Lecture 7

Независимости в байесовских сетях

Machine Learning

Ivan Smetannikov

15.04.2021

Lecture plan

- Условная независимость
- Независимость в Байесовских сетях
- Наивный Баес

Lecture plan

- Условная независимость
- Независимость в Байесовских сетях
- Наивный Баес

Независимость

Для событий $\alpha, \beta, P \overset{\text{удовлетворяет}}{\models} \alpha \overset{\text{независимы}}{\perp} \beta$ if:

Independence

For events $\alpha, \beta, P \overset{\text{satisfies}}{\vDash} \alpha \overset{\text{independence}}{\perp} \beta$ if:

- $P(\alpha, \beta) = P(\alpha)P(\beta)$
- $P(\alpha|\beta) = P(\alpha)$
- $P(\beta|\alpha) = P(\beta)$

For random variables $X, Y, P \vDash X \perp Y$ if:

- $P(X, Y) = P(X)P(Y)$ Universal or Factors
- $P(X|Y) = P(X)$
- $P(Y|X) = P(Y)$

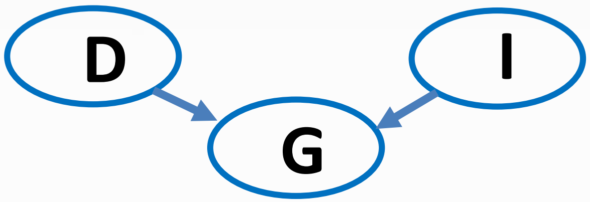
Independence

I	D	G	P(I,D,G)
i^0	d^0	g^1	0.126
i^0	d^0	g^2	0.168
i^0	d^0	g^3	0.126
i^0	d^1	g^1	0.009
i^0	d^1	g^2	0.045
i^0	d^1	g^3	0.126
i^1	d^0	g^1	0.252
i^1	d^0	g^2	0.0224
i^1	d^0	g^3	0.0056
i^1	d^1	g^1	0.06
i^1	d^1	g^2	0.036
i^1	d^1	g^3	0.024

I	D	P(I,D)
i^0	d^0	0.42
i^0	d^1	0.18
i^1	d^0	0.28
i^1	d^1	0.12

Independence

I	D	G	P(I,D,G)
i^0	d^0	g^1	0.126
i^0	d^0	g^2	0.168
i^0	d^0	g^3	0.126
i^0	d^1	g^1	0.009
i^0	d^1	g^2	0.045
i^0	d^1	g^3	0.126
i^1	d^0	g^1	0.252
i^1	d^0	g^2	0.0224
i^1	d^0	g^3	0.0056
i^1	d^1	g^1	0.06
i^1	d^1	g^2	0.036
i^1	d^1	g^3	0.024



P(I,D)=P(I)P(D)

I	D	P(I,D)
i^0	d^0	0.42
i^0	d^1	0.18
i^1	d^0	0.28
i^1	d^1	0.12

I	P(I)
i^0	0.6
i^1	0.4

D	P(D)
d^0	0.7
d^1	0.3

Conditional Independence

For (sets of) random variables X, Y, Z

$P \vDash (X \perp Y | Z)$ if:

Conditional Independence

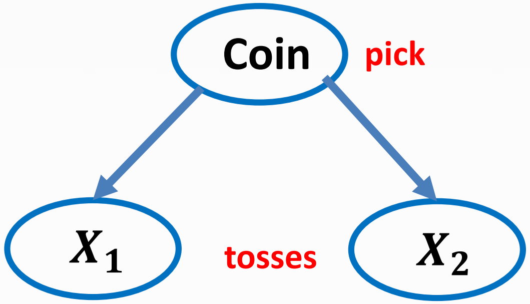
For (sets of) random variables X, Y, Z

$P \vDash (X \perp Y | Z)$ if:

- $P(X, Y | Z) = P(X|Z)P(Y|Z)$
- $P(X|Y, Z) = P(X|Z)$
- $P(Y|X, Z) = P(Y|Z)$
- $P(X, Y, Z) \propto \phi_1(X, Z)\phi(Y, Z)$

proportional

Conditional Independence

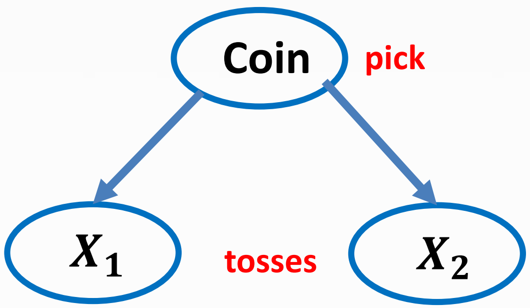


Two coins:

- Normal
- Biased with heads coming 90% of the time

- Pick a random coin, toss gives heads. Probability of heads on the second toss?

