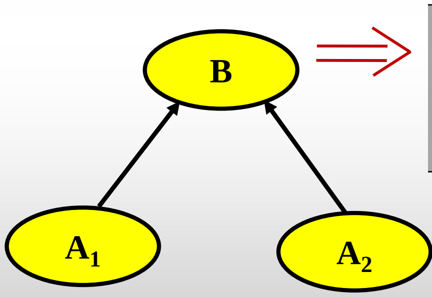
Elementy podstaw teorii sieci Bayesa

Katedra Sztucznej Inteligencji

WYŻSZA SZKOŁA INFORMATYKI i ZARZĄDZANIA w RZESZOWIE

- Sieć Bayesowska składa się ze zbioru *zmiennych* oraz zbioru *skierowanych łuków* pomiędzy zmiennymi
- Zmienne wraz ze skierowanymi łukami tworzą *acy-kliczny graf skierowany*
- Z każdą strukturą składającą się ze zmiennej B i jej rodziców $A_1, \ldots, A_n$ związana jest potencjalna tablica $P(B \mid A_1, \ldots, A_n)$



		B	
A_1	A_2	Jest	Nie_ma
Jest	Jest	0.4	0.6
Nie_ma	Jest	0.5	0.5
Jest	Nie_ma	0.2	0.8
Nie_ma	Nie_ma	0.7	0.3

$$P(A_1, A_2, B) = P(A_1)P(A_2)\boxed{P(B|A_1, A_2)}$$

[F.V. Jensen: *Bayesian Networks and Decision Graphs*, Springer-Verlag, Berlin Heidelberg 2002, pp. 19]

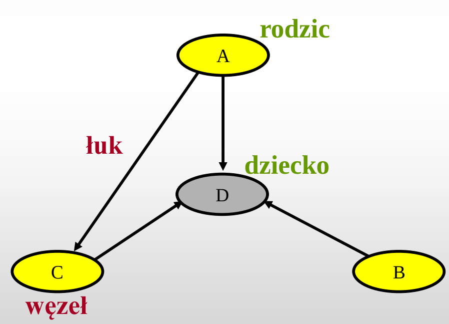
Sieć Bayesa

- Składnik jakościowy
- Składnik ilościowy

Sieć Bayesa

- Składnik jakościowy

- acykliczny graf skierowany zbudowany z:
  - węzłów
  - łuków



Sieć Bayesa

- Składnik jakościowy
- **Składnik ilościowy**
  - **prawdopodobieństwo a priori**
  - **prawdopodobieństwo warunkowe**
  - **prawdopodobieństwo marginalne**
  - **prawdopodobieństwo a posteriori**

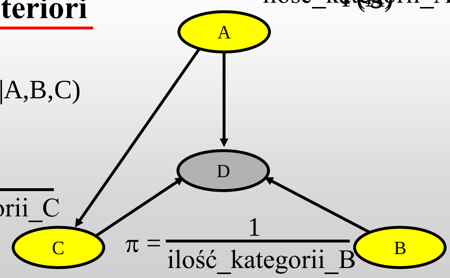
$$P(D|A) = \frac{P(D)*P(A|D)}{P(A)}$$

$$\pi = \frac{1}{ilość\_kategorii\_A}$$

$$P_c = P(A)*P(B)*P(C|A)*P(D|A,B,C)$$

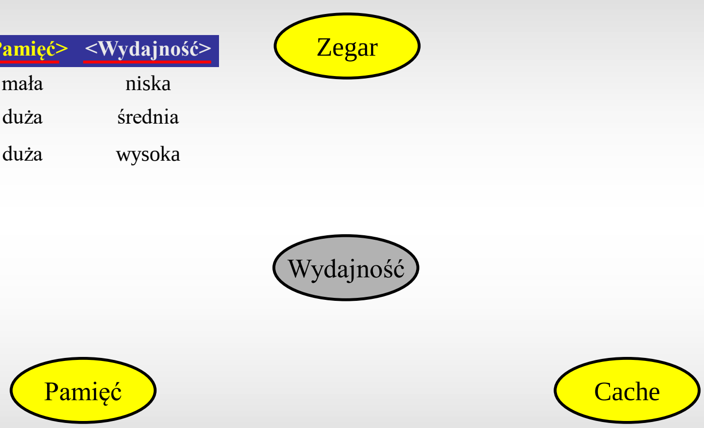
$$\pi = \frac{1}{ilość\_kategorii\_C}$$

$$\pi = \frac{1}{ilość\_kategorii\_B}$$



Przykład

<Zegar>	<Cache>	<Pamięć>	<Wydajność>
wolny	brak	mała	niska
szybki	brak	duża	średnia
szybki	jest	duża	wysoka



Prawdopodobieństwo marginalne

$$\textbf{ML} = \prod_{i=1}^{v}\prod_{j=1}^{q_i}\frac{\Gamma(\alpha_{ij})}{\Gamma(\alpha_{ij}-n_{ij})}\prod_{k=1}^{c_i}\frac{\Gamma(\alpha_{ijk}+n_{ijk})}{\Gamma(\alpha_{ijk})}$$

gdzie:

