Machine Learning: Probabilistic Graphical Models 1

Hao Tang

March 17, 2026

1. There are statistical independencies in probability distributions.
2. A graphical model is a compact representation of dependencies.
3. Statistical independencies can be reasoned on graphs.

Statistical independencies in probability distributions

- What are the statistical independencies in this distribution?

x	y	$p(x, y)$
0	0	12/21
0	1	2/21
1	0	6/21
1	1	1/21

- What are the statistical independencies in this distribution?

x	y	z	$p(x, y, z)$
0	0	0	$3/16$
0	1	0	$6/16$
1	0	0	$1/16$
1	1	0	$2/16$
0	0	1	$4/60$
0	1	1	$1/60$
1	0	1	$8/60$
1	1	1	$2/60$

- It's not at all obvious what statistical independencies there are even when we know the joint distribution.
- In this case,

$$p(x, y|z) = p(x|z)p(y|z), \tag{1}$$

or more succinctly

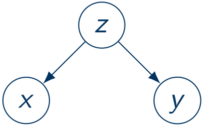
$$x \perp\!\!\!\perp y \mid z. \tag{2}$$

- How many possible independencies can we have for this distribution?

- A dumb algorithm for checking independencies in a distribution
    1. Compute $p(x|z)$, $p(y|z)$, and $p(x, y|z)$.
    2. Check if $p(x, y|z) = p(x|z)p(y|z)$.
    3. Repeat for all combination.
- This is inefficient, and we lack a language to talk about independencies.

Graphical models for encoding factorization

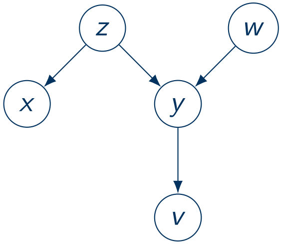
- The following graph



represents the distribution

$$p(x, y, z) = p(x|z)p(y|z)p(z). \quad (3)$$

- What factorization can we read off from this graph?



- The above graph represents the distribution

$$p(v, w, x, y, z) = p(v|y)p(x|z)p(y|z, w)p(z)p(w). \tag{4}$$

- **Definition.** A distribution p factorizes according to a directed graph G if

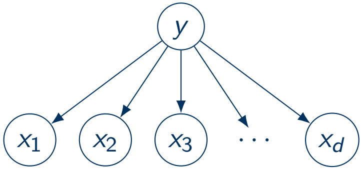
$$p(x_1, \ldots, x_n) = \prod_{i=1}^{n} p(x_i | \mathrm{Pa}_{x_i}) \tag{5}$$

  where Pa_x is the set of parent vertices of x in G.

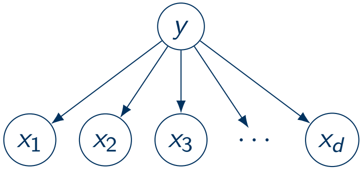
- Obviously, the graph G needs to have the same number of vertices as there are in the distribution p.

- The vertex u is one of v's parent if there is an edge from u to v.

- What is the distribution that factorizes according to the following graph?



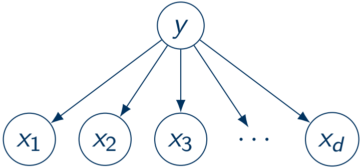
- What is the distribution that factorizes according to the following graph?



- The distribution that factorizes according to the graph has the form

$$p(x, y) = p(y)p(x_1|y)p(x_2|y)\cdots p(x_d|y) = p(y)\prod_{i=1}^{d} p(x_i|y). \tag{6}$$

- What is the distribution that factorizes according to the following graph?

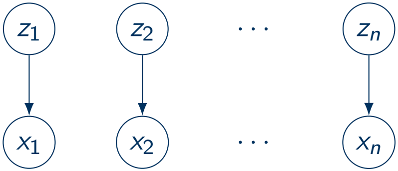


- The distribution that factorizes according to the graph has the form

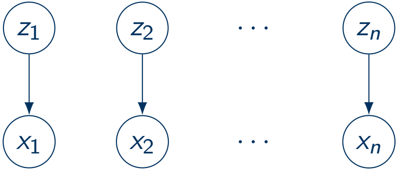
$$p(x, y) = p(y)p(x_1|y)p(x_2|y)\cdots p(x_d|y) = p(y)\prod_{i=1}^{d} p(x_i|y). \qquad (6)$$

- This is known as the naive Bayes model.

- What is the distribution that factorizes according to the following graph?



- What is the distribution that factorizes according to the following graph?