Conditional Independence

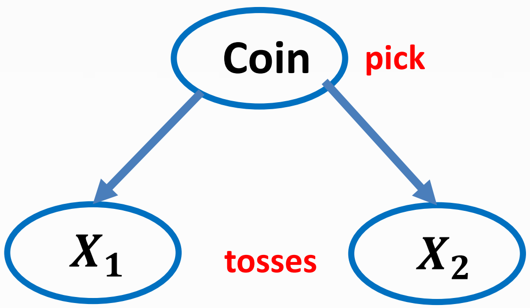


Two coins:

- Normal
- Biased with heads coming 90% of the time

- Pick a random coin, toss gives heads. Probability of heads on the second toss? It increases.
- Pick a Normal (or Biased) coin directly, toss gives heads. Probability of heads on the second toss? The same.

Conditional Independence



Two coins:

- Normal
- Biased with heads coming 90% of the time

- Pick a random coin, toss gives heads. Probability of heads on the second toss? It increases.
- Pick a Normal (or Biased) coin directly, toss gives heads. Probability of heads on the second toss? The same.

$$P \not\models (X_1 \perp X_2)$$
$$P \models (X_1 \perp X_2 | C)$$

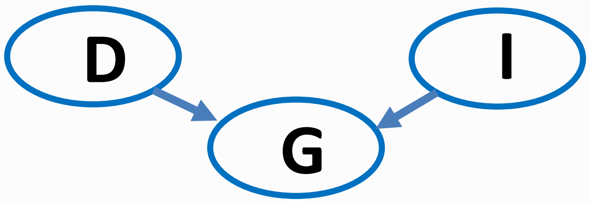
Conditioning can Lose Independences

I	D	G	P(I,D,G)
i^0	d^0	g^1	0.126
i^0	d^0	g^2	0.168
i^0	d^0	g^3	0.126
i^0	d^1	g^1	0.009
i^0	d^1	g^2	0.045
i^0	d^1	g^3	0.126
i^1	d^0	g^1	0.252
i^1	d^0	g^2	0.0224
i^1	d^0	g^3	0.0056
i^1	d^1	g^1	0.06
i^1	d^1	g^2	0.036
i^1	d^1	g^3	0.024

I	P(I|i^0)
i^0	0.6
i^1	0.4

I	D	P(I,D|i^0)
i^0	d^0	0.282
i^0	d^1	0.02
i^1	d^0	0.564
i^1	d^1	0.134

D	P(D|i^0)
d^0	0.7
d^1	0.3



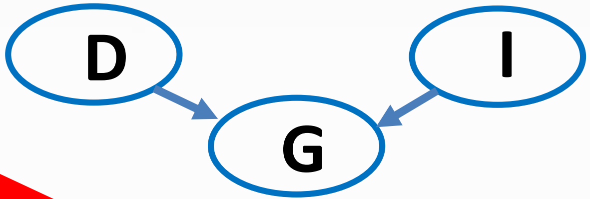
Conditioning can Lose Independences

I	D	G	P(I,D,G)
i^0	d^0	g^1	0.126
i^0	d^0	g^2	0.168
i^0	d^0	g^3	0.126
i^0	d^1	g^1	0.009
i^0	d^1	g^2	0.045
i^0	d^1	g^3	0.126
i^1	d^0	g^1	0.252
i^1	d^0	g^2	0.0224
i^1	d^0	g^3	0.0056
i^1	d^1	g^1	0.06
i^1	d^1	g^2	0.036
i^1	d^1	g^3	0.024

