

Bayes Nets

Dhruv Malik, Andy Palan

March 7th 2017

If $P(A|B, C) = P(A|B)$ is true, then A is said to be conditionally independent of C given B.

$$\begin{aligned}P(A|B, C) = P(A|B) &\iff P(C|A, B) = P(C|B) \\P(A|B, C) = P(A|B) &\iff P(A, C|B) = P(A|B) \times P(C|B).\end{aligned}$$

Conditional Independence: X is conditionally independent of Y given Z if and only if $P(X, Y|Z) = P(X|Z) \times P(Y|Z)$, or if and only if $P(X|Y, Z) = P(X|Z)$.

Computing a joint distribution for a system with n variables each with domain size d can take d^n space. We use Bayes nets to avoid this problem.

A Bayes Net is defined to be a directed acyclic graph (DAG) where each node represents a variable, and edges between nodes represent conditionally independent relationships. The key simplifying assumption is that each node is conditionally independent of all non-descendant nodes, given its parents. So, in a single node X , we store the conditional probability distribution $P(X|A_1, A_2, \dots, A_n)$ where A_i is the i^{th} parent of X .

This model enables us to compute the joint distribution for all the variables. Observe that since the model is a DAG, we can topologically sort it. Assume the model has n variables x_1, x_2, \dots, x_n and these are in order of the linearization. This is important because while looking at nodes in this order, we ensure that all parents of a node must be some subset of its predecessors, necessary for the following computation. Regardless of model, $P(x_1, x_2, \dots, x_n) = P(x_1) \times P(x_2|x_1) \times P(x_3|x_1, x_2) \times \dots \times P(x_n|x_1 \dots x_{n-1})$. Call this expression S . So to compute $P(x_1, x_2, \dots, x_n)$ we proceed in order of the sort. We make an inductive argument, the base case is trivial because the first variable x_1 in the sort has no parent so $P(x_1)$ is stored in the node. Now say we are considering some variable x_i , then we need to compute the term $P(x_i|x_1 \dots x_{i-1})$ in S . Assume that some subset of $x_1 \dots x_{i-1}$, say $x_A \dots x_G$ are parents of x_i . Then, $P(x_i|x_1 \dots x_{i-1}) = P(x_i|X_A \dots X_G)$ by our assumption that x_i is conditionally independent of all nodes in the graph besides its parents. This probability is stored in the node by the design of our model. Hence, we can compute the required probability for each term in S , and derive the joint distribution.