

Bayesovské sítě – rozšíření a zajímavosti

Jiří Kléma

Katedra kybernetiky,
FEL, ČVUT v Praze



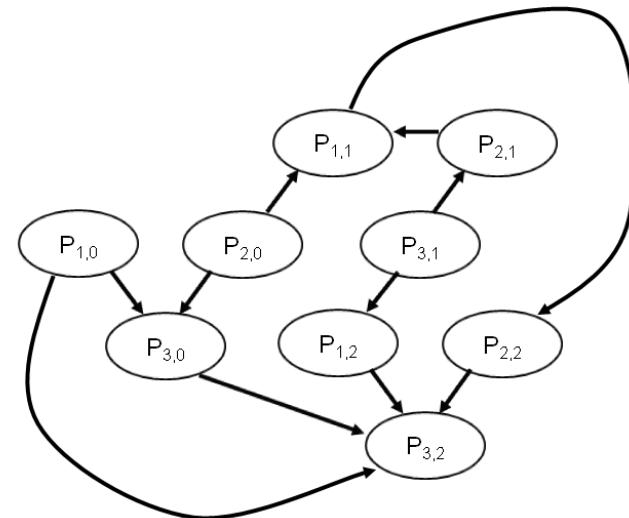
<http://ida.felk.cvut.cz>

Struktura přednášky

- Dosud výhradně “klasická” varianta BN
 - statický model,
 - binární proměnné.
- Nyní naznačíme
 - problematiku dynamických bayesovských sítí
 - * modelují vývoj systému v čase,
 - modelování dalších typů proměnných
 - * multinomiální diskrétní, spojité a jejich kombinace.
- Jiné typy grafů
 - orientované → neorientované grafy,
 - markovská náhodná pole (Markov Random Fields)
 - * jiné vlastnosti, jiné použití.

Dynamické bayesovské sítě – Dynamic Bayesian Networks (DBNs)

- uzly = proměnné odpovídající stavům reálných veličin v konkrétním čase,
- reálné veličiny se stále opakují (viz $P_{1,t}$, $P_{2,t}$, $P_{3,t}$ v obrázku),
- hrany opět vyjadřují příčinnou souvislost,
- bez dalších omezení příliš možných vazeb,

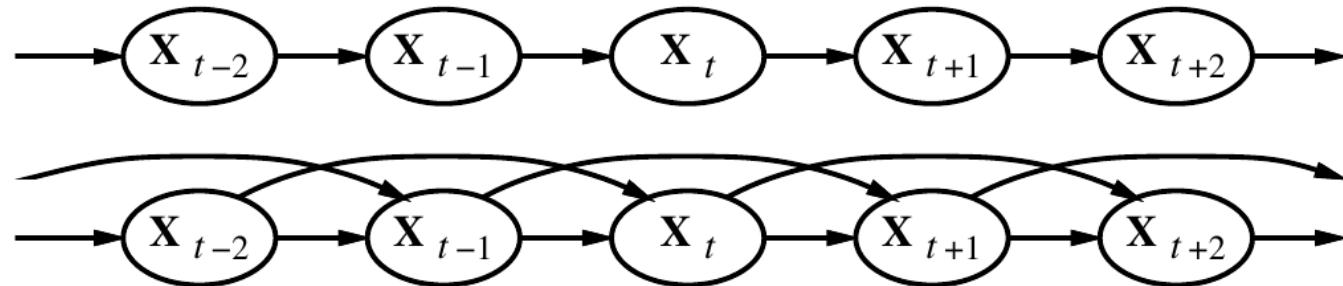


- veličiny a notace (a **implicitní omezení**)
 - t – **diskrétní čas** (volba velikosti kroku závisí na problému),
 - X_t – množina nepozorovatelných **stavových** veličin v čase t ,
* $KrevniCukr_t$, $ObsahZaludku_t$, $DusevniPohoda_t$, ...
 - E_t – množina pozorovatelných **senzorických** veličin v čase t ,
* $MereniKrevnihoCukru_t$, $PrijemTekutin_t$, $ReakceNaDotazy_t$, ...
 - $X_{a:b}$ – $X_a, X_{a+1}, \dots, X_{b-1}, X_b$,

Dynamické bayesovské sítě – Dynamic Bayesian Networks (DBNs)

- Omezení č. 1: **markovský předpoklad** řádu n

- přechodový model: $Pr(X_t|X_{0:t-1}) = Pr(X_t|X_{t-n:t-1})$



- senzorický model: $Pr(E_t|X_{0:t}, E_{0:t-1}) = Pr(E_t|X_t)$

- Omezení č. 2: **stacionarita**

- přechodový model ani senzorický model se v čase nemění

$$\forall t : Pr(X_t|X_{t-n:t-1}) = Pr(X_{t+1}|X_{t-n+1:t})$$

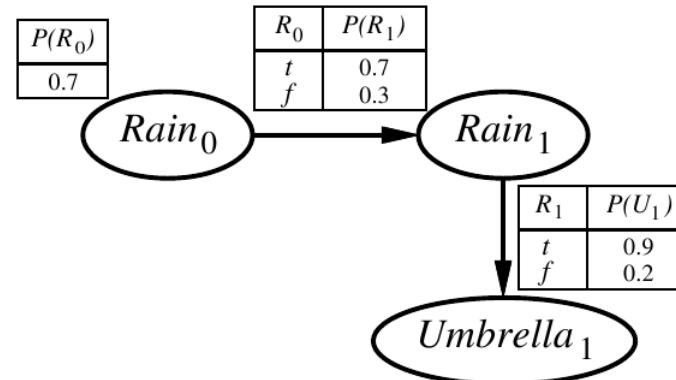
$$\forall t : Pr(E_t|X_t) = Pr(E_{t+1}|X_{t+1})$$

- K zadání DBN s markovským předpokladem řádu 1 stačí zadat

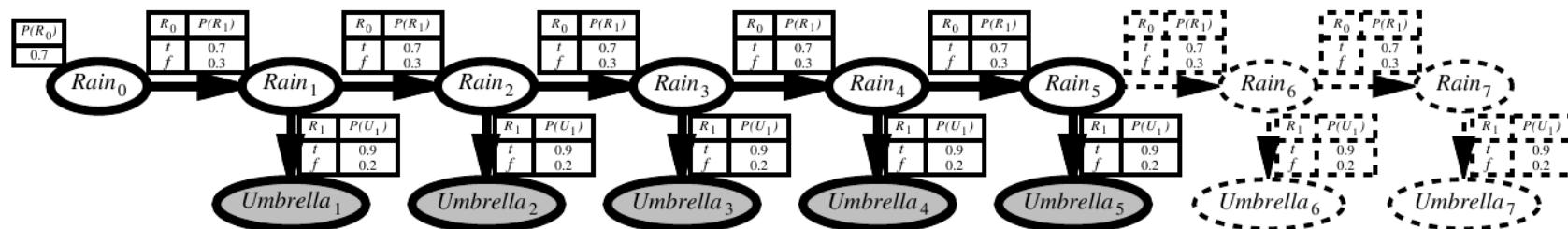
- počáteční podmínky: $Pr(X_0)$,
 - přechodový a senzorický model (pro jediné t).

