Probabilistic Graphical Models

1 Provide a simple way to visualize the structure of a probabilistic model

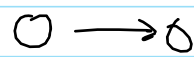
2 Can be used to design and motivate new models

3 Can reveal insights and properties of the model, including

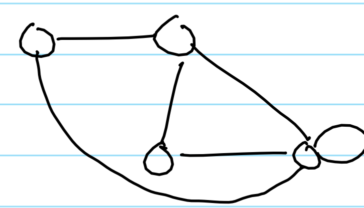
a) cond. independence through graph inspection

3 Computations required by inference and learning in sophisticated models can be formulated as operations on the graph

A graph consists of nodes (aka vertices) and links (and/or edges)

directed 

undirected 



Start with

Bayes nets, directed graphical models

Markov random fields, undirected graphical models

Directed graphs are good for "casual" relationships between variables, undirected graphs are better suited to expressing soft constraints between variables.

Bayesian Networks

Motivation:

Consider an arbitrary joint dist over three variables a, b, and c, $p(a, b, c)$.

a, b, c can be discrete, cont., etc.

We can use the prod. rule to write

$$p(a, b, c) = p(c \mid a, b)\, p(a, b)$$

and again

$$p(a, b, c) = p(c \mid a, b)\, p(b \mid a)\, p(a)$$

Note: this decomposition of the joint always holds!

Graphical Model Representation

Recipe:

1 node for each variable

2 for each conditional dist

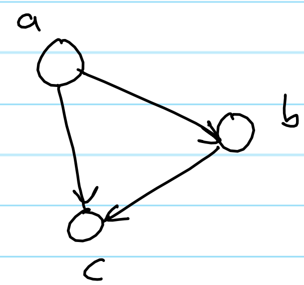
a add directed links to the graph from the nodes corresponding to the var's on which the dist. is conditioned

i.e. for $p(c \mid a, b)$ there should be edges from a and b to c

and $p(a)$ will have no incoming links

Example $p(a, b, c) = p(c|a,b)\,p(b|a)\,p(a)$

$\Downarrow$



Note that $p(a,b,c)$ is symmetrical w.r.t. the variables but $p(c|ab)\,p(b|a)\,p(a)$ is not!

* A different ordering would give rise to a different graphical representation

What about?:

$$p(x_1, \dots, x_K) = p(x_K | x_1, \dots, x_{K-1}) \cdots p(x_2|x_1)\,p(x_1)$$

This can be represented with a graph with K nodes, each of which is connected to all lower numbered nodes.

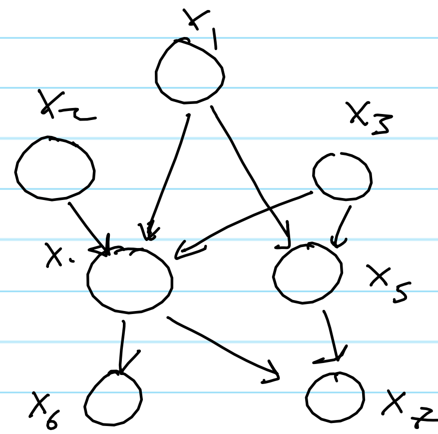
This is a fully connected graph because there is an edge btwn all pairs of nodes.

So far only complete joint dist's.

Absense of links in the graph conveys information about the properties of the class of distributions the graph represents.

PRINTABLE NOTEPAD 1

Example:



We can read the joint dist. off of this graph

$$p(x_1)\,p(x_2)\,p(x_3)\,p(x_4|x_1,x_2,x_3)\,p(x_5|x_1,x_3)$$
$$\times\ p(x_6|x_4)\,p(x_7|x_4,x_5)$$

Note that this is <u>not</u> equal to

$$p(x_1,\dots x_7) = p(x_7|x_6\cdots x_1)\,p(x_6|x_5\cdots x_1)\cdots p(x_2|x_1)\,p(x_1)$$

and thus it is a less flexible specification/ definition of the joint dist'n.

