EE 239AS.2 - Neural signal processing and machine learning
Prof. Jonathan Kao
(with notes adapted from Prof. Byron Yu, CMU)

Graphical models

1 An overview: why study graphical models?

Graphical models are diagrammatic representations of probability distributions.

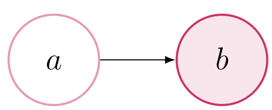
They are especially convenient because:

- They provide a simple way to visualize the structure of a probabilistic model.
- Properties of the model, such as conditional independence, can be obtained by inspection of the graph.

Aside: these will be especially useful when we consider e.g., Kalman filters, which make important use of conditional independencies to simplify distributions.

In a graphical model, we have *nodes* and *links*. Each *node* represents a random variable, and each *link* represents a probabilistic relationship between random variables.

Before we proceed, here are a few definitions.



Here, we have a directed graph with two nodes, a and b, where b is observed, and thus shaded.

- Child: node b is the child of node a if a link connects node a to node b, as in Figure 1.
- Generative models: generative models are models that explain how observed data was generated.

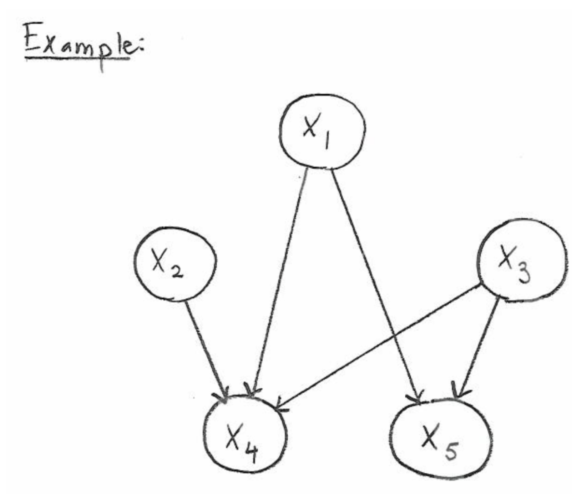
- Latent (or hidden) variables: a node is *latent* or *hidden* if we do not observe it in the data; the primary role of latent variables is to allow a complicated distribution over the observed variables in terms of far simpler conditional distributions. The latent variables need not have a physical interpretation.
- Observed variables (observations): a node is *observed* if we know its value; said differently, an observation is an experimental instantiation or data point of some random variable that we have collected. Observed nodes will be shaded in; thus in Figure 1, node b is *observed*.
- Parent: node a is the parent of node b if a link connects node a to node b, as in Figure 1.

Then, for a graph with K nodes, the joint distribution of the variables corresponding to the nodes, e.g. x_1, x_2, $\ldots$, x_k, is given by:

$$P(\mathbf{x}) = \prod_{k=1}^{K} P(x_k|\mathsf{par}(x_k)), \tag{1}$$

where $\mathsf{par}(x_k)$ denotes the parents of node x_k.

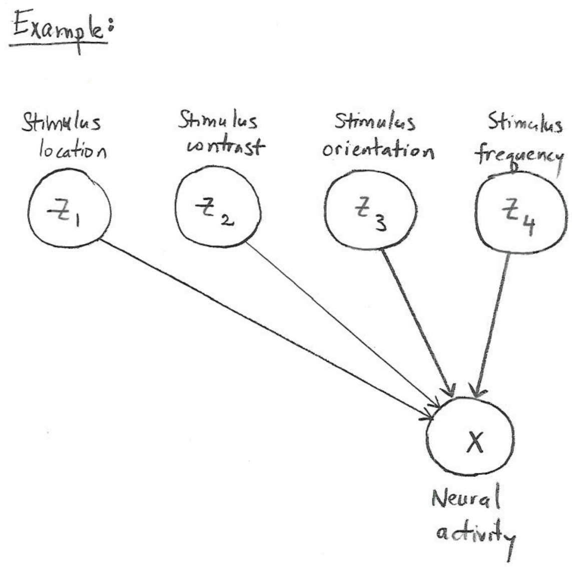
2 Directed graphical models (aka Bayesian networks)



Using our definition, we can read from this graph the joint distribution of $P(x_1, x_2, x_3, x_4, x_5)$. Concretely,

$$P(x_1, x_2, x_3, x_4, x_5) = P(x_1)P(x_2)P(x_3)P(x_4|x_1, x_2, x_3)P(x_5|x_1, x_3)$$

Aside: how are graphical models typically used in neuroscience? Usually, we record neural activity, x, and want to explain it in terms of some stimuli that led to that neural activity, $z_1, \ldots, z_M$.



How would you factorize this graph?