Příklad DBN (Russell, Norvig: AIMA)

- k zadání postačí počáteční podmínky a jeden řez (přechodový a senzorický model)

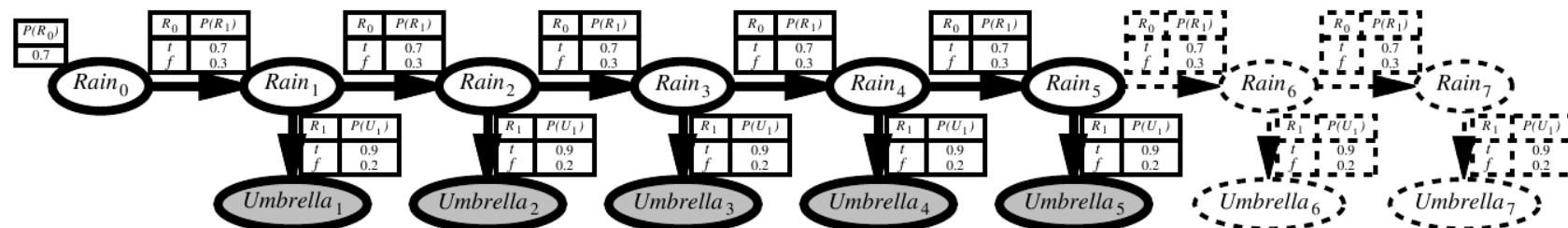


- síť pak lze rozbalit (rozvinout) až po libovolný časový krok



Příklad DBN (Russell, Norvig: AIMA) – pokračování

- pomocí klasické BN inference lze vyčíslit pravděpodobnost libovolného jevu za libovolných pozorování,
- mohu ignorovat dělení na stavové a senzorické veličiny, či dotazovat se na minulé stavy,
- jak se změní pravděpodobnost, že ve dni 1 pršelo vím-li, že ve dni 4 nepršelo a ve dni 2 se nosily deštníky?
 - $Pr(r_1) = Pr(r_0, r_1) + Pr(\neg r_0, r_1) = Pr(r_0)Pr(r_1|r_0) + Pr(\neg r_0)Pr(r_1|\neg r_0) = 0.58$,
 - Matlab BNT, jtreet_inf_engine: $Pr(r_1|\neg r_4, u_2) = 0.68$,
 - pravděpodobnost deště v den 1 pozorováním vzrostla,
- počet časových kroků nemusí být časově omezený
 - pro výpočetní zvládnutelnost omezení pouze na specializované úlohy.



DBN inferenční úlohy

- **filtrování** – $Pr(X_t|e_{1:t})$
 - zjištění pstrního rozložení aktuálního stavu za znalosti (všech) dřívějších senzorických veličin,
 - základní úloha při rozhodování racionálního agenta,
- **predikce** – $Pr(X_{t+k}|e_{1:t}), k > 0$
 - na rozdíl od filtrování nás zajímá budoucí stav,
- **vyhlazování** – $Pr(X_k|e_{1:t}), 0 \leq k < t$
 - lepší znalost o minulých stavech lze využít mj. při učení,
- **nejpravděpodobnější průchod** – $\arg \max_{x_{1:t}} Pr(x_{1:t}|e_{1:t})$
 - pozor, nejde o opakovanou aplikaci vyhlazování – stavy se navzájem ovlivňují,
 - využití například u rozpoznávání řeči (řečový signál → věta).

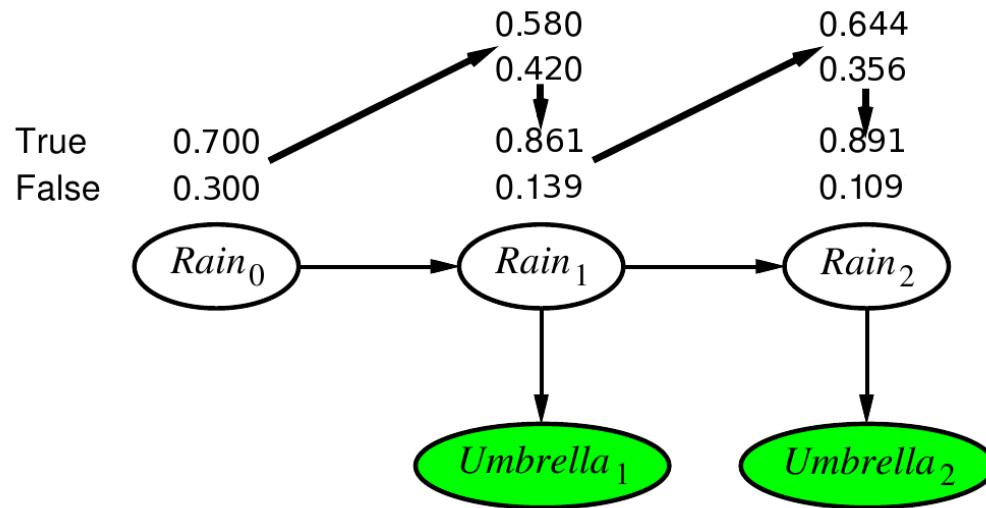
Filtrování

- cílem je zavést a implementovat rekurzivní definici
 - $Pr(X_{t+1}|e_{1:t+1}) = f(e_{t+1}, Pr(X_t|e_{1:t}))$,
 - filtrování pro čas t a pozorování v čase $t+1 \rightarrow$ filtrování pro čas $t+1$,

$$\begin{aligned} Pr(X_{t+1}|e_{1:t+1}) &= Pr(X_{t+1}|e_{1:t}, e_{t+1}) = \dots \text{ pouze rozděl pozorování} \\ &= \alpha Pr(e_{t+1}|X_{t+1}, e_{1:t}) Pr(X_{t+1}|e_{1:t}) = \dots \text{ použij Bayesův vzorec} \\ &= \alpha Pr(e_{t+1}|X_{t+1}) Pr(X_{t+1}|e_{1:t}) \dots \text{ plyne z markovské vlastnosti pozorování} \end{aligned}$$

