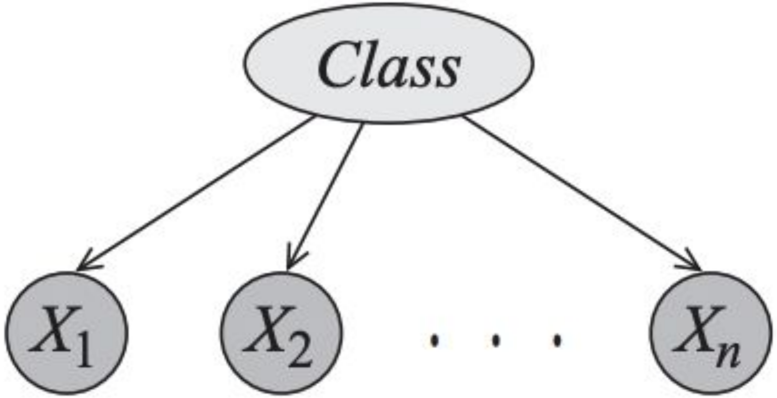
Advanced Machine Learning

Likhit Nayak

Naive Bayes Model



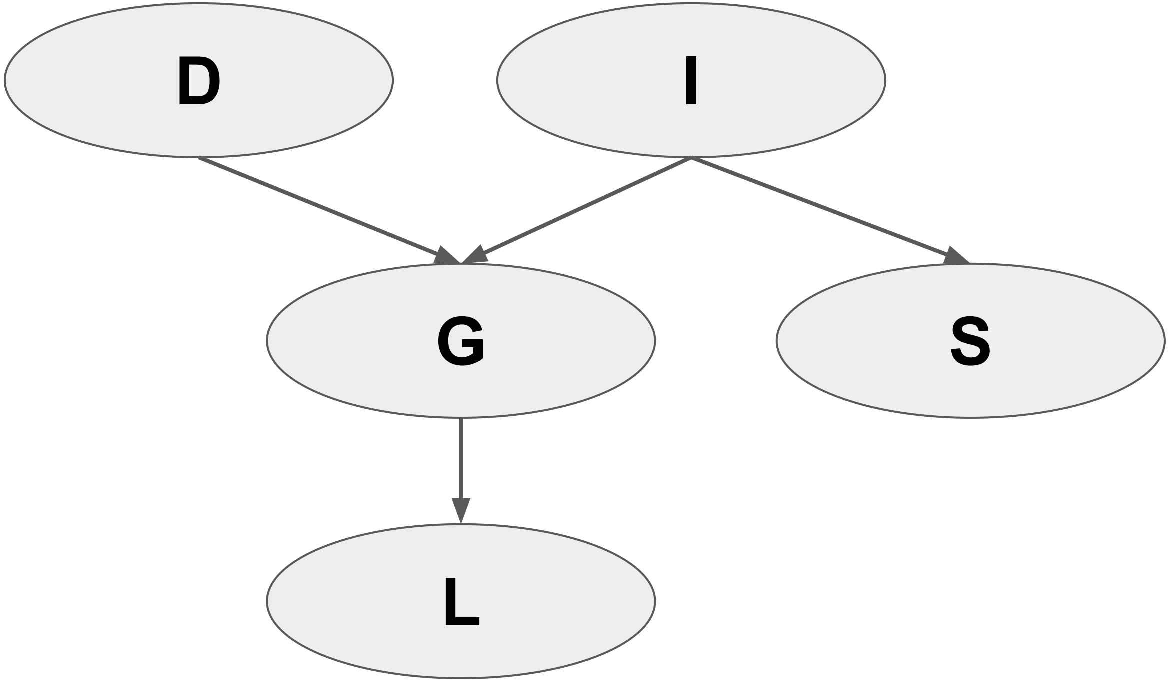
$$P(C, X_1, \ldots, X_n) = P(C) \prod_{i=1}^{n} P(X_i \mid C).$$

Bayesian Network

A Bayesian network structure $\mathcal{G}$ *is a directed acyclic graph whose nodes represent random variables* $X_1, \ldots, X_n$. *Let* $\mathrm{Pa}^{\mathcal{G}}_{X_i}$ *denote the parents of* X_i *in* $\mathcal{G}$, *and* $\mathrm{NonDescendants}_{X_i}$ *denote the variables in the graph that are not descendants of* X_i. *Then* $\mathcal{G}$ *encodes the following set of conditional independence assumptions, called the* local independencies, *and denoted by* $\mathcal{I}_\ell(\mathcal{G})$:

$$\text{For each variable } X_i: \quad (X_i \perp \mathrm{NonDescendants}_{X_i} \mid \mathrm{Pa}^{\mathcal{G}}_{X_i}).$$