$i=1,\ldots,v$, gdzie v jest liczbą atrybutów sieci Bayes'a
$j=1,\ldots,q_i$, gdzie q_i jest liczbą możliwych konfiguracji rodziców atrybutu X_{i}
$k=1,\ldots,c_{i,}$ gdzie c_i jest liczbą wartości atrybutu X_i,
n_{ijk} to liczba takich wierszy w bazie, w których rodzice atrybutu X_i przyjmują wartość j a atrybut przyjmuje wartość k,
α_{ijk}, α_{ij} to parametry początkowego rozkładu *Dirichleta*

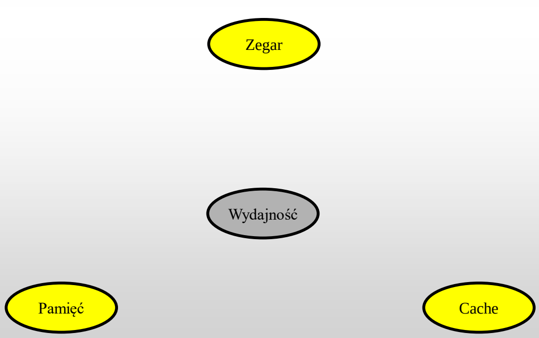
$$\alpha_{ij} = \frac{\alpha}{q_i} \qquad \alpha_{ijk} = \frac{\alpha}{q_i \cdot c_i}$$

Γ - jest funkcją: $\Gamma(n) = (n-1)!$

Prawdopodobieństwo marginalne

$$\textbf{ML} = \prod_{i=1}^{v}\prod_{j=1}^{q_i}\frac{\Gamma(\alpha_{ij})}{\Gamma(\alpha_{ij}-n_{ij})}\prod_{k=1}^{c_i}\frac{\Gamma(\alpha_{ijk}+n_{ijk})}{\Gamma(\alpha_{ijk})}$$

$v = 4$ $\quad i = 1..4$

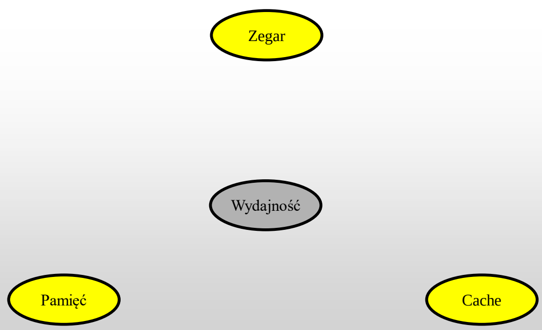


Prawdopodobieństwo marginalne

$$\text{ML} = \prod_{i=1}^{v}\prod_{j=1}^{q_i}\frac{\Gamma(\alpha_{ij})}{\Gamma(\alpha_{ij}-n_{ij})}\prod_{k=1}^{c_i}\frac{\Gamma(\alpha_{ijk}+n_{ijk})}{\Gamma(\alpha_{ijk})}$$

$v = 4$ $i = 1..4$

$q_i = 1$ $j = 1$

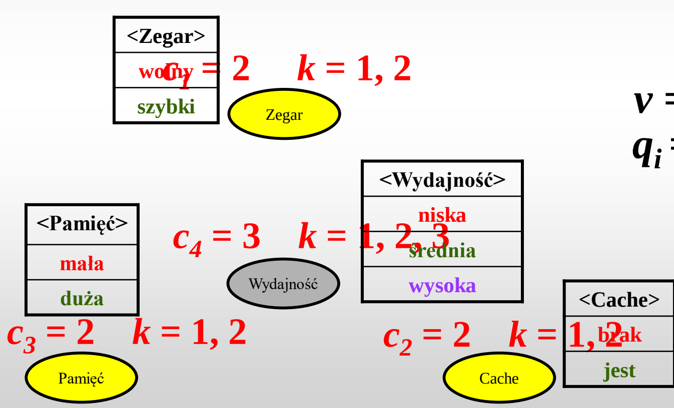


Prawdopodobieństwo marginalne

$$\text{ML} = \prod_{i=1}^{v}\prod_{j=1}^{q_i}\frac{\Gamma(\alpha_{ij})}{\Gamma(\alpha_{ij}-n_{ij})}\prod_{k=1}^{c_i}\frac{\Gamma(\alpha_{ijk}+n_{ijk})}{\Gamma(\alpha_{ijk})}$$

$v = 4$ $i = 1..4$

$q_i = 1$ $j = 1$



Prawdopodobieństwo marginalne

$$\text{ML} = \prod_{i=1}^{v}\prod_{j=1}^{q_i}\frac{\Gamma(\alpha_{ij})}{\Gamma(\alpha_{ij}-n_{ij})}\prod_{k=1}^{c_i}\frac{\Gamma(\alpha_{ijk}+n_{ijk})}{\Gamma(\alpha_{ijk})}$$