- filtrování zatím rozloženo na predikci ($Pr(X_{t+1}|e_{1:t})$) a zpracování pozorování ($Pr(e_{t+1}|X_{t+1})$)
- predikci vyřešíme vysčítáním přes rozdělení X_t
 - $Pr(X_{t+1}|e_{1:t}) = \sum_{x_t} Pr(X_{t+1}|x_t, e_{1:t}) Pr(x_t|e_{1:t}) = \sum_{x_t} Pr(X_{t+1}|x_t) Pr(x_t|e_{1:t})$,
- a dostáváme tak časově konstantní a rekurzivní funkci f
 - $f_{1,t+1} = \text{FORWARD}(f_{1:t}, e_{t+1})$, kde $f_{1:t} = Pr(X_t|e_{1:t})$

Filtrování (příklad)



- pozorujeme u_1 a u_2 – jaká je prst deště v čase 1 (nebereme-li v úvahu u_2) a v čase 2?

$$- Pr(R_1|u_1) = \alpha_1 Pr(u_1|R_1) \sum_{R_0 \in \{r_0, \neg r_0\}} Pr(R_1|R_0)Pr(R_0),$$

$$- Pr(r_1|u_1) = \alpha_1 \times 0.9 \times (0.7 \times 0.7 + 0.3 \times 0.3) = \alpha_1 \times 0.9 \times 0.58,$$

$$- Pr(\neg r_1 | u_1) = \alpha_1 \times 0.2 \times (0.3 \times 0.7 + 0.7 \times 0.3) = \alpha_1 \times 0.2 \times 0.42,$$

$$-\alpha_1 = 1.65 \rightarrow Pr(r_1|u_1) = 0.861, Pr(\neg r_1|u_1) = 0.139$$

$$- Pr(R_2|u_1, u_2) = \alpha_2 Pr(u_2|R_2) \sum_{R_1 \in \{r_1, \neg r_1\}} Pr(R_2|R_1) Pr(R_1|u_1),$$

$$- Pr(r_2|u_1, u_2) = \alpha_2 \times 0.9 \times (0.7 \times 0.861 + 0.3 \times 0.139) = \alpha_2 \times 0.9 \times 0.644,$$

$$- Pr(\neg r_2 | u_1, u_2) = \alpha_2 \times 0.2 \times (0.3 \times 0.861 + 0.7 \times 0.139) = \alpha_2 \times 0.2 \times 0.356,$$

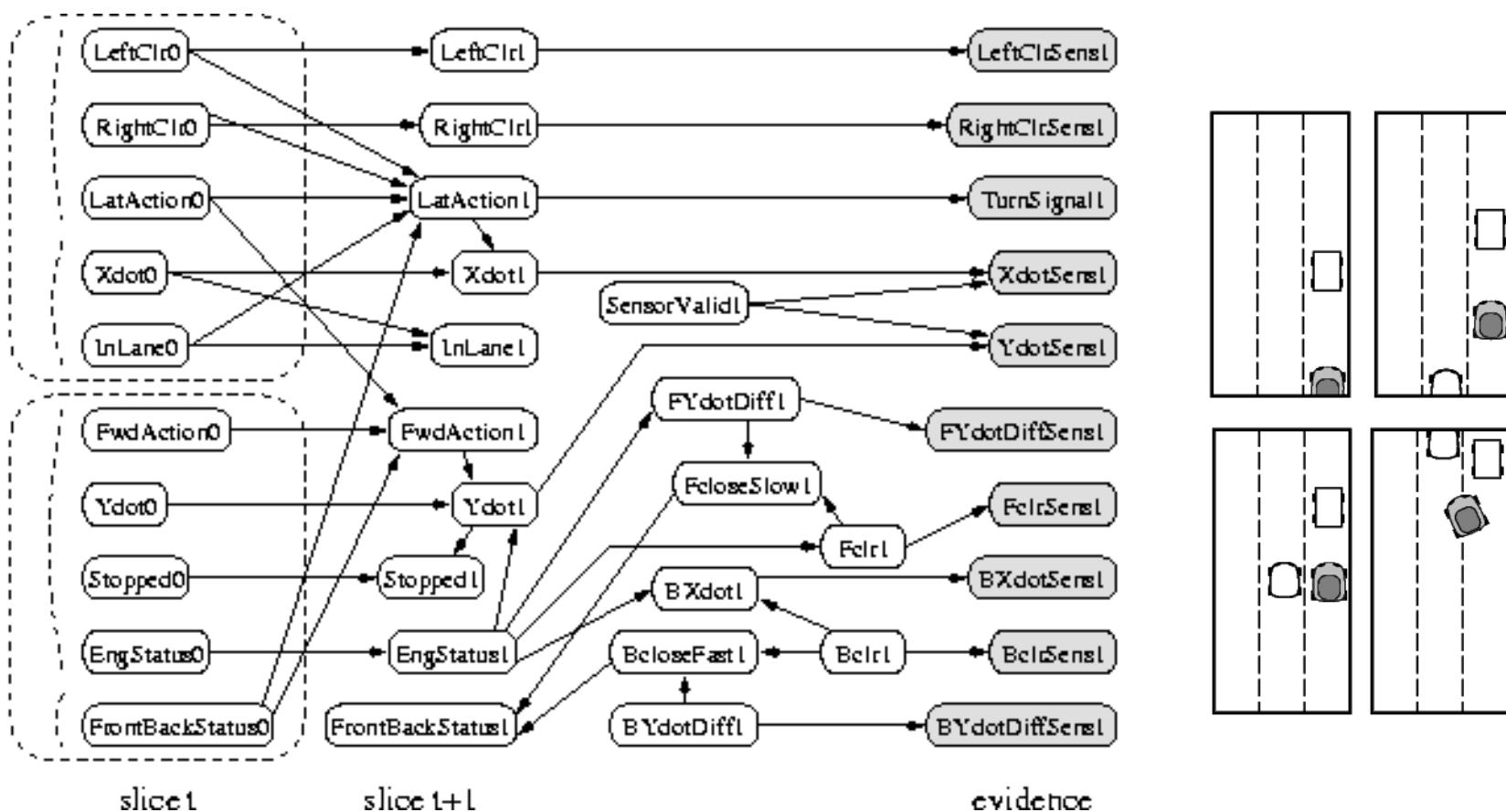
$$-\alpha_2 = 1.54 \rightarrow Pr(r_2|u_1, u_2) = 0.891, Pr(\neg r_2|u_1, u_2) = 0.109$$

DBN vs HMM

- Hidden Markov Model (HMM) – skrytý markovův model
 - časté použití v rozpoznávání řeči nebo biomedicíně,
- HMM je speciálním (a triviálním) případem dynamické bayesovské sítě
 - stav je popsán jedinou veličinou (senzorická veličina obvykle také jediná),
 - tj. příklad Rain-Umbrella je současně DBN a HMM,
 - mnohem častější použití HMMs než obecnějších DBNs
 - * HMM je statistický model zvládnutelný algoritmy s polynomiální složitostí.

