Probability And Bayesian Networks

DESKTOP-H800RKQ

October 29, 2024

Contents

1 Uncertainty **1**

2 Bayesian Probability **1**
- 2.1 Independence 2
- 2.2 Conditional Probability 2
- 2.3 Expected Values 3

3 Bayesian Networks **3**
- 3.1 Correlation and Causality 4
- 3.2 Conditional Independence 4

1 Uncertainty

Agents don't know everything but must still make decisions, whether in the absence of info or with noisy info.

An agent can know how uncertain it is.

2 Bayesian Probability

Frequentist probability is **ontological**: pertaining to the world (at this time).

Bayesian probability is **epistemological**: pertaining to knowledge (based on previous experience).

For a **random variable** X that contains a feature and attribute, it can take on values $x \in \text{Domain}(X)$. Assume x is discrete. **Joint probability** $P(x, y)$

is the probability that $X = x$ and $Y = y$ at the same time.

For probabilities:

- $P(X) \geq 0$
- $\sum_x P(X = x) = 1.0$
- $P(a \vee b) = P(a) + P(b)$ if a and b are mutually exclusive
- $P(a) = 0$ means definitely false
- $P(a) = 0$ means definitely true
- $0 < P(a) < 1$ means there is some belief about the truth of a, though this truth value is unknown (not that a is true to some degree)
- probability is a **measure of ignorance**

2.1 Independence

Describe a system with n features, so $2^n - 1$ probabilities. Can use independence to reduce the number of probabilities.

With independence, can sum the options for each feature rather than multiplying to describe all probabilities.

2.2 Conditional Probability

For random variables X and Y, $P(x \mid y)$ is the probability that $X = x$ given $Y = y$.

X and Y are **independent** if and only if $P(X) = P(X \mid Y), P(Y) = P(Y \mid X), P(X, Y) = P(X)P(Y)$, so learning about Y does not influence beliefs about X.

X and Y are **conditionally independent** given Z if and only if $P(X \mid Z) = P(X \mid Y, Z), P(Y \mid Z) = P(Y \mid X, Z), P(X, Y \mid Z) = P(X \mid Z)P(Y \mid Z)$ so learning about Y does not influence beliefs about X if Z is already known. But this does not mean X and Y are independent.

Incorporate Independence

$$P(X \mid Y, Z) = P(X \mid Y)$$

if X and Z are independent given Y

Product Rule

$$P(X, Y) = P(X \mid Y)P(Y)$$

which leads to **Bayes' Rule**

$$P(X \mid Y) = \frac{P(Y \mid X)P(X)}{P(Y)}$$

Sum Rule

$$\sum_x P(X = x, Y) = P(Y)$$

$P(Y)$ is the **marginal distribution** over Y.

2.3 Expected Values

Expected value of a function on X, $V(x)$ is

$$\mathbb{E}(V) = \sum_{x \in \mathrm{Dom}(X)} P(x), V(x)$$

This is useful in decision making, where $V(X)$ is the <u>utility</u> of situation X.

Bayesian Decision Making

$$\mathbb{E}(V(\text{decision})) = \sum_{\text{outcome}} P(\text{outcome} \mid \text{decision})V(\text{outcome})$$

3 Bayesian Networks

Complete independence reduces both representation and inference from $O(2^n)$ to $O(n)$.

Complete mutual independence is rare so more often use conditional independence.

Bayesian Network

- directed acyclic graph
- encodes independences in a graphical format
- edges give $P(X_i \mid \text{parents}(X_i))$

A Bayesian Network over variables $\{X_1, \ldots, X_n\}$ consists of:

- a DAG whose nodes are the variables
- a set of conditional probability tables (CPTs) giving $P(X_i \mid \text{parents}(X_i))$ for each X_i

The structure of a BN means that every X_i is conditionally independent of all its nondescendants given its parents

$$P(X_i \mid S, \text{parents}(X_i)) = P(X_i \mid \text{parents}(X_i))$$

for any subset $S \subseteq \text{nondescendants}(X_i)$.

The BN defines a factorization of the joint probability distribution. The joint distribution is formed by multiplying the conditional probability tables together

$$P(X_1, \ldots, X_n) = \prod_i P(X_i \mid \text{parents}(X_i))$$

3.1 Correlation and Causality

Directed links in a Bayes' network approximately causality, but this is not always the case.

In a Bayes network, this does not matter, though some structures will be easier to specify.

3.2 Conditional Independence

When no information is given, a later node in a Bayes network is not independent of its ancestors.

When info is given, a later node is conditionally independent of the ancestors of the given info.

The full joint probability can be specified using the local conditional probabilities.