<Zegar>
wolny
szybki
szybki

$n_{111} = 1$; $n_{112} = 2$

Zegar

$v = 4$ $i = 1..4$

$q_i = 1$ $j = 1$

<Wydajność>
niska
średnia
wysoka

<Pamięć>
mała
duża
duża

$n_{411} = 1$; $n_{412} = 1$; $n_{413} = 1$

Wydajność

$n_{311} = 1$; $n_{312} = 2$

Pamięć

$n_{211} = 2$; $n_{212} = 1$

Cache

<Cache>
brak
brak
jest

Prawdopodobieństwo marginalne

$$\text{ML} = \prod_{i=1}^{v}\prod_{j=1}^{q_i}\frac{\Gamma(\alpha_{ij})}{\Gamma(\alpha_{ij}-n_{ij})}\prod_{k=1}^{c_i}\frac{\Gamma(\alpha_{ijk}+n_{ijk})}{\Gamma(\alpha_{ijk})}$$

$$\alpha_{ij} = \frac{\alpha}{q_i}$$

$\propto_{11} = 1$

Zegar

$v = 4$ $i = 1..4$

$q_i = 1$ $j = 1$

$\propto = 1$

$\propto_{41} = 1$

Wydajność

$\propto_{31} = 1$

Pamięć

$\propto_{21} = 1$

Cache

Prawdopodobieństwo marginalne

$$\text{ML} = \prod_{i=1}^{v}\prod_{j=1}^{q_i}\frac{\Gamma(\alpha_{ij})}{\Gamma(\alpha_{ij}-n_{ij})}\prod_{k=1}^{c_i}\frac{\Gamma(\alpha_{ijk}+n_{ijk})}{\Gamma(\alpha_{ijk})}$$

$$\alpha_{ijk} = \frac{\alpha}{q_i \cdot c_i}$$

$v = 4$ $i = 1..4$

$q_i = 1$ $j = 1$

$\propto = 1$

$\propto_{111} = ½$; $\propto_{112} = ½$

Zegar $c_1 = 2$

$\propto_{411} = ⅓$; $\propto_{412} = ⅓$; $\propto_{413} = ⅓$

Wydajność $c_4 = 3$

$\propto_{311} = ½$; $\propto_{312} = ½$

Pamięć $c_3 = 2$

$\propto_{211} = ½$; $\propto_{212} = ½$

Cache $c_2 = 2$

Prawdopodobieństwo marginalne

$$\text{ML} = \prod_{i=1}^{v}\prod_{j=1}^{q_i}\frac{\Gamma(\alpha_{ij})}{\Gamma(\alpha_{ij}-n_{ij})}\prod_{k=1}^{c_i}\frac{\Gamma(\alpha_{ijk}+n_{ijk})}{\Gamma(\alpha_{ijk})}$$

$$\Downarrow$$

$$\ln(\text{ML}) = \sum_{i=1}^{v}\left\{\sum_{j=1}^{q_i}\left\{\ln\frac{\Gamma(\alpha_{ij})}{\Gamma(\alpha_{ij}-n_{ij})}+\sum_{k=1}^{c_i}\left\{\ln\frac{\Gamma(\alpha_{ijk}+n_{ijk})}{\Gamma(\alpha_{ijk})}\right\}\right\}\right\}$$

Prawdopodobieństwo marginalne

$$\ln(\mathrm{ML}) = \sum_{i=1}^{4}\left\{\sum_{j=1}\left\{\ln\frac{\Gamma(\alpha_{ij})}{\Gamma(\alpha_{ij}-n_{ij})}+\sum_{k=1}^{c_i}\left\{\ln\frac{\Gamma(\alpha_{ijk}+n_{ijk})}{\Gamma(\alpha_{ijk})}\right\}\right\}\right\}$$

$\Downarrow$

Zegar

$$\{[\ln\Gamma(\alpha_{11})-\ln\Gamma(\alpha_{11}-n_{11})]+[\ln\Gamma(\alpha_{111}-n_{111})-\ln\Gamma(\alpha_{111})]+[\ln\Gamma(\alpha_{112}-n_{112})-\ln\Gamma(\alpha_{112})]\}$$

Cache

$+$

$$\{[\ln\Gamma(\alpha_{21})-\ln\Gamma(\alpha_{21}-n_{21})]+[\ln\Gamma(\alpha_{211}-n_{211})-\ln\Gamma(\alpha_{211})]+[\ln\Gamma(\alpha_{212}-n_{212})-\ln\Gamma(\alpha_{212})]\}$$

Pamięć

$+$

$$\{[\ln\Gamma(\alpha_{31})-\ln\Gamma(\alpha_{31}-n_{31})]+[\ln\Gamma(\alpha_{311}-n_{311})-\ln\Gamma(\alpha_{311})]+[\ln\Gamma(\alpha_{312}-n_{312})-\ln\Gamma(\alpha_{312})]\}$$