Bayesian Automated Taxi (projekt, článek IJCAI 1995)

- projekt autonomního vozidla v běžném dálničním provozu
 - předjetí pomalejšího a objetí stojícího vozidla,
 - reakce na manévry jiných řidičů, změna počtu pruhů.



DBNs – přehled

- dosud jsme modelovali v BNs stálý vztah mezi různými veličinami
 - čas nehrál roli,
 - nebo model postihoval jeden časový okamžik.
- často je třeba sledovat proces změn v čase
 - popíšeme stav světa a modelujeme jeho dynamiku,
- kvůli zvládnutelnosti zavedeme omezení
 - markovský předpoklad: aktuální stav závisí pouze na časově omezené historii stavů,
 - stacionarita: pravděpodobnostní vztahy jsou v čase fixní, řezy se nemění,
 - diskrétní čas: lze oddělit řezy ($\text{řez} = \text{stav světa v konkrétním čase}$),
 - explicitní dělení veličin: skryté/stavové a pozorovatelné,
- pro každý čas znova kopírujeme celý řez = stavové a pozorovatelné veličiny
 - rozbalení sítě (unroll),
 - po rozbalení je DBN ekvivalentem běžné BN,
 - vzhledem k potenciálně nekonečnému rozvoji se tato analogie prakticky nevyužívá,
 - nové standardní úlohy: filtrování, predikce, vyhlazování, nejpravděpodobnější průchod.

Jiné typy proměnných

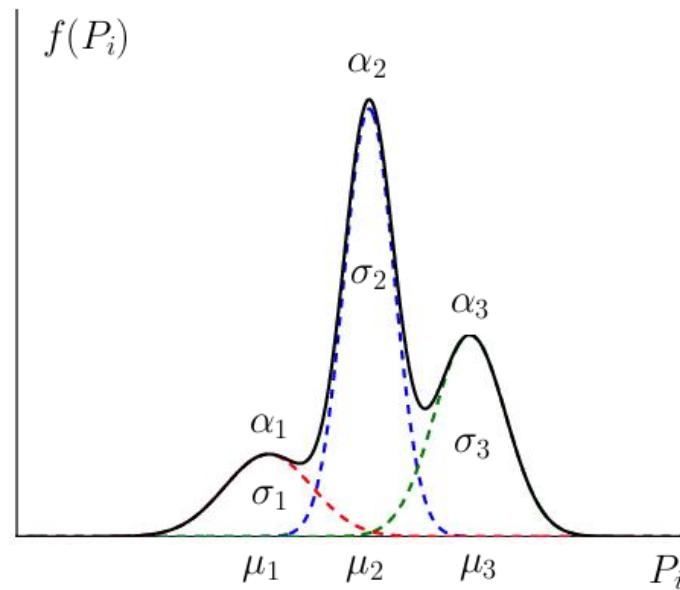
- BN ze své povahy musí určovat $Pr(P_i|rodice(P_i))$
 - CPT je zvláštní variantou obecnějšího **podmíněného pstního rozdělení** (conditional probability distribution – CPD),
- multinomiální proměnné
 - (diskrétní proměnné s více než dvěma možnými hodnotami),
 - reprezentace nijak odlišná od binárních veličin – CPTs,
 - CPTs pouze rychleji rostou s počtem rodičů
 - * $q_i \dots$ počet unikátních instanciací rodičů proměnné P_i ,
 - * $r_i \dots$ je počet hodnot P_i ,
 - * $r \dots$ max počet hodnot proměnné,
 - * celkový počet nezávislých parametrů BN

$$K = \sum_{i=1}^n q_i(r_i - 1),$$

- * pro binární proměnné $r_i = 2$ a $q_i = 2^{|rodice(P_i)|}$,
- * obecně $2 \leq r_i \leq r$ a $q_i \leq r^{|rodice(P_i)|}$.

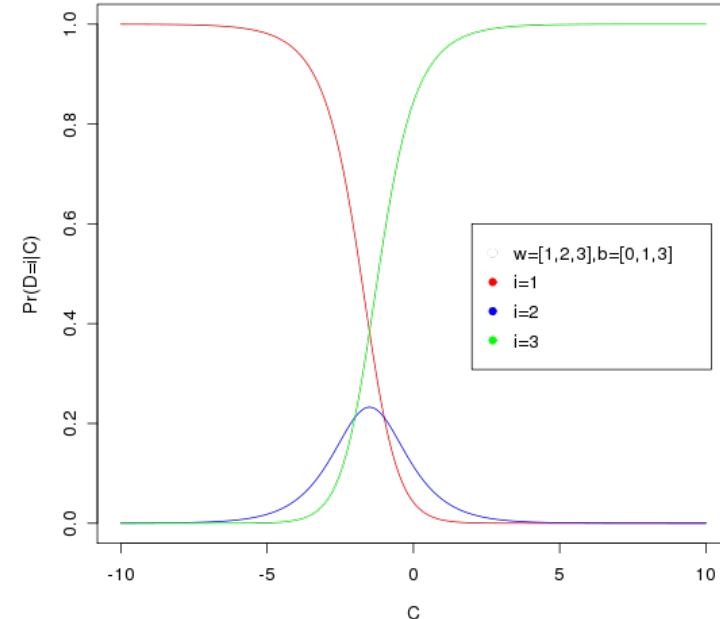
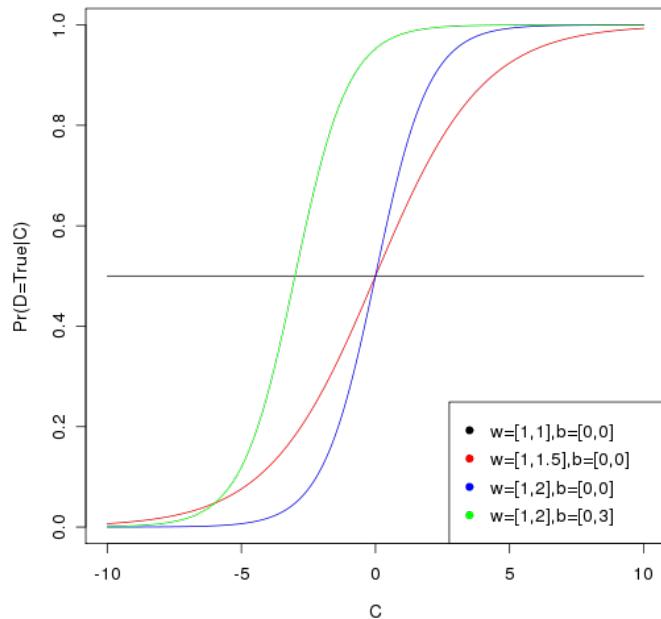
Jiné typy proměnných

