Uncertainty – probability

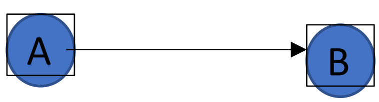
P(A) – marginal probability, formally P(A = a)
P(A, B) - joint probability, formally P(A = a, B = b), P(A) >= P(A, B)
P(A|B) – conditional probability, given B, what is the probability of A,
P(A|B) >= P(A, B), but we don't know P(A), P(B), P(A|B)

P(A, B) = P(A|B) P(B) = P(B|A) P(A) Bayes Rule

Bayesian Network

Why? A way to simplify the storage or computation (training) when we have lots of variables (events) involved through graph models. Assuming each variable has 3 possible values, there are P(A, …, I) has power(3, 9) combinations.

A	B	C	D	E	F	G	H	I
0	0	0						
1	0							
2	0							



The arrow edge indicates the causal relationship

Both of the two terms P(A|B) and P(B|A) are valid!

During training stage, most of the time P(B|A) is more convenient to obtain.

For example, A is a season variables {0, 1, 2, 3}, B is a weather variable {0, 1, 2}

P(B|A) is easier to be trained, e.g. when A = 0, which is Spring season, then you can train how many days of rain, sun shine, and wind. But P(A|B) can be still valid, especially during testing stage. P(B|A)*P(A) = P(A|B) * P(B)

If variables B and E are independent from each other, their joint probability

$P(B, E) = P(B)\, P(E)$ - factorization (decompose the large joint table into multiple products format)

If variables B and E are not independent: $P(B, E) = P(B)P(E|B) = P(E)P(B|E)$