# Pattern Recognition and Machine Learning Chapter 8 Summary*

CSCI-GA 3033 Bayesian Machine Learning
New York University
Yilun Kuang

Sep 23, 2022

## 8.0 Preliminaries

- Bayesian Networks → Directed Graphical Models
  - Bayesian Networks are suited for representing causal relationships between random variables
- Markov Random Fields → Undirected Graphical Models
  - Markov Random Fields can be used to express soft constraints between random variables
- Both Directed & Undirected Graphical Models can be converted into a Factor Graph.

## 8.1 Bayesian Networks

A fully-connected Directed Acyclic Graph (DAG) is associated with a joint distribution over all the nodes of the graph via the chain rule of probability

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

The following fully-connected DAG

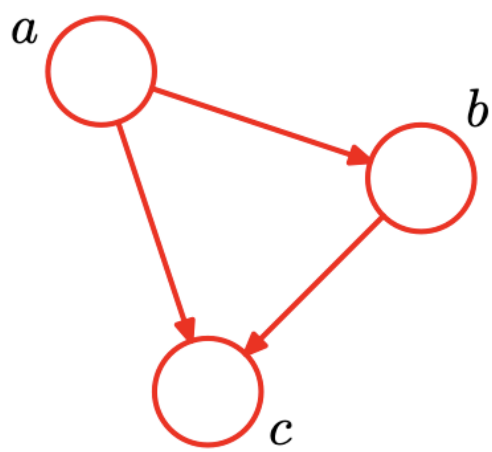


Figure 1: Fully-Connected DAG

has an associated joint distribution $p(a, b, c)$ and can be factorized as follows:

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

In the case of the absence of links, i.e. not fully-connected DAG, we have

*Section 8.1 - 8.2

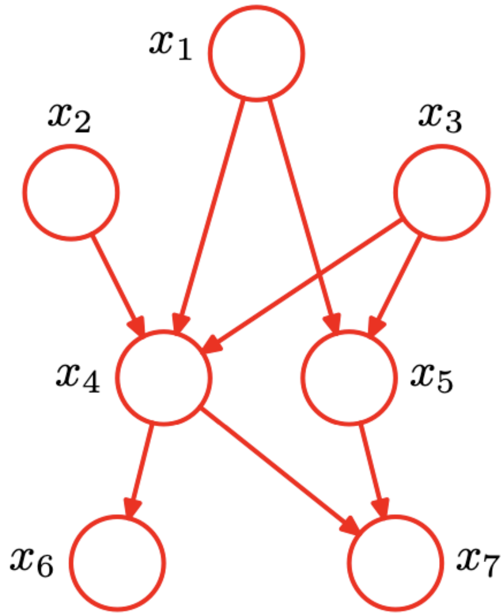


Figure 2: DAG

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

More generally, we have the following

$$p(\mathbf{x}) = \prod_{k=1}^{K} p(x_k|\mathrm{pa}_k)$$

### 8.1.1 - 8.1.4 Polynomial Regression, Generative Models, Discrete Variables, Linear-Gaussian Models

Skipped

## 8.2 Conditional Independence

### 8.2.0 Definition of Conditional Independence

If $p(a|b, c) = p(a|c)$, then we say $a$ is conditionally independent of $b$ given $c$, denoted $a \perp\!\!\!\perp b|c$. Alternatively, if $a \perp\!\!\!\perp b|c$, we have $p(a, b|c) = p(a|b, c)p(b|c) = p(a|c)p(b|c)$.
In general we can also think of conditional independence between sets of variables ****$\mathcal{X} \perp\!\!\!\perp \mathcal{Y}|\mathcal{V} \iff p(\mathcal{X}|\mathcal{V})p(\mathcal{Y}|\mathcal{V})$.
Marginal Independence is a trivial case of conditional independence: $\mathcal{X} \perp\!\!\!\perp \mathcal{Y} \iff \mathcal{X} \perp\!\!\!\perp \mathcal{Y}| \iff p(\mathcal{X}, \mathcal{Y}) = p(\mathcal{X})p(\mathcal{Y})$.