- spojité proměnné
 - diskretizace → multinomiální diskrétní proměnné,
 - * nevýhodné dilema: rozsáhlé CPTs vs ztráta informace,
 - reprezentace pomocí kanonických pstních rozdělení
 - * konečná, obvykle jednoduchá parametrizace,
 - * častý příklad: gaussovo rozdělení, případně jeho směsi,
 - * obrázek: definice spojitého uzlu bez rodičů 9 parametry,



Spojity uzel jako rodič

- CPD pro spojitého rodiče C a multinomiálního dítěte D
 - měkké prahování pomocí softmax funkce (multinomiální logistická fce)
$$Pr(D = d_i | C = c) = \frac{e^{w_i \times c + b_i}}{\sum_{j=1}^{|D|} e^{w_j \times c + b_j}}$$
 - \vec{w} – určuje měkkost (sklon) prahu, \vec{b} – určuje polohu prahu,
- lze i pro více spojitých rodičů
 - \mathbf{W} i \mathbf{B} ve formě matic, příspěvek rodičů se lineárně kombinuje,
- binární $D \rightarrow$ softmax odpovídá sigmoidě, ■ ternární D – psti po třídách,

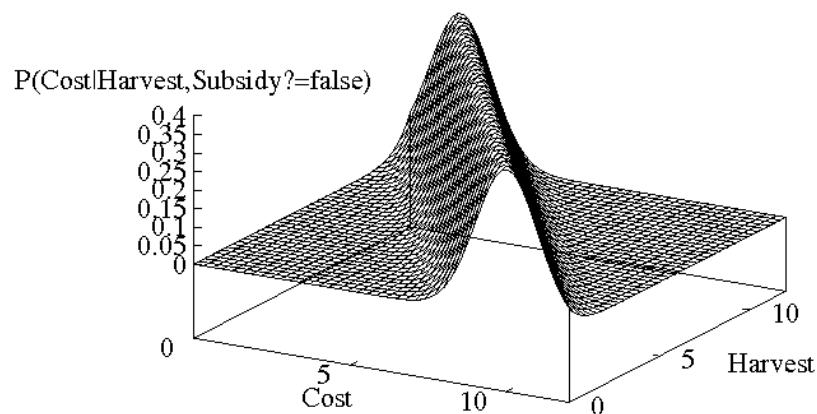
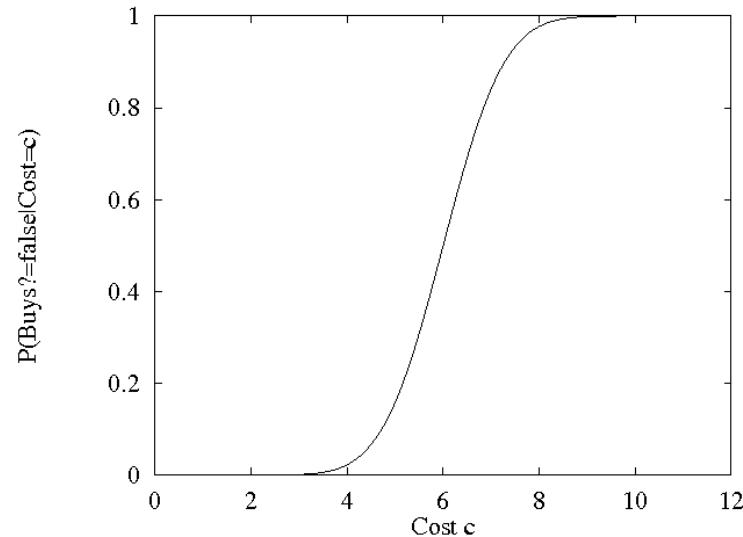
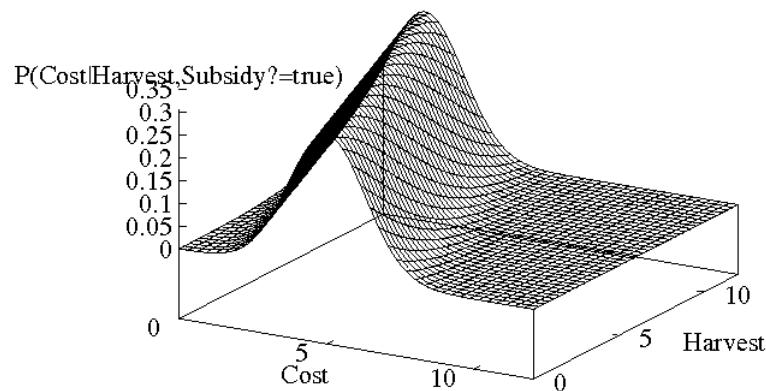
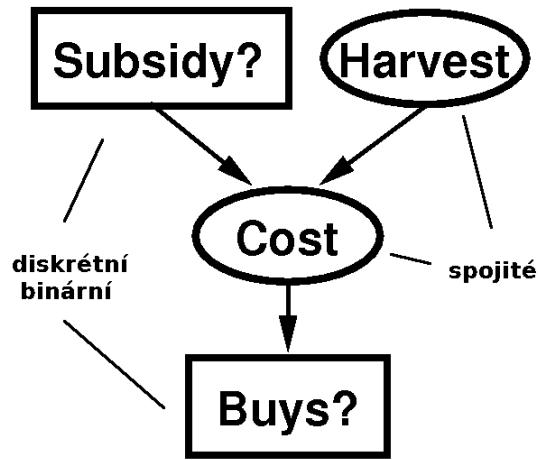


Spojity uzel jako dítě

- CPD pro spojité dítě C_p a kombinaci spojitého C_r a diskrétního D rodiče
 - nejobvyklejší je lineární gaussovský (LG) model
 - * střední hodnota je lineární fcí hodnot spojitého rodiče,
$$Pr(C_p = c_p | C_r = c_r, D = d) = N(\mu_d = a_d c_r + b_d, \sigma_d)$$
 - * je rozumným modelem, pokud se spojity rodič nemění příliš,
 - definujeme jednu LG funkci dítěte pro každou hodnotu diskrétního rodiče
 - * pro každou hodnotu d definujeme jinou sadu parametrů a_d , b_d a σ_d ,
- zobecnění pro více rodičů
 - střední hodnota je dána lineární kombinací hodnot spojitých rodičů,
 - nová fce CPD se musí definovat pro každou instanciaci diskrétních rodičů.