In general, most (if not all) of the models for the rest of this course can be expressed as a directed graphical model.

For any probability distribution, $P(x_1, \ldots, x_k)$ we can write:

$$P(x_1, \ldots, x_k) = P(x_1)P(x_2|x_1)P(x_3|x_1, x_2) \cdots P(x_k|x_1, \ldots, x_{k-1}).$$

If this represents the simplest factorized form of $P(x_1, \ldots, x_k)$, then we call the graph *fully connected*. It is the absence of links that conveys interesting properties about the probability distributions.

3 Conditional independence

Recall the definition of independence:

$$
\begin{aligned}
P(a|b) &= P(a) \\
&\text{or} \\
P(a,b) &= P(a)P(b)
\end{aligned}
$$

The definition of conditional independence is:

$$
\begin{aligned}
P(a|b,c) &= P(a|c) \\
&or \\
P(a,b|c) &= P(a|c)P(b|c)
\end{aligned}
$$

We would say that "a and b are conditionally independent given c."

For each of the following graphical models, let's ask:

1. What is the factored form of $P(a,b,c)$?
2. Are a and b independent?
3. Are a and b conditionally independent given c?

Here are three examples.

Example 1

Here, we have a directed graph with three nodes, a, b, and c. In this example, c could represent a stimulus, while a and b represent the firing rates of neuron 1 and neuron 2, respectively.

Let's answer the above questions.

1. $P(a, b, c) = P(c)P(a|c)P(b|c)$
2. No. Independence means $P(a, b) = P(a)P(b)$. Let's see if this is true.

$$\begin{aligned} P(a,b) &= \sum_c P(a,b,c) \\ &= \sum_c P(c)P(a|c)P(b|c) \\ &= \sum_c P(a,c)P(b|c) \\ &= \sum_c P(a)P(c|a)P(b|c) \\ &= P(a)\sum_c P(c|a)P(b|c). \end{aligned}$$

   For independence to hold, $P(c|a) = P(c)$. However, this is not the case.
3. Yes. Conditional independence means $P(a, b|c) = P(a|c)P(b|c)$.

$$\begin{aligned} P(a,b|c) &= \frac{P(a,b,c)}{P(c)} \\ &= \frac{P(c)P(a|c)P(b|c,a)}{P(c)} \\ &= \frac{P(c)P(a|c)P(b|c)}{P(c)} \\ &= P(a|c)P(b|c). \end{aligned}$$

Example 2

Here, we have another directed graph with three nodes, a, b, and c. Here, a could represent the stimulus at time step 1, c at time step 2, and b at time step 3.

Let's answer the above questions.

1. $P(a, b, c) = P(a)P(c|a)P(b|c)$

2. No.

$$
\begin{aligned}
P(a,b) &= \sum_c P(a,b,c) \\
&= P(a)\sum_c P(c|a)P(b|c).
\end{aligned}
$$

For independence to hold, $P(c|a) = P(c)$. However, this is not the case.

3. Yes.

$$
\begin{aligned}
P(a,b|c) &= \frac{P(a)P(c|a)P(b|c)}{P(c)} \\
&= \frac{P(a,c)P(b|c)}{P(c)} \\
&= P(a|c)P(b|c).
\end{aligned}
$$

Example 3

Here, we have yet another directed graph with three nodes, a, b, and c. Here, a could be stimulus attribute 1, b could be stimulus attribute 2, and c could be the neuron's firing rate. Let's answer the above questions.

1. $P(a,b,c) = P(a)P(b)P(c|a,b)$.
2. Yes. $P(a,b) = \sum_c P(a)P(b)P(c|a,b)$ simplifies straightforwardly.

3. No.

$$
\begin{aligned}
P(a,b|c) &= \frac{P(a,b,c)}{P(c)} \\
&= \frac{P(a)P(b)P(c|a,b)}{P(c)} \\
&= \frac{P(a)P(b)P(a|c,b)P(c|b)}{P(c)P(a|b)} \\
&= \frac{P(a)P(b,c)P(a|c,b)}{P(c)P(a|b)} \\
&= \frac{P(a)P(c)P(b|c)P(a|c,b)}{P(c)P(a|b)} \\
&= \frac{P(a)P(b|c)P(a|c,b)}{P(a)} \\
&= P(b|c)P(a|c,b)
\end{aligned}
$$

Summary

From these examples, we can construct independence and conditional independence statements for graphs. As we demonstrated through these examples:

1. In Example 1, a and b are dependent, but conditionally independent given c.
2. In Example 2, a and b are dependent, but conditionally independent given c.
3. In Example 3, a and b are independent, but conditionally dependent given c.