- The distribution that factorizes according to the graph has the form

$$p(x_1, \ldots, x_n, z_1, \ldots, z_n) = \prod_{i=1}^{n} p(x_i|z_i)p(z_i) \tag{7}$$

- We get to choose what $p(x_i|z_i)$ and $p(z_i)$ are, and giving them forms is called **parametrization**.

- We get to choose what $p(x_i|z_i)$ and $p(z_i)$ are, and giving them forms is called **parametrization**.
- What happens if we take

$$p(z_i) = \pi_{z_i} \tag{8}$$

$$p(x_i|z_i) = \frac{1}{(2\pi)^{d/2}|\Sigma_{z_i}|^{1/2}} \exp\left(-\frac{1}{2}(x_i - \mu_{z_i})^\top \Sigma_{z_i}^{-1}(x_i - \mu_{z_i})\right) \tag{9}$$

as our parametrization?

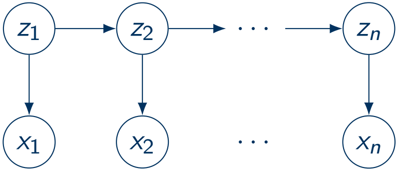
- We get to choose what $p(x_i|z_i)$ and $p(z_i)$ are, and giving them forms is called **parametrization**.
- What happens if we take

$$p(z_i) = \pi_{z_i} \tag{8}$$

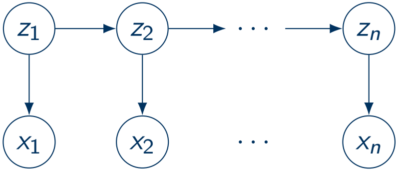
$$p(x_i|z_i) = \frac{1}{(2\pi)^{d/2}|\Sigma_{z_i}|^{1/2}} \exp\left(-\frac{1}{2}(x_i - \mu_{z_i})^\top \Sigma_{z_i}^{-1}(x_i - \mu_{z_i})\right) \tag{9}$$

  as our parametrization?
- This is the Gaussian mixture model.

- What is the distribution that factorizes according to the following graph?



- What is the distribution that factorizes according to the following graph?



- The distribution that factorizes according to the graph has the form

$$p(x_1, \dots, x_n, z_1, \dots, z_n) = p(z_1) \prod_{i=2}^{n} p(z_i | z_{i-1}) \prod_{i=1}^{n} p(x_i | z_i). \tag{10}$$

- This is known as the hidden Markov model.

- In general, a directed graph paired with a distribution that factorizes accordingly is called a **Bayesian network**.

- Can we read off statistical independencies from a graph?

Reading off statistical independencies from a directed graph

- There are three basic structures.

chain

$x \perp\!\!\!\perp z \mid y$

common cause

$x \perp\!\!\!\perp z \mid y$

v-structure

$x \perp\!\!\!\perp z$

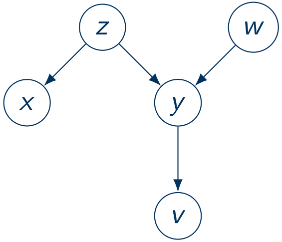
$x \not\perp\!\!\!\perp z \mid y$

- The basic structure $x \rightarrow y \rightarrow z$ is blocked when y is given.
- The basic structure $x \leftarrow y \rightarrow z$ is blocked when y is given.
- The basic structure $x \rightarrow y \leftarrow z$ is blocked if y and its descendants are not given.

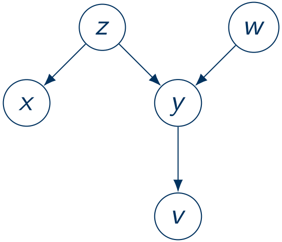
- The basic structure $x \to y \to z$ is blocked when y is given.
- The basic structure $x \leftarrow y \to z$ is blocked when y is given.
- The basic structure $x \to y \leftarrow z$ is blocked if y and its descendants are not given.
- A path is blocked if any basic structure along the path is blocked.
- If all paths connecting two vertices are blocked, then they are **d-separated**.

- Let $I(p)$ be the set of all statistical independencies in a distribution p.
- Let $I(G)$ be the set of all d-separation in a directed graph G.
- **Theorem.** For almost all distributions that factorize over G, $I(p) = I(G)$.

- What is the set of d-separations in the following graph?



- What is the set of d-separations in the following graph?



- The list of basic structures includes
  $x \perp\!\!\!\perp y \mid z$,
  $z \perp\!\!\!\perp w$, $z \not\perp\!\!\!\perp w \mid y$, $z \not\perp\!\!\!\perp w \mid v$
  $z \perp\!\!\!\perp v \mid y$, $w \perp\!\!\!\perp v \mid y$