Machine Learning for Graphs and Sequential Data

Recap: Bayesian Networks

lecturer: Prof. Dr. Stephan Günnemann
www.daml.in.tum.de

Summer Term 2020

Data Analytics and
Machine Learning

TUM

Challenge

How to model $\Pr(X_1, \dots, X_T)$, e.g. a distribution over sequences?

- Assume that X_t can take, e.g., values in $\{1, \dots, K\}$.

- Challenge: In general, we need $K^T - 1$ parameters to specify a distribution over the sequence $X_1, \dots, X_T$.

- To have a more compact representation and to reduce the number of parameters, we make additional assumptions about the joint distribution $\Pr(X_1, \dots, X_T)$.

  - These assumptions should be a natural-fit for the data/task.

  - We use **Bayesian networks** (i.e. directed graphical models) to illustrate the assumptions on $\Pr(X_1, \dots, X_T)$.

Data Analytics and
Machine Learning

Directed Graphical Models (Bayesian Networks)

- Bayesian network is a way to represent a joint distribution via a directed acyclic graph $G = (V, E)$.

  - Each node represents a random variable.

  - Directed edges are often interpreted as causal relations.
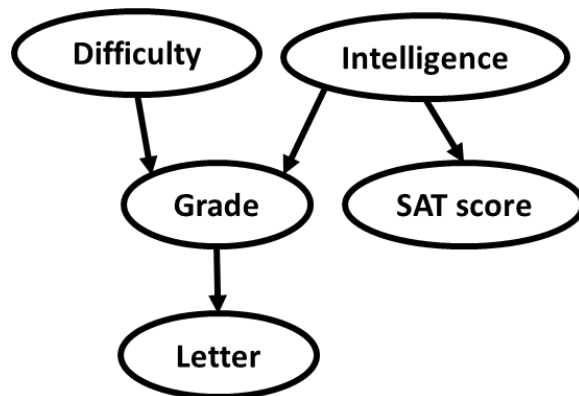
$variables \in \{0: low, 1: high\}$

- Graph G provides:

  1) a particular factorization for the joint distribution $Pr(X_1, \dots, X_{|V|})$.

  2) a set of conditional independencies, inferred by a routine $d\_separation(G)$.

Example from: [Daphne Koller and Nir Friedman. 2009. Probabilistic Graphical Models: Principles and Techniques. The MIT Press].
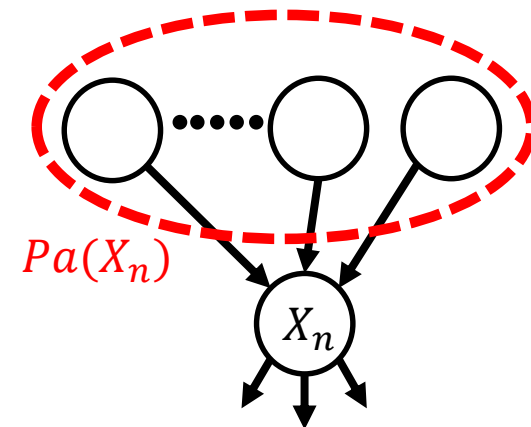
Recall: $X \perp Y | Z \iff \Pr(X, Y | Z) = \Pr(X | Z) \Pr(Y | Z)$

Data Analytics and
Machine Learning

Bayesian Networks - Factorization

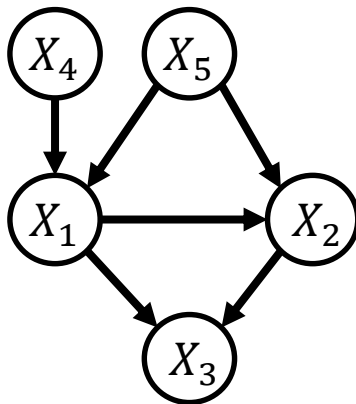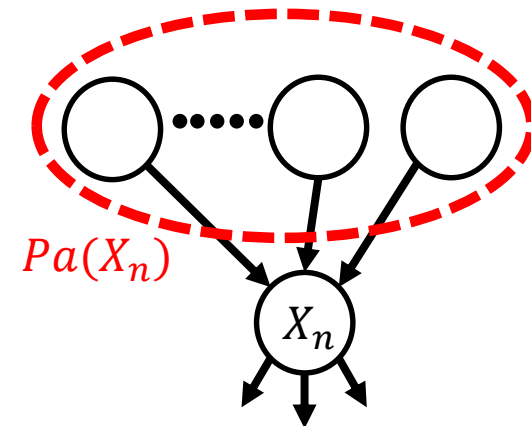- A Bayesian network specifies the following factorization for the joint distribution $\Pr(X_1, \ldots, X_{|V|})$.

$$\Pr(X_1, \ldots, X_{|V|}) = \prod_{n=1}^{|V|} \Pr(X_n | Pa(X_n))$$

  - $Pa(X_n)$: parents of node X_n in Graph G.