Spojité uzly – příklad (Russell, Norvig: AIMA)

- u ovoce určuje cena prodej, cena je dána sklizní a vládní dotací

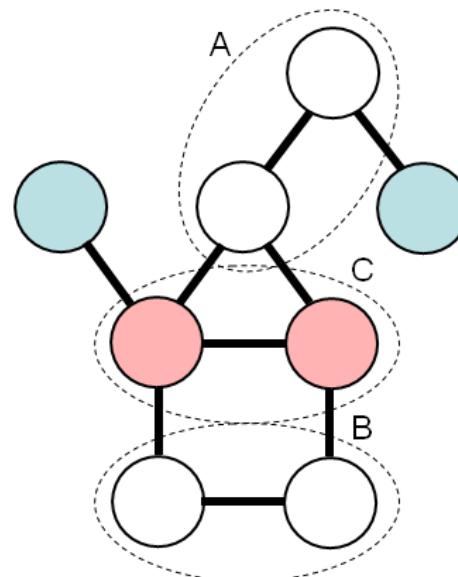


Spojité proměnné – přehled

- parametrizace spojitých proměnných a jejich CPDs
 - pomocí normálního rozdělení, jeho směsí, u CPD linearizace po instanciacích,
 - není obtížnější/náročnější než u proměnných diskrétních,
 - existuje řada variant uzlů a spojení (viz BNT for Matlab),
- celkový model sdružené pravděpodobnosti
 - pokud jsou uzly modelovány normálním rozdělením a CPDs LG modelem . . .
 - . . . sdružená pst má normální formu (více proměnných),
- způsoby usuzování se v principu nemění
 - nejsou použitelné všechny algoritmy,
 - je nutné zobecnit (mj. způsob nakládání s atomickými událostmi apod.).
- rozšiřuje se škála metod pro učení sítě
 - softmax je v BNT učen pomocí iteratively reweighted least squares (IRLS) algoritmu,
 - pro další typy vztahů také dedikované algoritmy.

Neorientované grafy

- neorientovaný grafický model = markovská síť = **markovské náhodné pole** (Markov Random Field – MRF),
 - stejný cíl jako orientované modely
 - rozložit (faktorizovat) sdruženou pst za využití podmíněné nezávislosti,
 - vyjádření podmíněné nezávislosti neorientovaným grafem
 - prosté grafické oddělení – jednodušší než d-oddělení u orientovaných modelů,
 - $A \perp\!\!\!\perp B|C$: odebráním množiny uzlů C z grafu zaniknou všechny cesty z A do B.



Neorientované grafy

- faktorizace sdružené psti vychází z definice podmíněné nezávislosti

- pro dva uzly nespojené hranou platí

$$Pr(P_i, P_j | P_{\setminus \{i,j\}}) = Pr(P_i | P_{\setminus \{i,j\}}) Pr(P_j | P_{\setminus \{i,j\}})$$

- Hammersley-Cliffordův teorém

$$Pr(P_1, \dots, P_n) = \frac{1}{Z} \prod_{C \in mcl\{G\}} \psi_C(P_C)$$

Z ... normalizační konst, $mcl\{G\}$... množina max klik grafu G , ψ_C ... potenciálová fce

- možné obtíže

- faktory (obvykle) nemají výklad (nejsou to psti),
 - součin faktorů musíme normalizovat (aby šlo o sdruženou pst),

- potenciálové funkce

- HC teorém předpokládá jejich striktní pozitivitu,
 - tradičně vyjádřeny exponenciálou funkce energie E

$$\psi_C(P_C) = \exp^{-E(P_C)}$$

- sdružená pst odpovídá součtu energií všech maximálních klik.

MRF: příklad (Bishop: Graphical models, chapter 8)

- odstranění šumu z binárního obrázku (2 barvy)

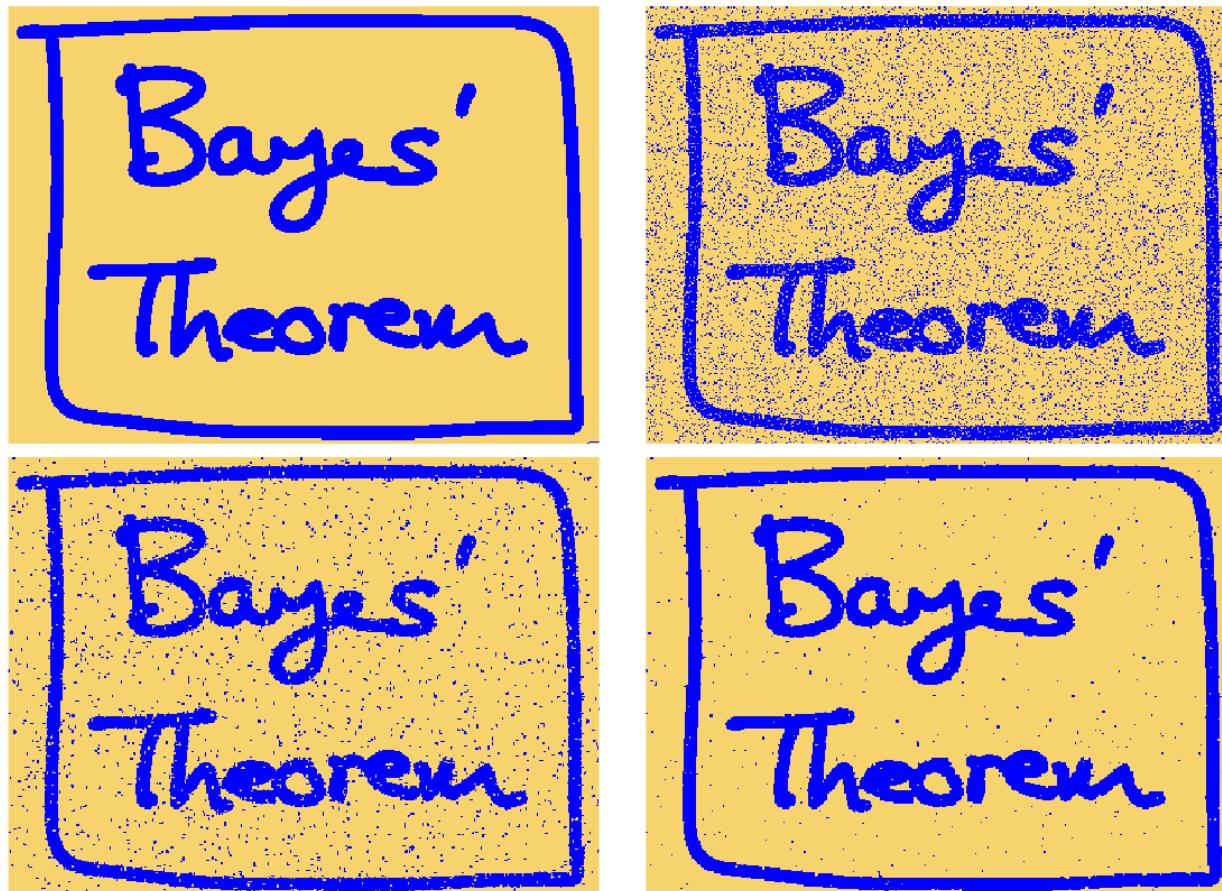
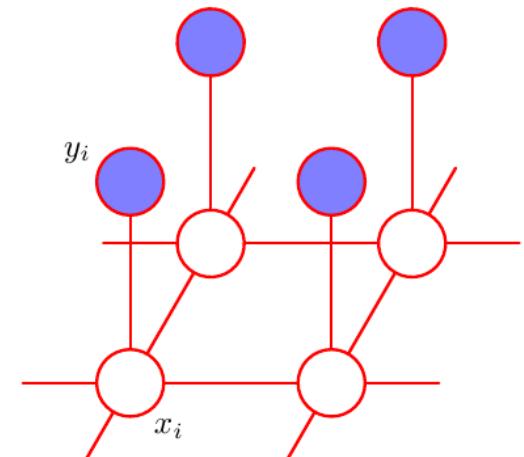


