Graphical Models

St. Joseph's University, Bengaluru

ML 2 BDA3321

Prerequisites

Probabilistic Ideas

Chain Rule

We can always represent a joint distribution as follows:

$$p(x_1, x_2 \ldots x_v) = p(x_1)p(x_2|x_1)p(x_3|x_2, x_1) \ldots p(x_v|x_{v-1} \ldots, x_1)$$

Conditional Independence

Two variables X and Y are conditionally independent given Z iff the conditional joit can be written as a product of conditional marginals, i.e.,

$$X \perp Y|Z \Leftrightarrow p(X, Y|Z) = p(X|Z)p(Y|Z)$$

Prerequisites

Graph Terminology I

Graph $G = (\mathcal{V}, \mathcal{E})$ consists of set of **nodes**, $\mathcal{V} = \{1, \ldots V\}$, and a set of **edges**, $\mathcal{E} = \{(s, t) : s, t \in \mathcal{V}\}$

Adjacency matrix A way to represent a graph in which we write $G(s, t) = 1$ to denote $(s, t) \in \mathcal{E}$

Parent $pa(s) \triangleq \{t : G(t, s) = 1\}$

Child $ch(s) \triangleq \{t : G(s, t) = 1\}$

Family $fam(s) = \{s\} \cup pa(s)$

Root A node with no parents.

Leaf A node with no children.

DAG This is a Directed Acyclic Graph.

Prerequisites

Graph Terminology II

Topological Ordering For a DAG, topological ordering is a numbering of nodes such that parents have lower numbers than their children.

Tree Undirected graph with no cycles, or, directed graph with no directed cycles.

Forest Set of trees.

Subgraph G_A is the graph created by using the nodes in A and their corresponding edges, $G_A = (\mathcal{V}_A, \mathcal{E}_A)$

Clique A set f nodes that are all neighbours of each other.

Directed Graphical Models

Ordered Markov Property

Given a topological order, this is the assumption tht a node only depends on it's immediate parents, not all the predecessors in the ordering, i.e.

$$x_s \perp \mathbf{x}_{pred(s)\ \ pa(s)} | \mathbf{x}_{pa(s)}$$

In general, for a DAG we have

$$p(\mathbf{x}_{1:V}|G) = \prod_{t=1}^{V} p(x_t|\mathbf{x}_{pa(t)})$$

DAGs: Examples

Naive Bayes Classifiers

Assumption

The features are conditionally independent given the class label. This allows us to write the joint distribution as follows:

$$p(y, \mathbf{x}) = p(y) \prod_{j=1}^{D} p(x_j | y)$$

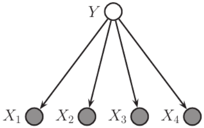

Figure: [Mur12]

References I

[Mur12] Kevin P Murphy. *Machine Learning: A Probabilistic Perspective*. MIT Press, 2012.

[Mar14] Stephen Marsland. *Machine Learning, An Algorithmic Perspective*. CRC Press, 2014.

[Ian17] Aaron Courville Ian Goodfellow Yoshua Bengio. *Deep Learning*. MIT Press, 2017.