- Example:

$$\Pr(X_1, X_2, X_3, X_4, X_5) =$$

Data Analytics and
Machine Learning
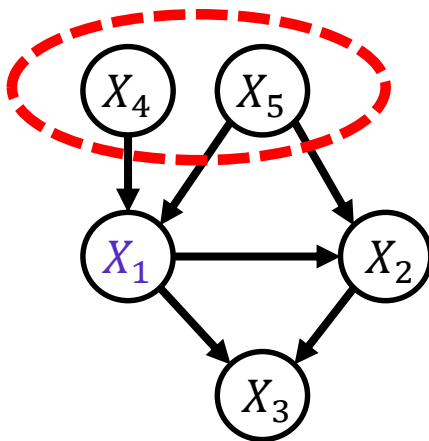
Bayesian Networks - Factorization

- A Bayesian network specifies the following factorization for the joint distribution $\Pr(X_1, \ldots, X_{|V|})$.

$$\Pr(X_1, \ldots, X_{|V|}) = \prod_{n=1}^{|V|} \Pr(X_n | Pa(X_n))$$

  – $Pa(X_n)$: parents of node X_n in Graph G.

- Example:

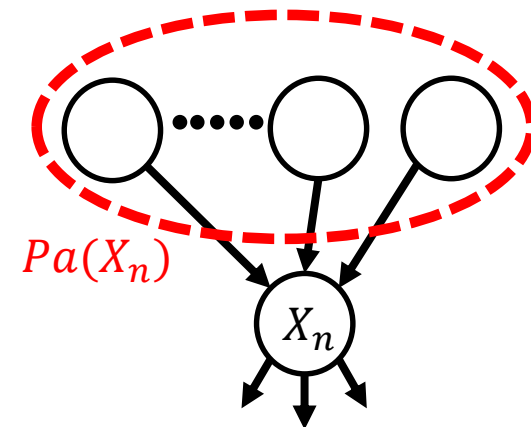$$\Pr(X_1, X_2, X_3, X_4, X_5) =$$

$$\Pr(X_1 | X_4, X_5) \times$$

$$Pa(X_1) = \{X_4, X_5\}$$

Data Analytics and
Machine Learning
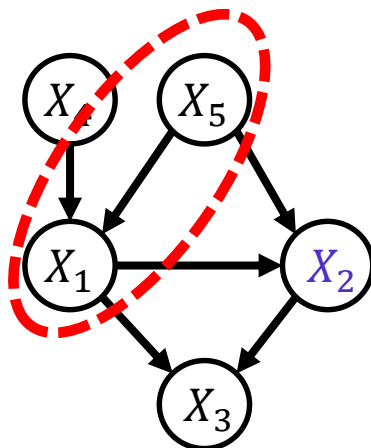
Bayesian Networks - Factorization

■ A Bayesian network specifies the following factorization for the joint distribution $\Pr(X_1, \ldots, X_{|V|})$.

$$\Pr(X_1, \ldots, X_{|V|}) = \prod_{n=1}^{|V|} \Pr(X_n | Pa(X_n))$$

 – $Pa(X_n)$: parents of node X_n in Graph G.

$Pa(X_n)$

■ Example:

$\Pr(X_1, X_2, X_3, X_4, X_5) =$

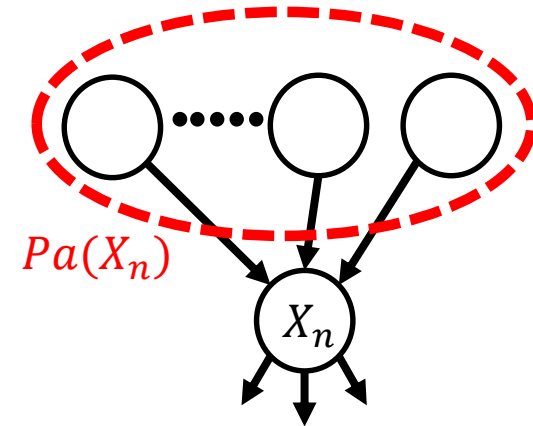$\Pr(X_1 | X_4, X_5) \times \Pr(X_2 | X_1, X_5) \times$

