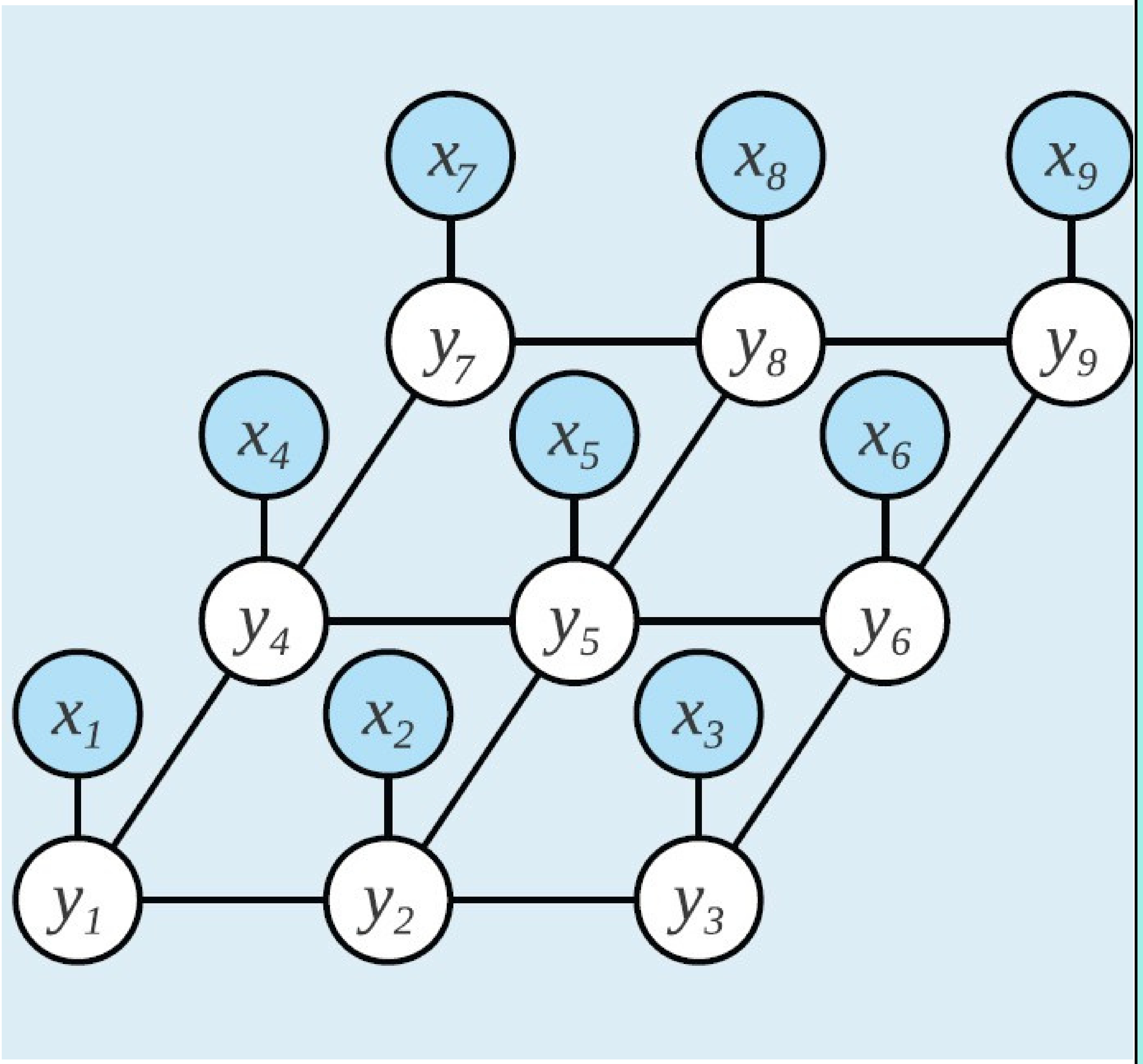
Some Fundamental Rules in Probability

Probability Theory

1. Introduction
2. Sum Rule
3. Product Rule
4. Bayes' Theorem

Introduction

Probabilistic modeling provides a principled foundation for designing machine learning methods. Once we have defined probability distributions corresponding to the uncertainties of the data and our problem, it turns out that there are only two fundamental rules, the **sum rule**, and the **product rule**.



Sum Rule

The sum rule is also known as the **marginalization property**. The sum rule relates the joint distribution to a marginal distribution. In general, when the joint distribution contains more than two random variables, the sum rule can be applied to any subset of the random variables, resulting in a marginal distribution of potentially more than one random variable.

$$p(\boldsymbol{x}) = \begin{cases} \sum_{\boldsymbol{y} \in \mathcal{Y}} p(\boldsymbol{x}, \boldsymbol{y}) & \text{if } \boldsymbol{y} \text{ is discrete} \\ \int_{\mathcal{Y}} p(\boldsymbol{x}, \boldsymbol{y}) \mathrm{d}\boldsymbol{y} & \text{if } \boldsymbol{y} \text{ is continuous} \end{cases}$$

If x = $[x_1, x_2, x_3, \ldots x_n]^T$, we obtain the marginal

$$p(x) = \int p(x_1, x_2, x_3, \ldots x_n)\, dx_{\setminus i}$$

Product Rule

The product rule can be interpreted as the fact that every joint distribution of two random variables can be factorized (written as a product) of two other distributions.

The two factors are the marginal distribution of the first random variable p(x), and the conditional distribution of the second random variable given the first $p(y \mid x)$.

Since the ordering of random variables is arbitrary in p(x, y), the product rule also implies $p(x, y) = p(x \mid y)p(y)$.

- $p(x, y) = p(y \mid x)\, p(x)$
- $p(x, y) = p(x \mid y)\, p(y)$

Bayes' Theorem

In machine learning and Bayesian statistics, we are often interested in making inferences of unobserved (latent) random variables given that we have observed other random variables.

Let us assume we have some prior knowledge p(x) about an unobserved random variable x and some relationship p(y | x) between x and a second random variable y, which we can observe.

If we observe y, we can use Bayes' theorem to draw some conclusions about x given the observed values of y.

$$\underbrace{P(A \mid B)}_{\text{Posterior}} = \frac{\overbrace{P(B \mid A)}^{\text{Likelihood}}\,\overbrace{P(A)}^{\text{Porior}}}{\underbrace{P(B)}_{\text{Evidence}}}$$