Wydajność

$+$

$$+\{[\ln\Gamma(\alpha_{41})-\ln\Gamma(\alpha_{41}-n_{41})]+[\ln\Gamma(\alpha_{411}-n_{411})-\ln\Gamma(\alpha_{411})]+[\ln\Gamma(\alpha_{412}-n_{412})-\ln\Gamma(\alpha_{412})]$$
$$+[\ln\Gamma(\alpha_{413}-n_{413})-\ln\Gamma(\alpha_{413})]\}$$

Prawdopodobieństwo marginalne

Zegar

ln(ML) = -13,40

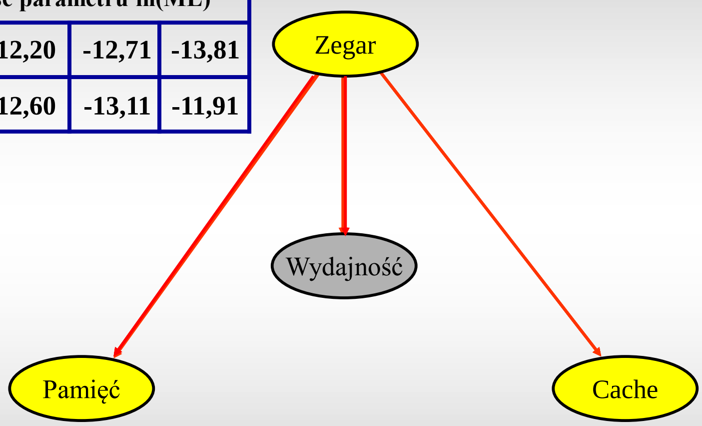
Wydajność

Pamięć

Cache

Algorytm K2 – KROK I

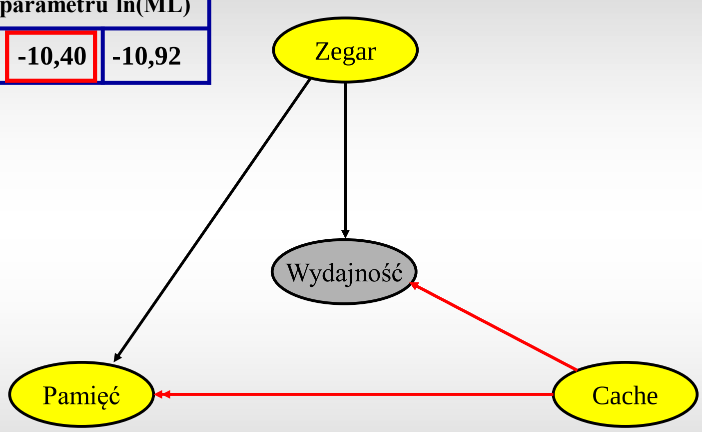
Wartość parametru ln(ML)			
-13,40	-12,20	-12,71	-13,81
-11,50	-12,60	-13,11	-11,91



ln(ML) = -11,50

Algorytm K2 – KROK II

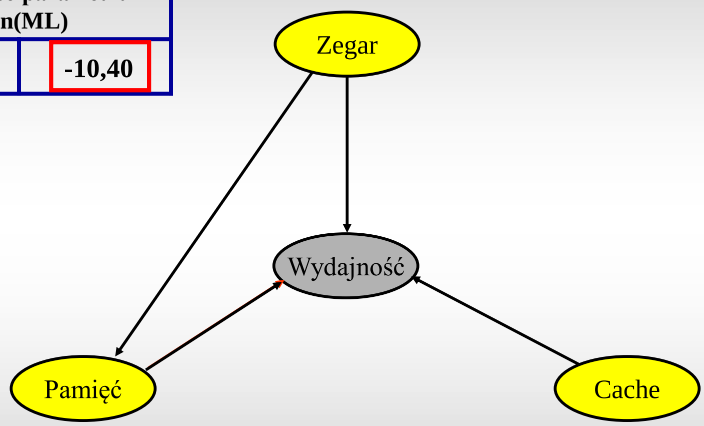
Wartość parametru ln(ML)		
-12,01	-10,40	-10,92



ln(ML) = -10,40

Algorytm K2 – KROK III

Wartość parametru ln(ML)	
-10,40	-10,40



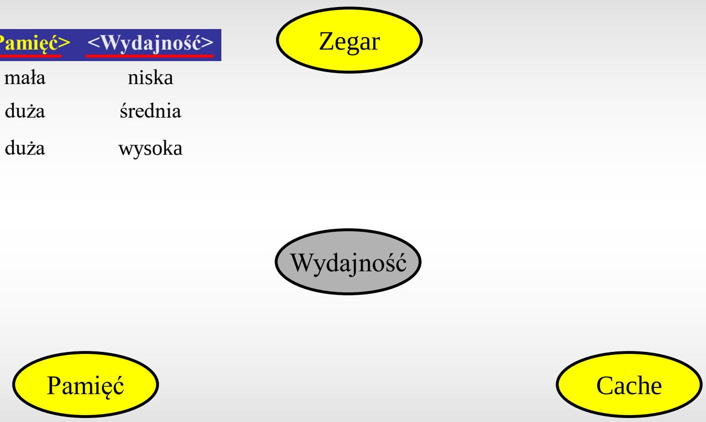
ln(ML) = -10,40

Rozkład prawdopodobieństwa

- Rozkład prawdopodobieństwa początkowego (a priori π)
- Rozkład prawdopodobieństwa dla sieci bez połączeń
- Rozkład prawdopodobieństwa dla węzłów posiadających „rodziców”