Figure 8.30 Illustration of image de-noising using a Markov random field. The top row shows the original binary image on the left and the corrupted image after randomly changing 10% of the pixels on the right. The bottom row shows the restored images obtained using iterated conditional models (ICM) on the left and using the graph-cut algorithm on the right. ICM produces an image where 96% of the pixels agree with the original image, whereas the corresponding number for graph-cut is 99%.

MRF: příklad (Bishop: Graphical models, chapter 8)

- návrh struktury modelu z apriorní znalosti
 - barva sousedních bodů silně korelovaná,
 - vzdálené uzly ignorujeme (první řád modelu)
 - * pravidelná mřížka uzlů X_i , $X_i \in \{-1, 1\}$
 - šum modelujeme další "vrstvou" uzlů Y_i , $Y_i \in \{-1, 1\}$
 - * slabý šum ($Pr(noise) \ll 0.5$) → silná korelace s X_i ,
 - dva typy maximálních klik o 2 uzlech
 - * sousední X_i a X_j , spojená Y_i a X_i ,
 - odpovídá často používanému Isingovu modelu.



MRF: příklad (Bishop: Graphical models, chapter 8)

- určení sdružené psti $Pr(\mathbb{X}, \mathbb{Y})$

- energie klik: $-\eta X_i Y_i$ a $-\beta X_i X_j$,
 - * η a β kladné parametrizační konstanty,
 - * při shodě znamének nižší energie a vyšší pravděpodobnost,
 - úplná funkce energie

$$E(\mathbb{X}, \mathbb{Y}) = h \sum_i X_i - \beta \sum_{\{i,j\}} X_i X_j - \eta \sum_i X_i Y_i,$$

* první člen upřednostňuje jednu barvu před druhou (pozadí je apriorně častější),

- cílová pravděpodobnost

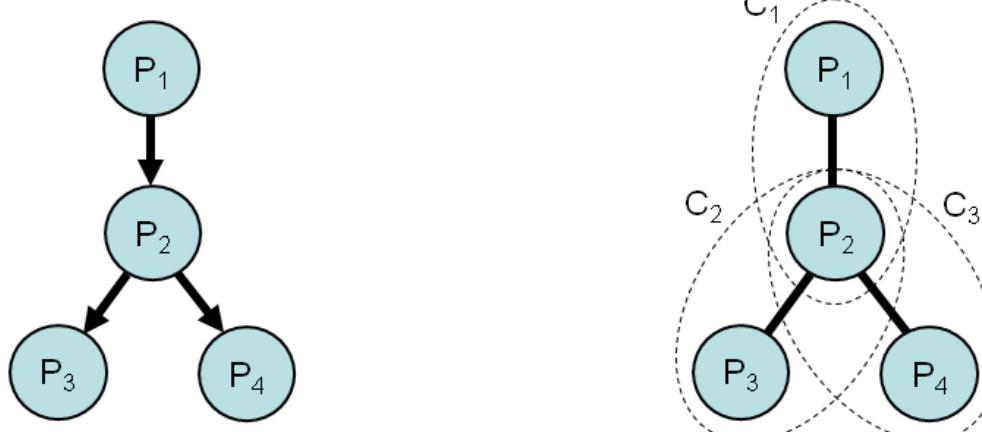
$$Pr(\mathbb{X}, \mathbb{Y}) = \frac{1}{Z} \prod_C \psi_C(P_C) = \frac{1}{Z} \exp -E(\mathbb{X}, \mathbb{Y})$$

- odstranění šumu = učení modelu

- nastavíme pozorované Y_i , hledáme X_i , aby $Pr(\mathbb{X}, \mathbb{Y})$ byla maximální,
 - lze provést gradientním algoritmem (který bere v úvahu souřadnice)
 1. inicializace \mathbb{X} ($\forall i X_i = Y_i$),
 2. $\forall i$ postupně a opakovaně (v náhodném pořadí nebo systematickým průchodem)
 - (a) urči celkovou energii pro $X_i = +1$ a $X_i = -1$,
 - (b) vyber stav s nižší energií,
 3. skonči pokud energie dále neklesla (po celém kole $\forall i$).

Převod od orientovaného k neorientovanému modelu

- důležitá součást junction tree inferenčního algoritmu,
 - nutno zajistit, aby potenciály klik odpovídaly podmíněným pstem v orientovaném grafu,
 - to je triviální v grafech, kde uzly mají jediného rodiče,