<u>General</u>

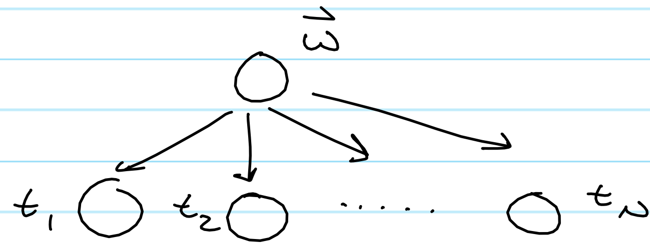
$$p(\vec{x}) = \prod_{k=1}^{K} p(x_k|pa_k)$$

where pa_k stands for the parents of x_k in the graph. model

<u>Note</u> $p(\vec{x})$ is easy to normalize if CPT's are normalized.

Example Graphical Models

Polynomial regression



R.V.'s $\quad \vec{t} = (t_1, \ldots, t_N)^T \quad$ observed data

$\quad \vec{w} \quad$ vector of poly. coefficients

parameters

σ^2 noise variance

α precision of Gaussian prior ofer $\vec{w}$

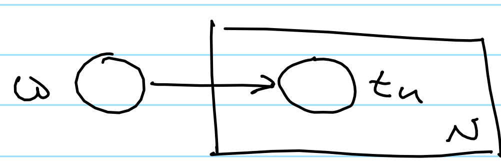
$\vec{x}$ input data

Joint dist

$$p(\vec{w}, \vec{t}) = p(\vec{w}) \cdot \prod_{n=1}^{N} p(t_n \mid \vec{w})$$

Plate notation

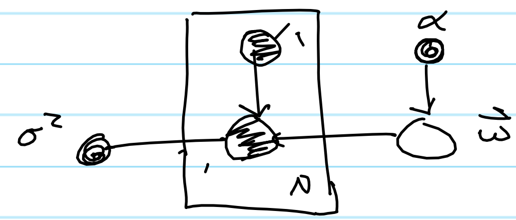
Allows for graphical replication, ie.



Making parameters explicit

Making params explicit cont.

Filled circles indicate observed variables



$\{t_n\}_1^N$ are observed, $\sigma^2, \{x_n\}_1^N$, and α are "observed" parameters

$\vec{\omega}$ is a "latent" or "hidden" variable

Posterior inference

We will often be interested in the posterior distribution of the latent variables in the model, here $\vec{\omega}$

Note, via Bayes rule

$$p(\vec{\omega} \mid \vec{t}) \propto p(\vec{\omega}) \prod_{n=1}^{N} p(t_n \mid \vec{\omega})$$

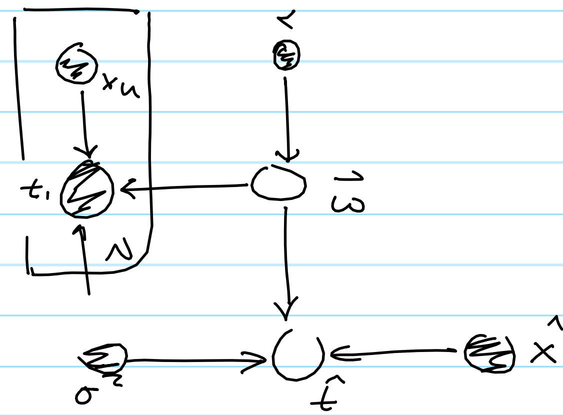
For prediction and other related tasks we may not care about the model parameters

Given a new input value $\hat{x}$ we might wish to predict $\hat{t}$ (really the dist of $\hat{t}$)

The joint dist is

$$p(\hat{t}, \vec{t}, \vec{\omega} \mid \hat{x}, \vec{x}, \alpha, \sigma^2) = \prod p(t_n \mid x_n, \vec{\omega}, \sigma^2)\, p(\vec{\omega} \mid \alpha) \times p(\hat{t} \mid \hat{x}, \vec{\omega}, \sigma^2)$$

The corresponding graphical model is



The predictive dist'n for $\hat{t}$ can be arrived at through the sum rule

$$p(\hat{t} \mid \vec{x}, \vec{t}, \alpha, \sigma^2) = \int p(\hat{t}, \vec{\omega} \mid \hat{x}, \vec{t}, \alpha, \sigma^2)\, d\vec{\omega}$$

Generative models and sampling

Will cover general sampling techniques later, graphical models make ancestral sampling easy.

Recipe

Start at root and sample from conditional distributions after all of their parents have been sampled.

Requires

Being able to sample from cond. dist's

use? Monte Carlo integration see eg.