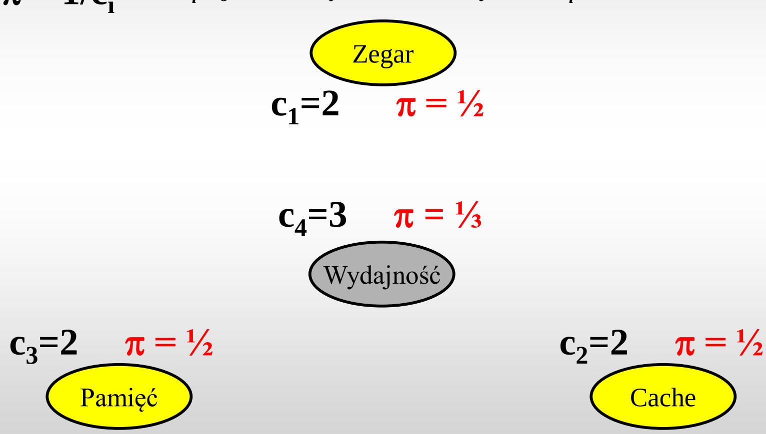
Przykład

<Zegar>	<Cache>	<Pamięć>	<Wydajność>
wolny	brak	mała	niska
szybki	brak	duża	średnia
szybki	jest	duża	wysoka



Rozkład prawdopodobieństwa

$\pi = 1/c_i$ **c_i jest liczbą wartości atrybutu X_i**



Rozkład prawdopodobieństwa

$P(X_i) = \frac{\pi * \alpha + x_i}{\alpha + n}$; gdzie x_i ilość wystąpień określonej wartości atrybutu

Zegar

$$P(X_1) = \begin{cases} \frac{\pi * \alpha + x_{i=szybki}}{\alpha + n} = \frac{\frac{1}{2}*1+2}{1+3} = 0{,}625 \\ \frac{\pi * \alpha + x_{i=wolny}}{\alpha + n} = \frac{\frac{1}{2}*1+1}{1+3} = 0{,}375 \end{cases}$$

$$P(X_4) = \begin{cases} \frac{\pi * \alpha + x_{i=niska}}{\alpha + n} = \frac{\frac{1}{3}*1+1}{1+3} = 0{,}333 \\ \frac{\pi * \alpha + x_{i=średnia}}{\alpha + n} = \frac{\frac{1}{3}*1+1}{1+3} = 0{,}333 \\ \frac{\pi * \alpha + x_{i=wysoka}}{\alpha + n} = \frac{\frac{1}{3}*1+1}{1+3} = 0{,}333 \end{cases}$$

$$P(X_3) = \begin{cases} \frac{\pi * \alpha + x_{i=mała}}{\alpha + n} = \frac{\frac{1}{2}*1+2}{1+3} = 0{,}625 \\ \frac{\pi * \alpha + x_{i=duża}}{\alpha + n} = \frac{\frac{1}{2}*1+1}{1+3} = 0{,}375 \end{cases}$$

$$P(X_2) = \begin{cases} \frac{\pi * \alpha + x_{i=brak}}{\alpha + n} = \frac{\frac{1}{2}*1+2}{1+3} = 0{,}625 \\ \frac{\pi * \alpha + x_{i=jest}}{\alpha + n} = \frac{\frac{1}{2}*1+1}{1+3} = 0{,}375 \end{cases}$$

