Probabilities

AI and ML

Bayes' theorem

$$\mathrm{P(H|E)} = \frac{\mathrm{P(H|E) \cdot P(H)}}{\mathrm{P(E)}}$$

$\mathrm{P(H|E)}$: Posterior probability of H given E - Probability after the data or evidence has been taken into account.
$\mathrm{P(E|H)}$: Probability of E given H.
$\mathrm{P(H)}$: Prior probability of H - Probability before the data or evidence has been taken into account.
$\mathrm{P(E)}$: Probability of E / Marginal probability / Normalizing constant.

Example

Let's say we have two events:

- Event A: There is a fire
- Event B: There is smoke

And we know the following probabilities:

- $P(B|A)$: The probability of observing smoke given that there is a fire (likelihood) - 0.9
- $P(A)$: The prior probability of a fire occurring (prior) - 0.01
- $P(B)$: The probability of observing smoke (evidence) - 0.05

=> $P(A \mid B) \approx 0.18$

So, after observing smoke, the updated probability of there being a fire is approximately 18%. This shows how our initial belief in the occurrence of a fire (prior probability) is updated based on the new evidence of observing smoke.

Bayesian network

> A Bayesian network is a **probabilistic graphical model** that represents **a set of random variables and their conditional dependencies using a directed acyclic graph**. It is a powerful tool for reasoning under uncertainty and making predictions based on observed evidence.

Components of a Bayesian network:

- **Nodes**: Represent random variables or events in the domain of interest. Each node corresponds to a variable that can take on different values.
- **Edges**: Directed edges between nodes represent probabilistic dependencies. They indicate the direct influence of one variable on another, representing causal relationships or dependencies.
- **Conditional Probability Tables (CPTs)**: Associated with each node, CPTs quantify the conditional probabilities of each variable given its parents in the network. These tables encode the probabilistic dependencies represented by the edges.
  Bayesian networks allow for efficient inference of probabilities and predictions by propagating evidence through the network using probabilistic principles, such as Bayes' theorem, to update beliefs about variables based on observed data.
- Queries and diagnosis: $P(cause|evidence)$.
- Explanations and prediction: $P(evidence|cause)$.
- Classification: maxclass $P(class|data)$.
- Decision-making with a cost/loss/risk function.

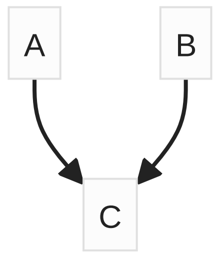
Conditional independence

- **Dependence**: Think of dependence as "interconnectedness." Variables are dependent if they influence each other.
- **Independence**: Think of independence as "self-sufficiency." Variables are independent if they do not rely on each other and their probabilities are not affected by each other's occurrence.
  There exists three canonical cases of conditional independence: head-to-tail, tail-to-tail and head-to-head.

Head-to-tail

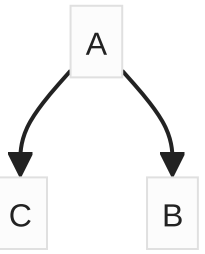
A and C are independent given B.

Tail-to-tail



A and B are dependent given C.

Head-to-Head



B and C are independent given A.