### 8.2.1 Three Examples

#### 1) Tail-to-Tail

Consider the simple three-node graphical models below,

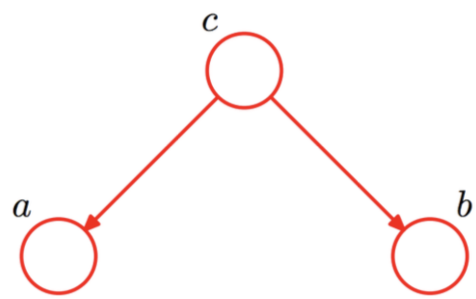


Figure 3: Tail-to-Tail

where none of the variables are observed.
**Exercise**: Is $a$ and $b$ marginally independent?
**Solution**: In general no. Given the above graph, we can factorize the joint distribution into

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

**Digression:**

Notice that by the chain rule of probability, we have

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

This implies that $p(a|b, c) = p(a|c) \implies a \perp\!\!\!\perp b|c$. We'll see in the Tail-Tail Observed example below this is indeed the case so the factorization of the joint distribution based on the graph readout is correct.

To see if $a$ and $b$ are marginally independent, we can check

$$p(a, b) = \sum_c p(a, b, c) = \sum_c p(c)p(b|c)p(a|c) \neq p(a)p(b) \text{ in general.}$$

So $a \not\perp\!\!\!\perp b|$. $\square$.
**Remark:** The node $c$ is said to be *tail-to-tail* with respect to this path because the node is connected to the tails of the two arrows, and the presence of such a path connecting nodes $a$ and $b$ causes these nodes to be dependent.

#### Tail-Tail Observed

**Exercise:** Now if the variable $c$ is observed, we can go on and ask if $a$ and $b$ are conditionally independent given the observation $c$.
**Solution:** Yes. Notice that the joint distribution of $a$ and $b$ conditioned on $c$ is given by

$$p(a, b|c) = \frac{p(a, b, c)}{p(c)} = \frac{p(c)p(b|c)p(a|c)}{p(c)} = p(b|c)p(a|c)$$

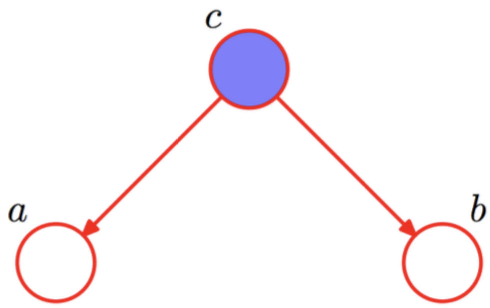


Figure 4: Tail-Tail Observed

So $a \perp\!\!\!\perp b|c$. $\square$.
**Remark:** The conditioned node $c$ blocks the path from $a$ to $b$ and causes $a$ and $b$ to become conditionally independent.

**2) Tail-Head**

Consider another example of Tail-Head.

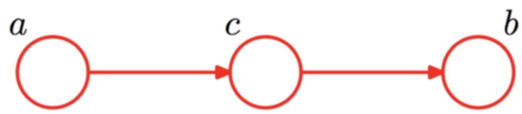


Figure 5: Tail-Head

**Exercise:** Is $a$ and $b$ marginally independent?
**Solution:** In general no. Given the graph, we have the factorization of the joint distribution as follows

$$p(a,b,c) = p(a)p(c|a)p(b|c).$$

Then

$$p(a,b) = \sum_c p(a,b,c) = p(a)\sum_c p(c|a)p(b|c) = p(a)\sum_c p(b,c|a) = p(a)p(b|a) \neq p(a)p(b) \text{ in general.}$$

So $a \not\perp\!\!\!\perp b|$. $\square$.