Wydajność

Pamięć

Cache

Rozkład prawdopodobieństwa

$P(X_i) = \frac{\pi * \alpha + x_i}{\alpha + n}$

Zegar

Zegar	
szybki	**wolny**
0,625	**0,375**

Wydajność		
mała	**średnia**	**duża**
0,333	**0,333**	**0,333**

Cache	
brak	**jest**
0,625	**0,375**

Wydajność

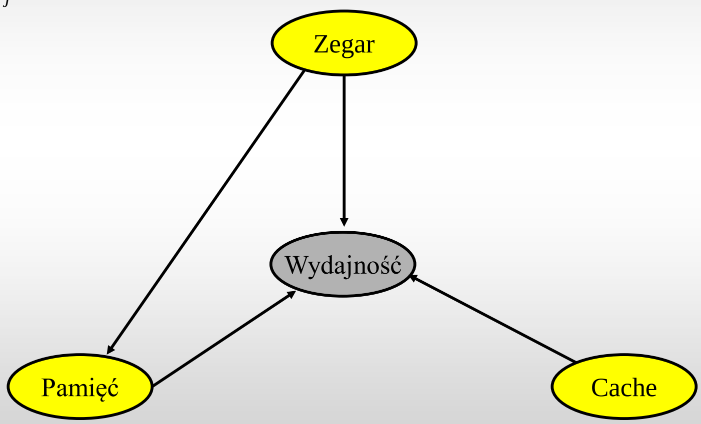
Pamięć

Cache

Pamięć	
mała	**duża**
0,625	**0,375**

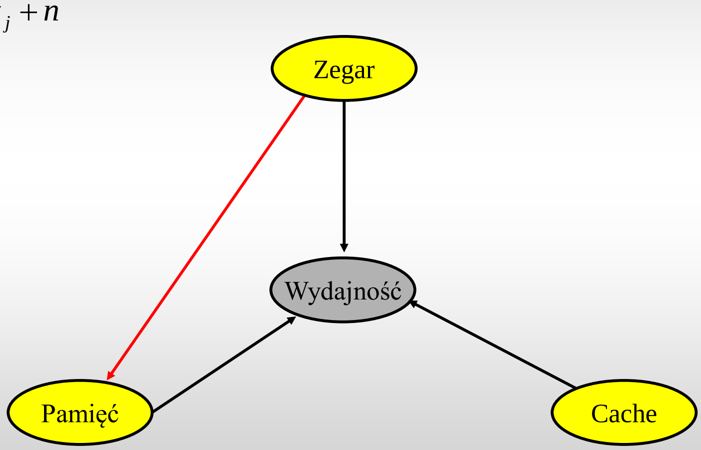
Rozkład prawdopodobieństwa

$$P(X_i) = \frac{\pi * \alpha_j + n_j}{\alpha_j + n} \quad ;$$



Rozkład prawdopodobieństwa

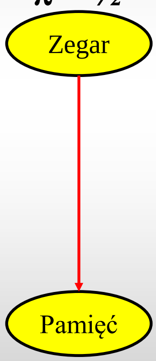
$$P(X_i) = \frac{\pi * \alpha_j + n_j}{\alpha_j + n}$$



Rozkład prawdopodobieństwa

$$P(X_i) = \frac{\pi * \alpha_j + n_j}{\alpha_j + n}$$

$\pi = ½$



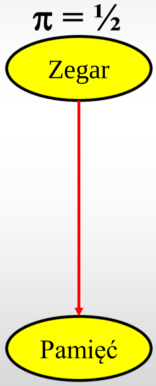
$\pi = ½$

<Zegar>	<Cache>	<Pamięć>	<Wydajność>
wolny	brak	mała	niska
szybki	brak	duża	średnia
szybki	jest	duża	wysoka

Rozkład prawdopodobieństwa

$$P(X_i) = \frac{\pi * \alpha_j + n_j}{\alpha_j + n}$$

$\pi = ½$



$\pi = ½$

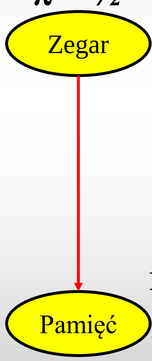
<Zegar>	<Cache>	<Pamięć>	<Wydajność>
wolny	brak	mała	niska
szybki	brak	duża	średnia
szybki	jest	duża	wysoka

Rozkład prawdopodobieństwa

$P(X_i) = \dfrac{\pi * \alpha_j + n_j}{\alpha_j + n}$; gdzie n_j to liczba wszystkich przypadków takich, że dana kombinacji atrybutów występuje w bazie oraz że wartość jest taka jak oczekujemy

<Zegar>	<Cache>	<Pamięć>	<Wydajność>
wolny	brak	mała	niska
szybki	brak	duża	średnia
szybki	jest	duża	wysoka

$\pi = ½$



$\pi = ½$

P(Z=wolny|P=mała) = (0,5 * 0,5 + 1)/(0,5 + 1) = 0,833

P(Z=wolny|P=duża) = (0,5 * 0,5 + 0)/(0,5 + 1) = 0,167

P(Z=szybki|P=mała) = (0,5 * 0,5 + 0)/(0,5 + 2) = 0,1

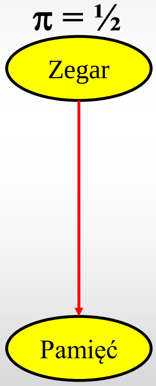
P(Z=szybki|P=duża) = (0,5 * 0,5 + 2)/(0,5 + 2) = 0,9

Rozkład prawdopodobieństwa

$P(X_i) = \dfrac{\pi * \alpha_j + n_j}{\alpha_j + n}$

<Zegar>	<Cache>	<Pamięć>	<Wydajność>
wolny	brak	mała	niska
szybki	brak	duża	średnia
szybki	jest	duża	wysoka