Bayesian Network



$$(L \perp I, D, S \mid G)$$

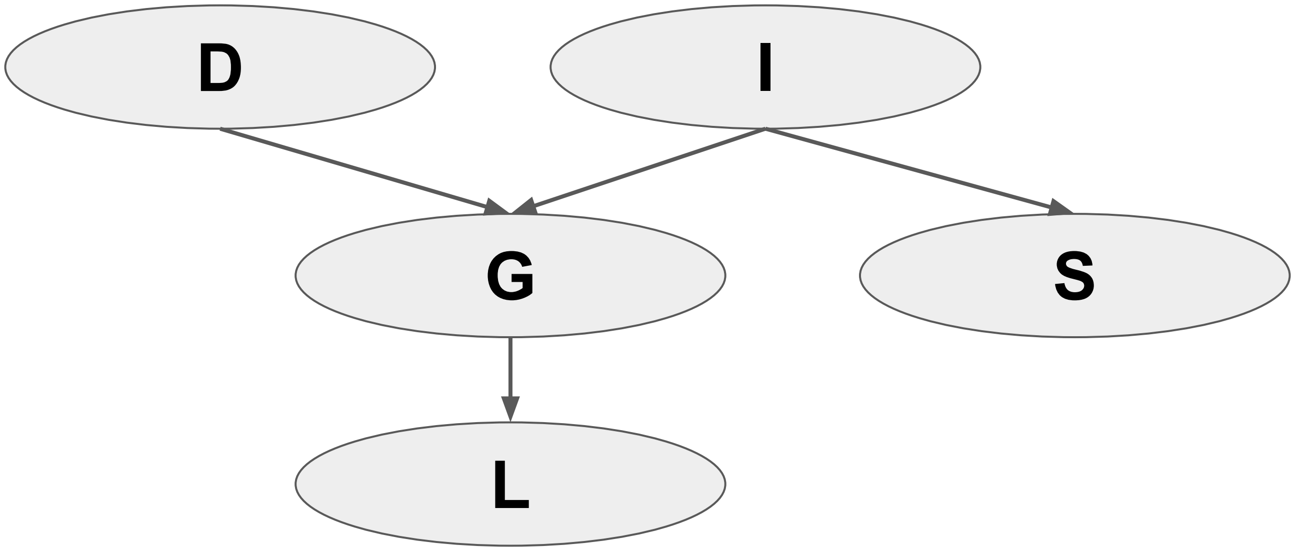
$$(S \perp D, G, L \mid I)$$

$$(G \perp S \mid I, D)$$

$$(I \perp D)$$

$$(D \perp I, S)$$

Bayesian Network



$$P(I, D, G, L, S) = P(I)P(D \mid I)P(G \mid I, D)P(L \mid I, D, G)P(S \mid I, D, G, L)$$

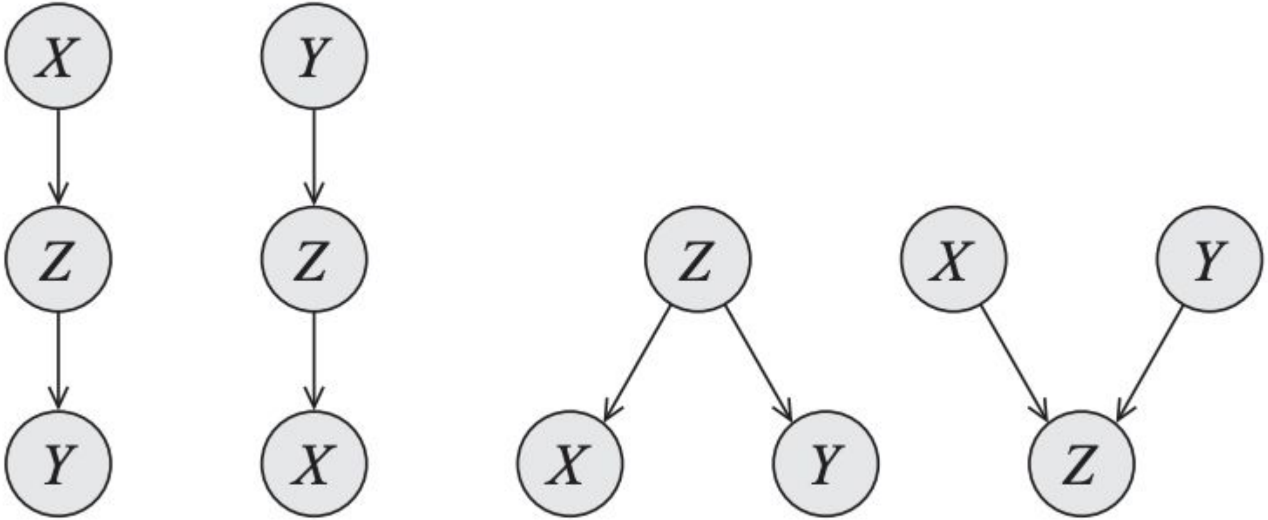
Bayesian Network (BN)

Let $\mathcal{G}$ be a BN graph over the variables $X_1, \ldots, X_n$. We say that a distribution P over the same space factorizes *according to $\mathcal{G}$ if P can be expressed as a product*

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid \mathrm{Pa}^{\mathcal{G}}_{X_i})$$

This equation is called the chain rule for Bayesian networks*. The individual factors $P(X_i \mid \mathrm{Pa}^{\mathcal{G}}_{X_i})$ are called* conditional probability distributions (CPDs) *or* local probabilistic models.

Independencies in Bayesian Networks



(a) (b) (c) (d)

$X \perp Y \mid Z$ $\quad$ $X \perp Y \mid Z$ $\quad$ $X \perp Y \mid Z$ $\quad$ $X \perp Y$

References

Koller, Daphne, and Nir Friedman. Probabilistic graphical models: principles and techniques. MIT press, 2009.