I	P(I|i^0)
i^0	0.6
i^1	0.4

I	D	P(I,D|i^0)
i^0	d^0	0.282
i^0	d^1	0.02
i^1	d^0	0.564
i^1	d^1	0.134

D	P(D|i^0)
d^0	0.7
d^1	0.3



I	P(I|i^0)
i^0	0.284
i^1	0.698

D	P(D|i^0)
d^0	0.846
d^1	0.154

Lecture plan

- Условная независимость
- Независимость в Байесовских сетях
- Наивный Баес

Independence and Factorization

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

X, Y independent

$P(X, Y, Z) \propto \phi_1(X, Z)\phi(Y, Z)$

$(X \perp Y | Z)$

Independence and Factorization

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

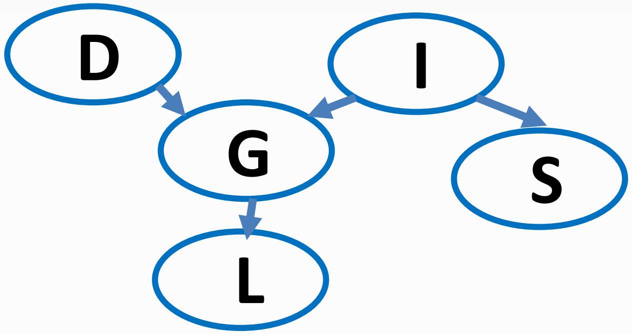
X,Y independent

$$P(X,Y,Z) \propto \phi_1(X,Z)\phi(Y,Z)$$

$$(X \perp Y | Z)$$

- Factorization of a distribution P implies independencies that hold in P
- If P factorizes over G, can we read these independencies from the structure of G?

Flow of influence and d-separation

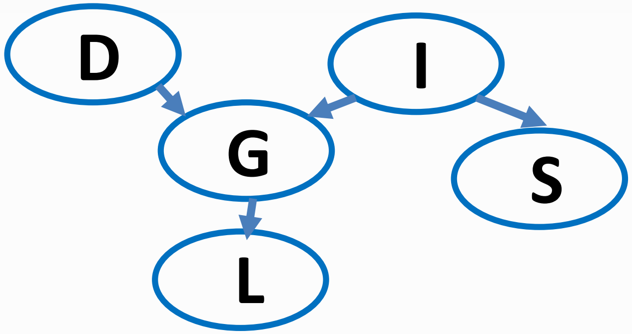


Definition: ***X*** and ***Y*** are <u>d-separated</u> in ***G*** given ***Z*** if there is no active trail in ***G*** between ***X*** and ***Y*** given ***Z***

Notation: $dsep_G(X, Y|Z)$

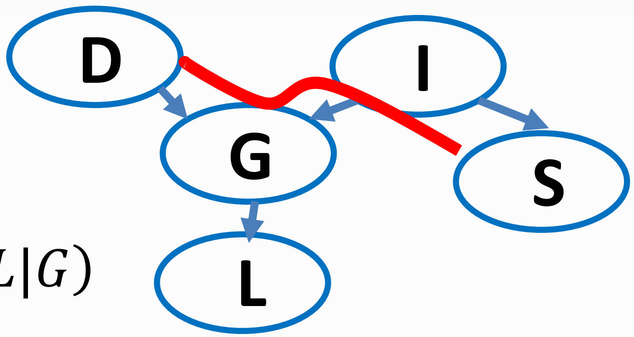
Factorization $\Longrightarrow$ Independence: BNs

Theorem: If P factorizes over G, and $dsep_G(X, Y|Z)$ then P satisfies $(X \perp Y|Z)$



Factorization $\Longrightarrow$ Independence: BNs

Theorem: If P factorizes over G, and $dsep_G(X, Y|Z)$ then P satisfies $(X \perp Y|Z)$

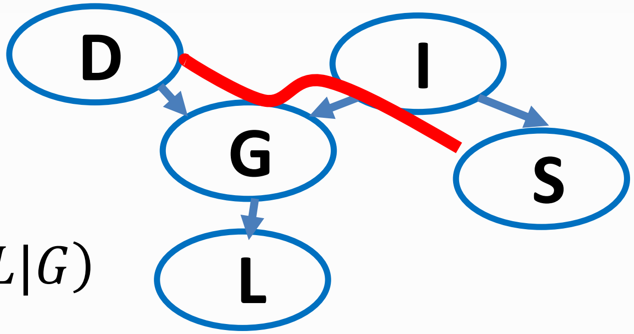


chain rule

$$P(D, I, G, S, L) = P(D)P(I)P(G|D, I)P(D|I)P(L|G)$$

Factorization $\Longrightarrow$ Independence: BNs

Theorem: If P factorizes over G, and $dsep_G(X, Y|Z)$ then P satisfies $(X \perp Y|Z)$