$\pi = ½$

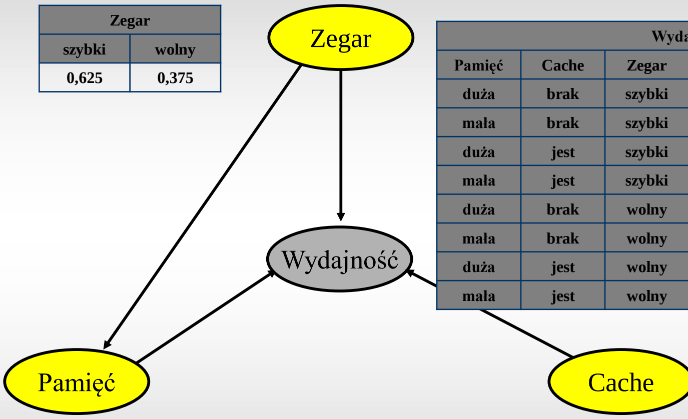


$\pi = ½$

	Pamięć	
Zegar	mała	duża
wolny	0,833	0,167
szybki	0,9	0,1

Rozkład prawdopodobieństwa

Zegar	
szybki	wolny
0,625	0,375

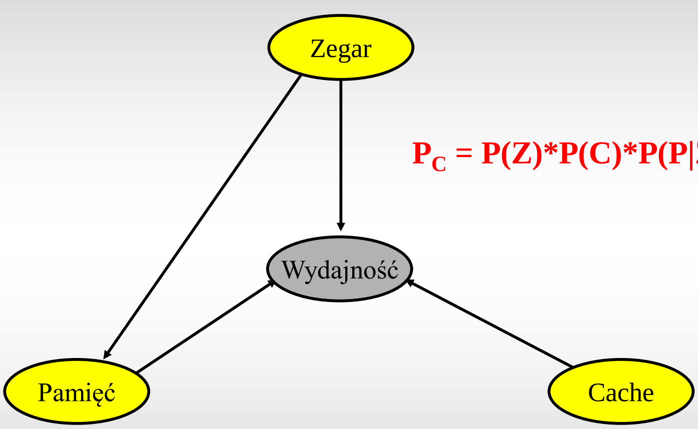


Wydajność					
Pamięć	Cache	Zegar	niska	wysoka	średnia
duża	brak	szybki	0,037	0,037	0,926
mała	brak	szybki	0,333	0,333	0,333
duża	jest	szybki	0,037	0,926	0,037
mała	jest	szybki	0,333	0,333	0,333
duża	brak	wolny	0,333	0,333	0,333
mała	brak	wolny	0,926	0,037	0,037
duża	jest	wolny	0,333	0,333	0,333
mała	jest	wolny	0,333	0,333	0,333

Pamięć		
Zegar	mała	duża
wolny	0,833	0,167
szybki	0,9	0,1

Cache	
brak	jest
0,625	0,375

Rozkład marginalny



P_C = P(Z)*P(C)*P(P|Z)*P(W|Z,C,P)

