given C?
- Not good enough to just look locally.
- Consider I(x,y|z)…

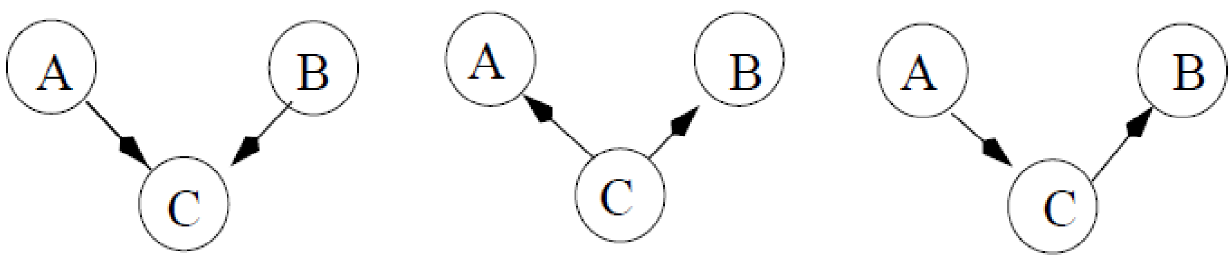


Rules of D-Separation

- To find out if things are independent we use D-Separation.
- If *every* path from a set of notes *X* to a set of *Y* is blocked by set *Z*, then we have

$$I(\mathbf{x}_X, \mathbf{x}_Y | \mathbf{x}_Z)$$

- A path is blocked iff
  - A node in *Z* is on the path and is head to tail wrt the path.
  - A node in Z is on the path and is tail to tail wrt the path.
  - There is a node on the path that is head to head, and neither that node nor any of its descendents is in *Z.*



C is head-to-head C is tail-to-tail C is head-to-tail

D-Seperation

What a painful procedure.

- It is... but it makes sense...

More later

- Next week. More belief networks.
- In the meantime.
  - Read Barber Chapter 2 and 3
  - Practice questions at the end of the chapters.
  - Preparation: read Barber Chapter 4
  - Try out a kaggle challenge.