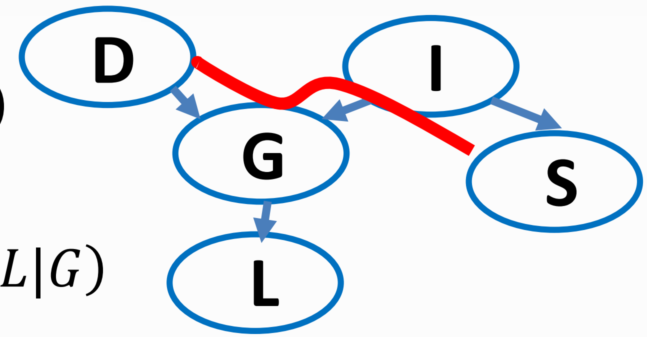
chain rule

$$P(D, I, G, S, L) = P(D)P(I)P(G|D, I)P(D|I)P(L|G)$$

$$P(D, S) = \sum_{G,L,I} P(D)P(I)P(G|D, I)P(S|I)P(L|G) =$$
$$\sum_I P(D)P(I)P(S|I) \sum_G (P(G|D, I) \sum_L P(L|G)) =$$

Factorization $\Longrightarrow$ Independence: BNs

Theorem: If P factorizes over G, and $dsep_G(X, Y|Z)$ then P satisfies $(X \perp Y|Z)$



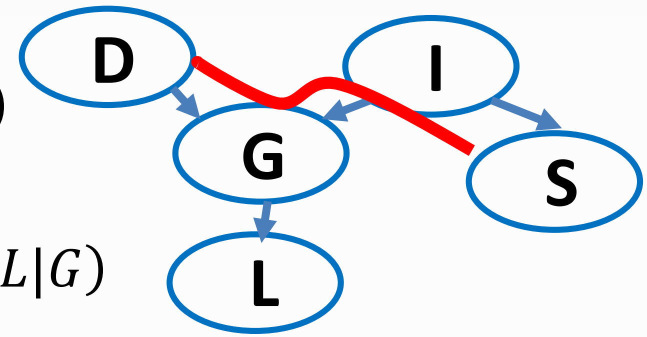
chain rule

$$P(D,I,G,S,L) = P(D)P(I)P(G|D,I)P(D|I)P(L|G)$$

$$P(D,S) = \sum_{G,L,I} P(D)P(I)P(G|D,I)P(S|I)P(L|G) =$$
$$\sum_I P(D)P(I)P(S|I) \sum_G (P(G|D,I) \sum_L P(L|G)) =$$
$$P(D)(\sum_I P(I)P(S|I))$$

Factorization $\Longrightarrow$ Independence: BNs

Theorem: If P factorizes over G, and $dsep_G(X, Y|Z)$ then P satisfies $(X \perp Y|Z)$



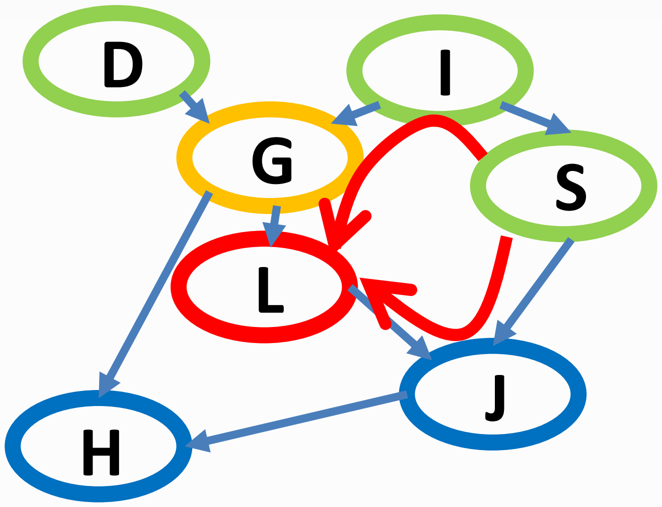
$$P(D, I, G, S, L) = \overset{\textbf{chain rule}}{P(D)P(I)P(G|D, I)P(D|I)P(L|G)}$$

$$P(D, S) = \sum_{G,L,I} P(D)P(I)P(G|D, I)P(S|I)P(L|G) =$$
$$\sum_I P(D)P(I)P(S|I) \sum_G (P(G|D, I) \sum_L P(L|G)) =$$
$$\underbrace{P(D)}_{\phi_1(D)}(\underbrace{\sum_I P(I)P(S|I)}_{\phi_2(S)})$$