$$P(W=niska) = \sum_{z,c,p} P(Z)*P(C)*P(P|Z)*P(W|Z,C,P)$$

Rozkład marginalny

$$P(W{=}niska) = \sum_{z,c,p} P(Z)*P(C)*P(P|Z)*P(W|Z,C,P) =$$

$$= \sum_{z,p} P(Z)*P(P|Z)* \sum_{c} P(C)*P(W|Z,C,P) =$$

$$= \sum_{z,p} P(Z)*P(P|Z)*[P(C{=}brak)*P(W|Z,C{=}brak,P)+$$

$$+P(C{=}jest)*P(W|Z,C{=}jest,P)] =$$

$$= \sum_{z} P(Z)*\sum_{p} P(P|Z)*[P(C{=}brak)*P(W|Z,C{=}brak,P)+$$

$$+P(C{=}jest)*P(W|Z,C{=}jest,P)] =$$

Rozkład marginalny

$$= \sum_{z} P(Z)*\{P(P{=}duża|Z)*[P(C{=}brak)*P(W|Z, C{=}brak,P{=}duża)+$$

$$+P(C{=}jest)*P(W|Z, C{=}jest,P{=}duża)]+$$

$$P(P{=}mała|Z)*[P(C{=}brak)*P(W|Z, C{=}brak,P{=}mała)+$$

$$+P(C{=}jest)*P(W|Z, C{=}jest,P{=}mała)]\}=$$

Rozkład marginalny

= P(Z=szybki)*({P(P=duża|Z=szybki)*

Pamięć		
Zegar	mała	duża
wolny	0,833	0,167
szybki	0,9	0,1

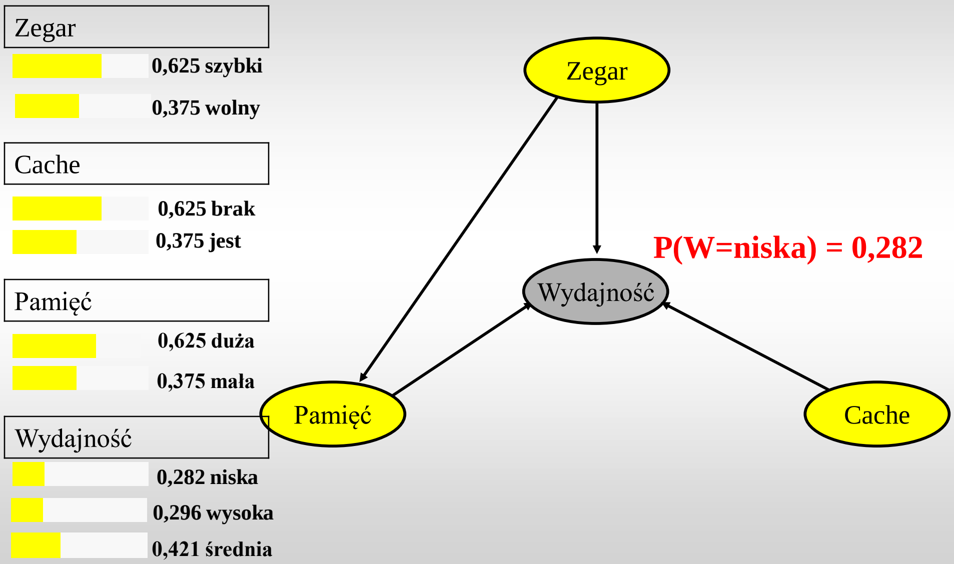
Rozkład marginalny

= P(Z=szybki)*({P(P=duża|Z=szybki)*

*[P(C=brak)*P(W|Z=szybki,C=brak,P=duża)+

			Wydajność		
Pamięć	Cache	Zegar	niska	wysoka	średnia
duża	brak	szybki	0,037	0,037	0,926
mała	brak	szybki	0,333	0,333	0,333
duża	jest	szybki	0,037	0,926	0,037
mała	jest	szybki	0,333	0,333	0,333
duża	brak	wolny	0,333	0,333	0,333
mała	brak	wolny	0,926	0,037	0,037
duża	jest	wolny	0,333	0,333	0,333
mała	jest	wolny	0,333	0,333	0,333

Rozkład marginalny



Uczenie sieci polega na:

obliczeniu rozkładu prawdopodobieństwa
oraz
obliczeniu rozkładu marginalnego

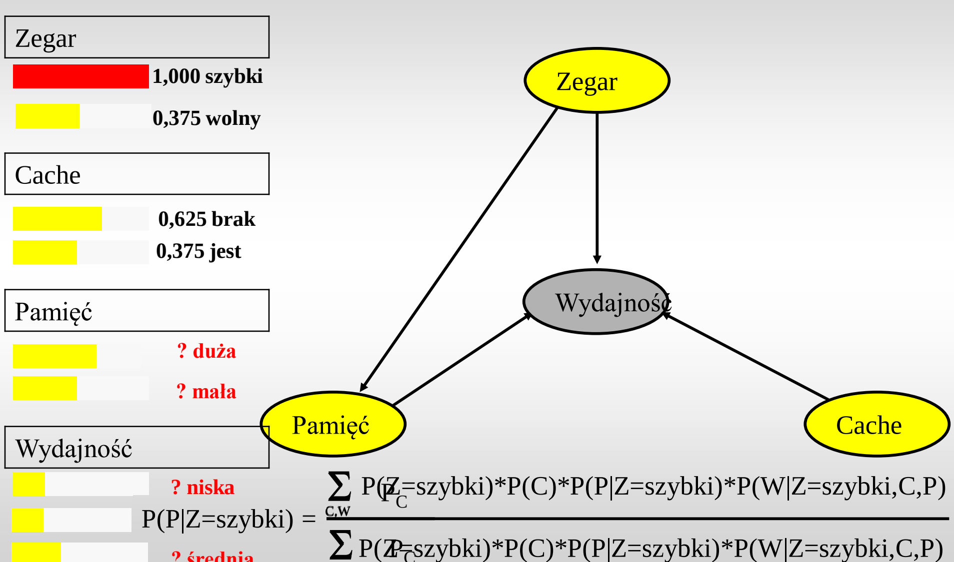
dla badanego zbioru danych,
zaś kolejnym etapem badań jest

wnioskowanie

Ze względu na to, że BN określa wspólny rozkład prawdopodobieństwa, można go wykorzystać do wnioskowania. Najpopularniejsze metody wnioskowania to:

- **wnioskowanie predykcyjne (od przyczyn do skutków) - od nowych informacji o przyczynach do nowych przekonań o skutkach, zgodnie z kierunkami łuków sieci**
- **wnioskowanie diagnostycznego - (od skutków do przyczyn), od nowych skutków do nowych przekonaniach o przyczynach, w przeciwnym kierunku niż łuki sieci (Hagmayer i in., 2007; Pearl, 2009; Korb i in., 2014).**

Wnioskowanie predykcyjne



$$P(P|Z=szybki) = \frac{\sum_{C,W} P(Z=szybki)*P(C)*P(P|Z=szybki)*P(W|Z=szybki,C,P)}{\sum_{C,P,W} P(Z=szybki)*P(C)*P(P|Z=szybki)*P(W|Z=szybki,C,P)}$$