**Tail-Head Observed**

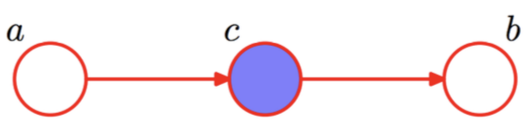


Figure 6: Tail-Head Observed

**Exercise**: Now if $c$ is observed, is $a$ and $b$ conditionally independent given $c$ ?
**Solution**: Yes. As we'll see, the observation of $c$ blocks the path from $a$ to $b$. To show this, notice that

$$p(a,b|c) = \frac{p(a,b,c)}{p(c)} = \frac{p(a)p(c|a)p(b|c)}{p(c)} = \frac{p(a,c)p(b|c)}{p(c)} = p(a|c)p(b|c)$$

So $a \perp\!\!\!\perp b|c$. $\square$.

**3) Head-Head**

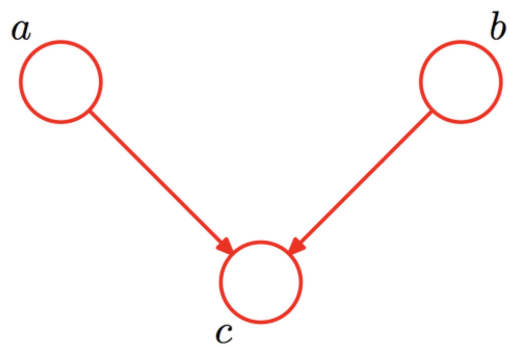


Figure 7: Head-Head

**Exercise**: Is $a$ and $b$ marginally independent?
**Solution**: Yes, contrary to the two previous examples. Notice that

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

So $a \perp\!\!\!\perp b|$. $\square$.

**Head-Head Observed**

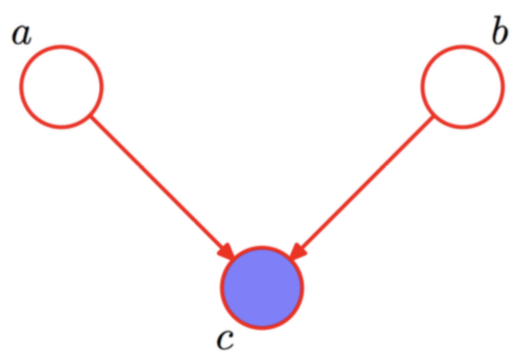


Figure 8: Head-Head Observed

**Exercise**: Given that $c$ is now observed, is $a$ and $b$ conditionally independent given $c$ ?
**Solution**: In general no, contrary to the two previous examples. Notice that

$$p(a,b|c) = \frac{p(a,b,c)}{p(c)} = \frac{p(a)p(b)p(c|a,b)}{p(c)} \neq p(a|c)p(b|c) \text{ in general.}$$

So $a \not\!\perp\!\!\!\perp b|c$. $\square$.

**Remark**: The Head-Head example has the opposite behavior from the first two. The node $c$ is *head-to-head* with respect to the path from $a$ to $b$ as it connects to the heads of the two arrows. When node $c$ is not observed, $c$ blocks the path and $a$ and $b$ are marginally independent. When $c$ is observed, conditioning on $c$ unblocks the path and the variables $a$ and $b$ are conditionally dependent.
**Remark**: The node $y$ is a descendant of node $x$ if there is a path from $x$ to $y$ in which each step of the path follows the directions of the arrows. A *head-to-head* path will become unblocked if either the node, or *any of its descendants*, is observed.

#### Summary

In general, a *tail-to-tail* or a *head-to-tail* node leaves a path unblocked unless it's observed in which case it blocks the path.
A *head-to-head* node blocks a path if it's unobserved, but once the node, and/or at least one of its descendants, is observed the path becomes unblocked.

#### Car Fuel System Examples for the Head-Head Explaining-Away Phenomenon

Skipped

### 8.2.2 D-Separation

D-Separation follows directly from the three examples above.