Machine Learning

Lecture 25: Statistical dependencies 1

Hao Tang

December 9, 2022

Recap of statistical independence

- Two variables x and y are independent if

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

- Equivalently, two variables x and y are independent if

$$p(x|y) = p(x) \tag{2}$$

- We will use $x \perp y$ to denote the independence of x and y.

Independence of many variables

- If $\{x_1, \dots, x_n\} \perp \{y_1, \dots, y_m\}$ then

$$p(x_1, \dots, x_n, y_1, \dots, y_m) = p(x_1, \dots, x_n) p(y_1, \dots, y_m) \tag{3}$$

- Independence implies factorization.

- For example, suppose $x \in \mathcal{X}$, $y \in \mathcal{Y}$, $z \in \mathcal{Z}$. If $\{x, y\} \perp z$,

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

Independence of many variables

- If $\{x_1, \ldots, x_n\} \perp \{y_1, \ldots, y_m\}$ then

$$p(x_1, \ldots, x_n, y_1, \ldots, y_m) = p(x_1, \ldots, x_n)p(y_1, \ldots, y_m) \tag{3}$$

- Independence implies factorization.

- For example, suppose $x \in \mathcal{X}$, $y \in \mathcal{Y}$, $z \in \mathcal{Z}$. If $\{x, y\} \perp z$,

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

- The original domain is $\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$, but after factorization, the domain we need to consider, $\mathcal{X} \times \mathcal{Y}$ and $\mathcal{Z}$, is much smaller than $\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$.

Mutual independence vs pairwise independence

- The variables x_1, x_2, x_3 are mutually independent if

$$p(x_1, x_2, x_3) = p(x_1)p(x_2)p(x_3). \tag{5}$$

- If $x_1 \perp x_2$, $x_2 \perp x_3$, and $x_1 \perp x_3$, then x_1, x_2, x_3 are pairwise independent.

- Mutual independence implies pairwise independence, but the converse is not necessarily true.

Conditional independence

- The variables x and y are conditionally independent given z if

$$p(x, y|z) = p(x|z)p(y|z). \tag{6}$$

- In this case, we write $x \perp y \mid z$.

- The sets of variables $\{x_1, \ldots, x_n\}$ and $\{y_1, \ldots, y_m\}$ are conditionally independent given $\{z_1, \ldots, z_t\}$ if

$$\begin{aligned} &p(x_1, \ldots, x_n, y_1, \ldots, y_m|z_1, \ldots, z_t) \\ &\quad = p(x_1, \ldots, x_n|z_1, \ldots, z_t)p(y_1, \ldots, y_m|z_1, \ldots, z_t). \end{aligned} \tag{7}$$

Testing independence

- By definition of marginalization,

$$p(x|z) = \sum_{y} p(x, y|z) \tag{8}$$

$$p(y|z) = \sum_{x} p(x, y|z) \tag{9}$$

- Check if

$$p(x, y|z) = p(x|z)p(y|z) \tag{10}$$

  for all x, y, and z.

- The above algorithm is slow. In general, testing independence is a hard problem.

“Chain rule” of conditional probabilies

- Any joint probability $p(x_1, x_2, \ldots, x_n)$ can be factorized in any order.

- For example,

$$p(x_1, x_2, \ldots, x_n) = p(x_1)p(x_2|x_1)p(x_3|x_1, x_2) \cdots p(x_n|x_1, \ldots, x_{n-1}). \qquad (11)$$

- Or

$$p(x_1, x_2, \ldots, x_n) = p(x_n)p(x_{n-1}|x_n)p(x_{n-2}|x_{n-1}, x_n) \cdots p(x_1|x_2, \ldots, x_n). \qquad (12)$$

Applying independence

- Every Thursday there is a alarm testing (t).
- The alarm (a) goes off when there is fire (f).
- If the alarm goes off, people in the building should meet at the front door (g) on the ground floor.
- People gathers in front the building when there is a strike (s).

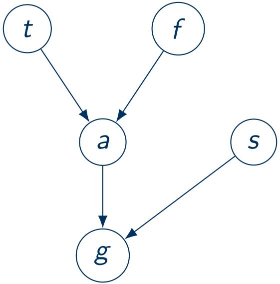
Applying independence

- Alarm testing is independent of a fire ($t \perp f$).
- A strike is independent of what happens in the building ($s \perp \{a, f, t\}$).
- People gathering is independent of fire and alarm testing if we know whether the alarm goes off or whether there is a strike ($g \perp \{f, t\} \mid s, a$).
- Combining the above, we have

$$p(a, t, f, s, g) = p(t)p(f|t)p(a|f, t)p(s|a, f, t)p(g|s, a, f, t) \tag{13}$$

$$= p(t)p(f)p(a|f, t)p(s)p(g|s, a) \tag{14}$$

A (directed) graph representation



$$p(a, t, f, s, g) = p(t)p(f)p(a|f, t)p(s)p(g|s, a) \quad (15)$$

A (directed) graph representation

- Each vertex is a variable.
- A parent has edges pointing from itself to its children.
- The graph is directed and acyclic.
- A distribution factorizes according to a graph if

$$p(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} p(x_i|\mathrm{Pa}(x_i)). \tag{16}$$

- Instead of describing independencies, the graph describes a factorization.

Two objects

- Graph
- Probability distribution
  - A probability distribution has a set of independencies.
  - A probability distribution can factorize according to a graph.
- Can we read off independencies from a graph?

Basic structures

chain

$$x \perp z \mid y$$

common cause

$$x \perp z \mid y$$

v-structure

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

chain

$x \perp z \mid y$

$$p(x, z|y) = \frac{p(x, y, z)}{p(y)} = \frac{p(x)p(y|x)p(z|y)}{p(y)} = p(x|y)p(z|y) \tag{17}$$

common cause

$x \perp z \mid y$

$$p(x, z|y) = \frac{p(x, y, z)}{p(y)} = \frac{p(y)p(x|y)p(z|y)}{p(y)} = p(x|y)p(z|y) \tag{18}$$

v-structure

$x \perp z$

$x \not\perp z \mid y$

$$p(x,z) = \sum_y p(x,y,z) = \sum_y p(x)p(z)p(y|x,y) = p(x)p(z) \tag{19}$$

v-structure

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

If $x \perp z \mid y$,

$$p(x, z) = \sum_y p(x, y, z) = \sum_y p(y)p(x, z|y) = \sum_y p(y)p(x|y)p(z|y)$$
$$= \sum_y p(x|y)p(y, z). \tag{20}$$

But

$$p(x, z) = p(x)p(z) = \sum_y p(x)p(y, z). \tag{21}$$

This can hold only when $p(x|y) = p(x)$, but x and y are not independent; a contradiction.