$Pa(X_2) = \{X_1, X_5\}$
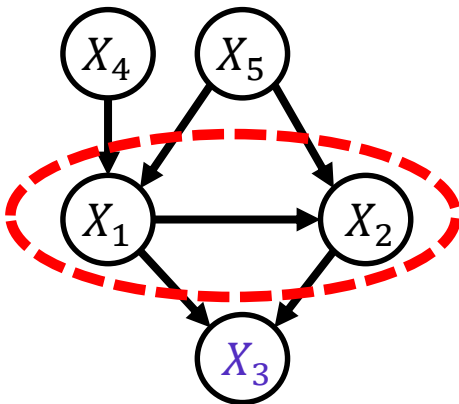
Bayesian Networks - Factorization

- A Bayesian network specifies the following factorization for the joint distribution $\Pr(X_1, \ldots, X_{|V|})$.

$$\Pr(X_1, \ldots, X_{|V|}) = \prod_{n=1}^{|V|} \Pr(X_n | Pa(X_n))$$

  - $Pa(X_n)$: parents of node X_n in Graph G.

- Example:

$$\Pr(X_1, X_2, X_3, X_4, X_5) =$$

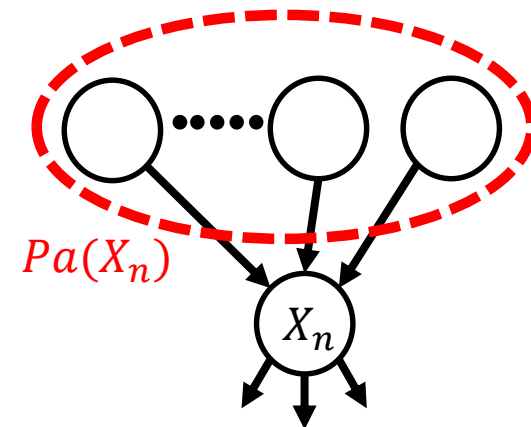$$\Pr(X_1 | X_4, X_5) \times \Pr(X_2 | X_1, X_5) \times \Pr(X_3 | X_1, X_2) \times$$

$$Pa(X_3) = \{X_1, X_2\}$$

Data Analytics and
Machine Learning
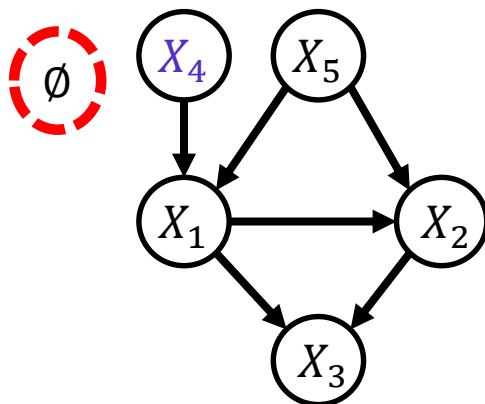
Bayesian Networks - Factorization

- A Bayesian network specifies the following factorization for the joint distribution $\Pr(X_1, \ldots, X_{|V|})$.

$$\Pr(X_1, \ldots, X_{|V|}) = \prod_{n=1}^{|V|} \Pr(X_n | Pa(X_n))$$

  - $Pa(X_n)$: parents of node X_n in Graph G.

$Pa(X_n)$

X_n

- Example:

$\emptyset$

X_4 X_5

X_1 X_2

X_3

$$\Pr(X_1, X_2, X_3, X_4, X_5) =$$

$$\Pr(X_1 | X_4, X_5) \times \Pr(X_2 | X_1, X_5) \times \Pr(X_3 | X_1, X_2) \times \Pr(X_4) \times$$
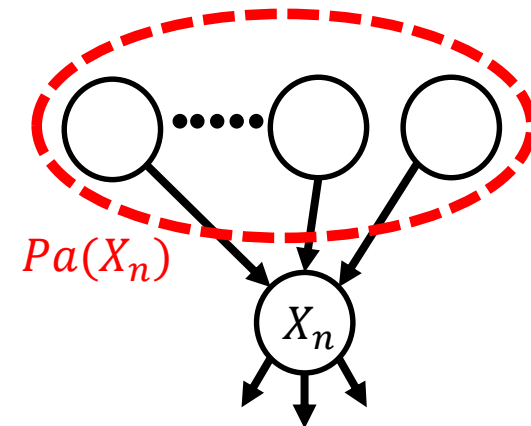
$$Pa(X_4) = \emptyset$$
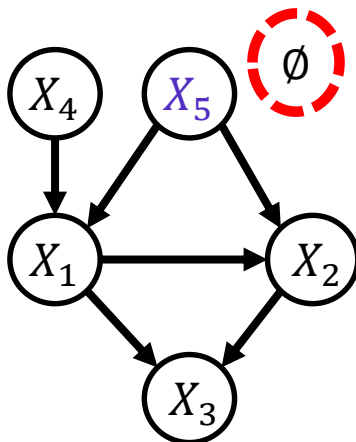
Bayesian Networks - Factorization

- A Bayesian network specifies the following factorization for the joint distribution $\Pr(X_1, \ldots, X_{|V|})$.

$$\Pr(X_1, \ldots, X_{|V|}) = \prod_{n=1}^{|V|} \Pr(X_n | Pa(X_n))$$

  - $Pa(X_n)$: parents of node X_n in Graph G.

$Pa(X_n)$

- Example:

$$\Pr(X_1, X_2, X_3, X_4, X_5) =$$

$$\Pr(X_1 | X_4, X_5) \times \Pr(X_2 | X_1, X_5) \times \Pr(X_3 | X_1, X_2) \times \Pr(X_4) \times \Pr(X_5)$$
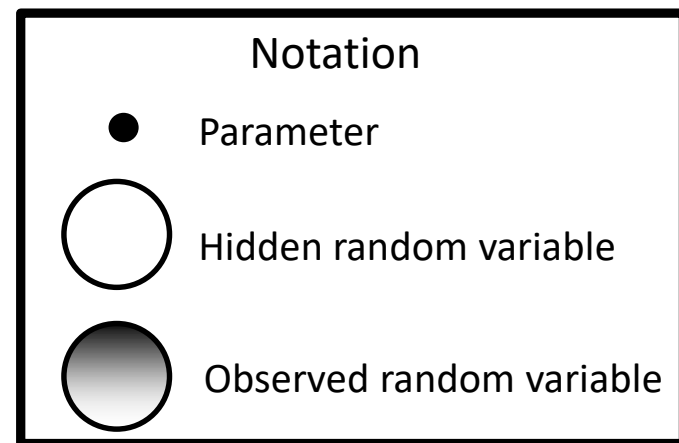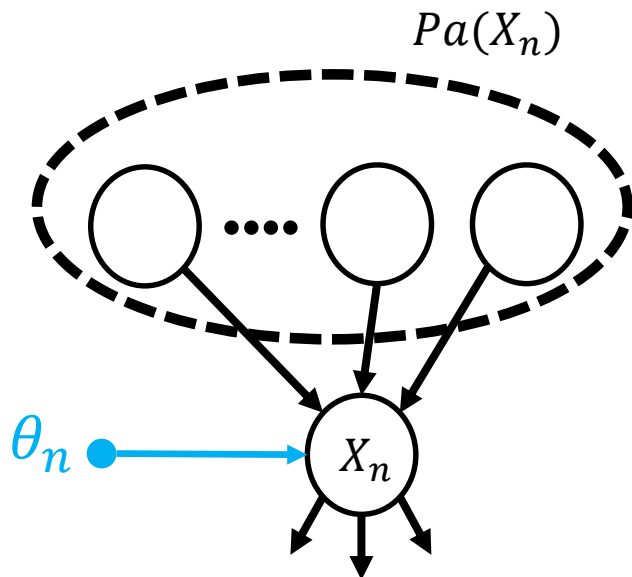
$$Pa(X_5) = \emptyset$$

Bayesian Networks - Model Parameters

- A set of parameters $\{\theta_1, \dots, \theta_{|V|}\}$ parametrizes the joint distribution.

