

Probabilistic Graphical Models 10-708

Recitation 1 Handout

1 Probability Overview

Conditional probability:

$$P(X|Y) = \frac{P(X, Y)}{P(Y)}$$

Chain rule:

$$P(X, Y) = P(X|Y)P(Y) = P(Y|X)P(X)$$

Bayes rule (comes from definition of conditional probability + chain rule):

$$P(X|Y) = \frac{P(X, Y)}{P(Y)} = \frac{P(Y|X)P(X)}{P(Y)}$$

Marginalization:

$$P(X) = \sum_y P(X, Y = y)$$

Law of total probability (comes from marginalization + chain rule):

$$P(X) = \sum_y P(X, Y = y) = \sum_y P(X | Y = y)P(Y = y)$$

Useful property:

$$P(X|Y_1, Y_2) = \frac{P(X, Y_1, Y_2)}{P(Y_1, Y_2)} = \frac{P(X, Y_1, Y_2)}{\sum_x P(X = x, Y_1, Y_2)}$$

2 Bayesian Networks

2.1 Independencies in BNs

- Types of three-variable structures: chain (aka causal trail or evidential trail), common parent (aka common cause), v-structure (aka common effect)
- Property: A variable X is independent of its non-descendants given its parents
- Property: A variable X is independent of all other variables in the network given its Markov blanket, which consists of its parents, its children, and its co-parents

2.2 I-Maps

- I-Map: A graph G is an I-map for a distribution P if $I(G) \subseteq I(P)$
- Minimal I-Map: A graph G is a minimal I-map for a distribution P if you cannot remove any edges from G and have it still be an I-map for P
- Perfect I-Map: A graph G is a perfect I-map for a distribution P if $I(G) = I(P)$
- I-Equivalence: Two graphs G_1 and G_2 are I-equivalent if $I(G_1) = I(G_2)$

2.3 Parameterization of BNs

- Bayesian networks are parameterized by a set of conditional probability distributions of the form $P(X \mid \text{parents of } X)$
- The joint probability distribution of the variables in a BN can be written in factorized form as a product of CPDs, i.e. $P(X_1, \dots, X_n) = \prod_i P(X_i \mid \text{parents of } X_i)$

3 Markov Random Fields

3.1 Independencies in MRFs

- Two variables X and Y are independent if there is no active trail between them; a trail is active if it doesn't contain any observed variables
- Property: A variable X is independent of all other variables in the network given its Markov blanket, which consists of its direct neighbors in the graph

3.2 Parameterization of MRFs

- Markov random fields are parameterized by a set of factors defined over cliques in the graph; factors are not distributions as they do not have to sum to 1
- The joint probability distribution of the variables in an MRF can be written in factorized form as a normalized product of factors, i.e. $P(X_1, \dots, X_n) = \frac{1}{Z} \phi_i(C_i)$ where C_i is the set of variables in the i th clique, and Z is the partition function