Flow of influence and d-separation

Any node is d-separated from its non-descendants given its parents

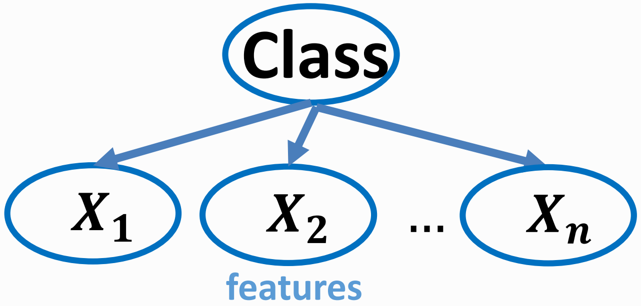
If P factorizes over G, then in P, any **variable** is independent of its **non-descendants** given its **parents**



Lecture plan

- Условная независимость
- Независимость в Байесовских сетях
- Наивный Баес

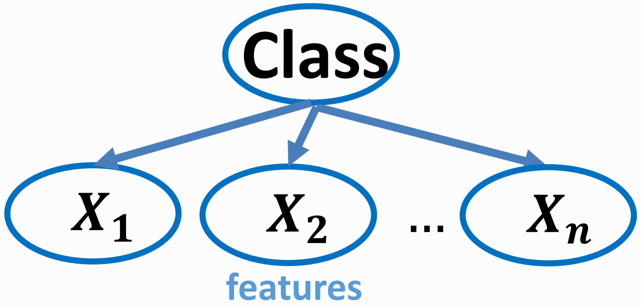
Naïve Bayes Model



$\left(X_i \perp X_j | C\right)$ for all X_i, X_j

X_i, X_j are conditionally independent given C

Naïve Bayes Model

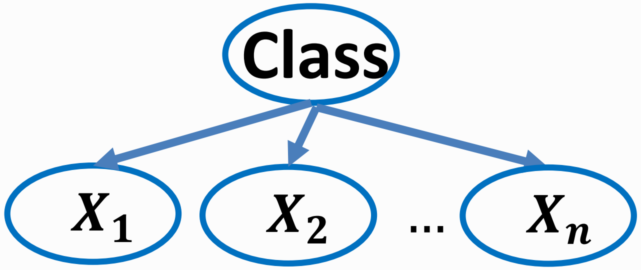


$\left(X_i \perp X_j | C\right)$ for all X_i, X_j

X_i, X_j are conditionally independent given C

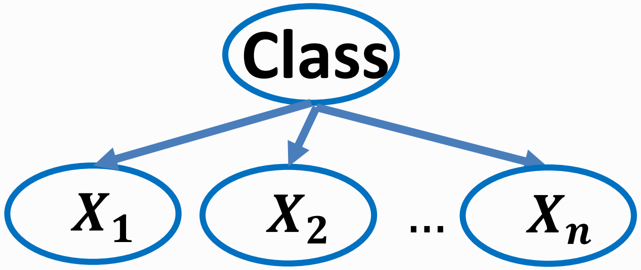
$$P(C, X_1, \dots, X_n) = P(C) \sum_{i=1}^{n} P(X_i | C)$$

Naïve Bayes Model



$$\frac{P(C = c^1 | x_1, \ldots, x_n)}{P(C = c^2 | x_1, \ldots, x_n)} =$$

Naïve Bayes Model



odds ratios

$$\frac{P(C=c^1|x_1,\ldots,x_n)}{P(C=c^2|x_1,\ldots,x_n)} = \frac{P(C=c^1)}{P(C=c^2)}\prod_{i=1}^{n}\frac{P(x_i|C=c^1)}{P(x_i|C=c^2)}$$

Summary

- Simple approach for classification
  - Computationally efficient
  - Easy to construct
- Surprisingly effective in domains with many weakly relevant features
- Strong independence assumptions reduce performance when many features are strongly correlated