- θ_n parametrizes the factor $\Pr(X_n | Pa(X_n))$

  - Indeed, the factor is a function of θ_n as well.
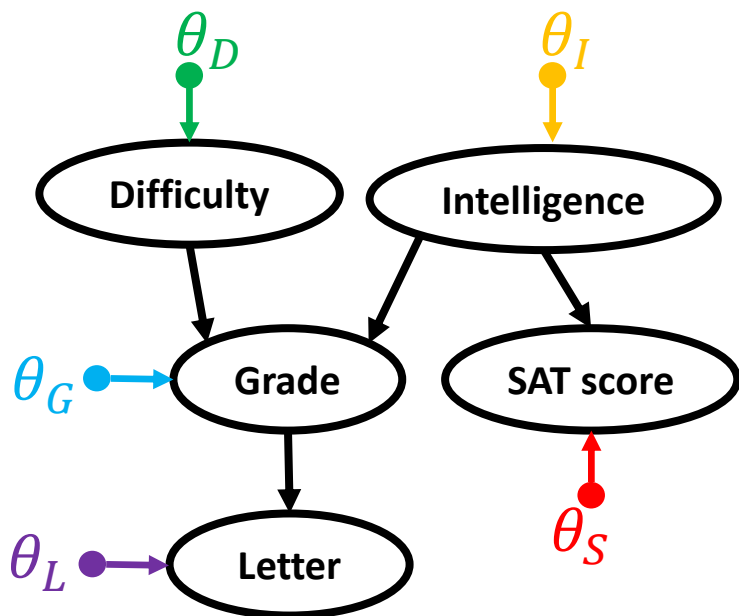
    We can write $\Pr(X_n | Pa(X_n) ; \theta_n)$.

  - In a graphical model, we show θ_n using a filled circle connected to X_n.

Bayesian Networks - Model Parameters

- Example:

$variables \in \{0: low, 1: high\}$

$\Pr(D = 0)$	$\Pr(D = 1)$
0.5	0.5

$\Pr(I = 0)$	$\Pr(I = 1)$
0.5	0.5

D	I	$\Pr(G = 0|D, I)$	$\Pr(G = 1|D, I)$
0	0	0.5	0.5
0	1	0.1	0.9
1	0	0.9	0.1
1	1	0.5	0.5

I	$\Pr(S = 0|I)$	$\Pr(S = 1|I)$
0	0.9	0.1
1	0.1	0.9

G	$\Pr(L = 0|G)$	$\Pr(L = 1|G)$
0	0.9	0.1
1	0.1	0.9

Example from: [Daphne Koller and Nir Friedman. 2009. Probabilistic Graphical Models: Principles and Techniques. The MIT Press].

Bayesian Networks – Generative Process

- We can think of the model as a generative process
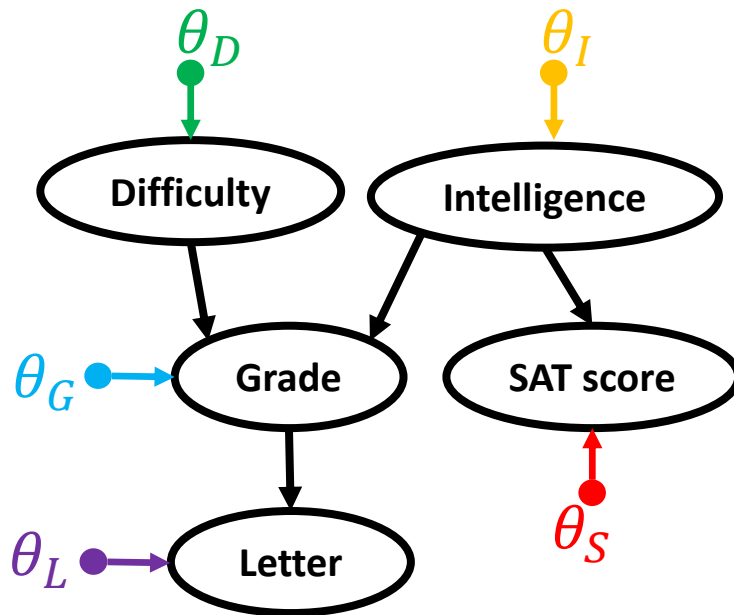
$$Difficulty \sim p(D|\theta_D)$$
$$Intelligence \sim p(I|\theta_I)$$
$$Grade \sim p(G|D, I, \theta_G)$$
$$SAT \sim p(S|I, \theta_S)$$
$$Letter \sim p(L|G, \theta_L)$$

Example from: [Daphne Koller and Nir Friedman. 2009. Probabilistic
Graphical Models: Principles and Techniques. The MIT Press].

Data Analytics and
Machine Learning