$$Pr(P_1, \dots, P_4) = \frac{1}{Z} \prod_i \psi_{C_i}(P_{C_i})$$

$$Pr(P_1, \dots, P_4) = \\ = Pr(P_1)Pr(P_2|P_1)Pr(P_3|P_2)Pr(P_4|P_2)$$

$$\psi_{C_1}(P_{C_1}) = \psi_{1,2}(P_1, P_2) = Pr(P_1)Pr(P_2|P_1)$$

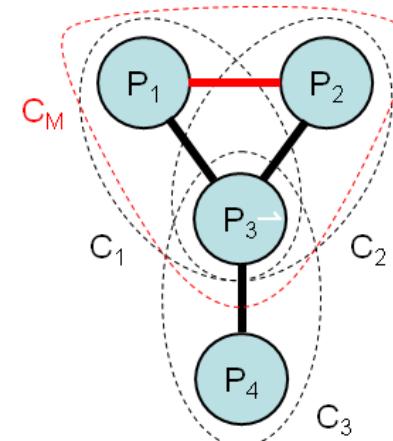
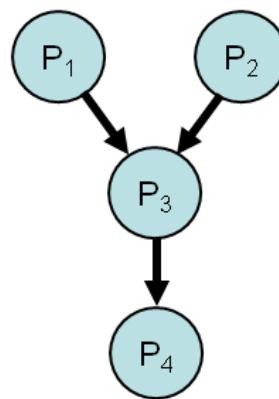
$$\psi_{C_2}(P_{C_2}) = \psi_{2,3}(P_2, P_3) = Pr(P_3|P_2)$$

$$\psi_{C_3}(P_{C_3}) = \psi_{2,4}(P_2, P_4) = Pr(P_4|P_2)$$

Z = 1

Převod od orientovaného k neorientovanému modelu

- u sítí s konvergentními uzly nutná moralizace (viz junction tree algoritmus),
 - jinak některé podmíněné psti nejsou použitelné ani v jedné z klik!
 - neorientovaný graf kóduje jinou množinu vztahů (podmíněné) nezávislosti než graf původní.



$$Pr(P_1, \dots, P_4) = \frac{1}{Z} \prod_i \psi_{C_i}(P_{C_i}) \quad (Z = 1)$$

$$\begin{aligned} Pr(P_1, \dots, P_4) &= \\ &= Pr(P_1)Pr(P_2)Pr(P_3|P_1, P_2)Pr(P_4|P_3) \end{aligned}$$

$$\begin{aligned} \psi_{C_M}(P_{C_M}) &= \psi_{1,2,3}(P_1, P_2, P_3) = \\ &= Pr(P_1)Pr(P_2)Pr(P_3|P_1, P_2) \\ \psi_{C_3}(P_{C_3}) &= \psi_{3,4}(P_3, P_4) = Pr(P_4|P_3) \end{aligned}$$

- převod od neorientovaného grafu k orientovanému obtížnější a méně častý.

Shrnutí

- BNs patří do velké třídy grafických stochastických modelů,
- zaměření BNs
 - znalostní systémy, kauzální diagramy,
- změny formalismu či omezení vedou na modely známé v různých oblastech
 - DBN vs HMM
 - * řeč, signály, časové řady,
 - BN vs MRF
 - * plyny, společnost, obrazy,
- veličiny mohou být diskrétní i spojité
 - z výukových důvodů jsme se soustředili na diskrétní.

Použití bayesovských sítí

- PATHFINDER (Stanford, 1980')
 - diagnóza chorob lymfatických uzlin,
 - testován na 53 případech pacientů, diagnostikoval lépe než několik z nejlepších patologů,
 - následován řadou lékařských diagnostických systémů,
- další diagnostické systémy/aplikace
 - interpretace elektromyografie (EMG),
 - plynové turbíny,
 - výnosy obilí,
- modelování dat genové exprese
 - vzájemné interakce mezi geny (gene regulatory networks),
 - G1 kód pro enzym který podporuje expresi jiného genu G2,
 - G1 "způsobí" G2.
- interaktivní řešení problémů v systému MS Windows
 - řeší HW i SW problémy,
 - podobně diagnostika poruch aut, tryskových motorů apod,
- modelování v geologii, hydrologii, sociologii,
- mj. BN Weka klasifikátor.

BN – shrnutí

- pravděpodobnost je rigorózním nástrojem pro modelování neurčitosti,
- každá **atomická** událost je popsána rozdelením sdružené pravděpodobnosti,
- dotazy zodpovíme agregací (součtem, popřípadě následným podílem) atomických událostí,
- pro netriviální domény je třeba sdruženou pravděpodobnost modelovat zjednodušeně,
- nástrojem je nezávislost a podmíněná nezávislost
 - na nich jsou založené **bayesovské sítě**,
 - i v nich je ale inference založená na vyčíslení psti NP-těžká k velikosti sítě,
 - řešením jsou speciální typy sítí (singly connected), např. stromy – pouze jeden rodič,
 - nebo přibližná inference,
- z pohledu učení (indukce sítí z dat) rozlišujeme dvě úlohy
 - jednoduší: optimalizaci parametrů existující sítě,
 - obtížnější: učení struktury sítě.

Výhody bayesovských sítí

- struktura vyjádřená grafem je přehledná a sémantika srozumitelná pro uživatele
 - přirozeně ji lze kauzálně interpretovat (příčina – následek) a také manuálně tvořit,
 - kauzální usuzování postupuje od příčin k následkům,
 - diagnostická usuzování slouží ke zjištění příčin k pozorovaným následkům,
 - pro srovnání: neuronová síť jako “černá skříňka”,
- schopnost spojovat apriorní znalost s daty
 - na základě dat lze automaticky upravit existující síť (např. vytvořenou expertem),
 - učení sítě výhradně z dat je speciálním případem výše uvedeného (prázdný init model),
- teoretická podloženost, robustnost, šíře využití
 - malé změny modelu mají zanedbatelný vliv na výsledek, současně se nepřeučuje,
 - použití v explorační analýze, ale i při klasifikaci či regresi,
- schopnost pracovat s chybějícími daty a skrytými proměnnými
 - model umí odhadnout pravděpodobnou hodnotu chybějící položky ve vztahu k ostatním,
 - skrytá proměnná může vysvětlit “podezřelou” závislost mezi dvěma pozorovanými veličinami,
- co stojí proti?
 - učení je obtížné, neexistuje všeobecně uznávaný a široce aplikovatelný algoritmus.

Doporučené doplňky – zdroje přednášky

:: Četba

- Russel, Norvig: **AI: A Modern Approach**, Uncertain Knowledge and Reasoning (Part IV)
 - reprezentace CPD se spojitymi promennymi,
 - pravdepodobnostni usuzovani v case (kap. 15, neni v prvni vydani),
 - Google books (isbn:0136042597),
- Murphy: **A Brief Introduction to Graphical Models and Bayesian Networks**.
 - co se spojitymi promennymi umi BNT?
 - <http://www.cs.ubc.ca/~murphyk/Bayes/bayes.html>,
- Bishop: **Pattern Recognition & Machine Learning**, Graphical Models (Chapter 8)
 - obecná kapitola, v této přednášce inspiruje pasáž o neorientovaných grafech,
 - <http://research.microsoft.com/en-us/um/people/cmbishop/prml/Bishop-PRML-sample.pdf>.
- Forbes et al.: **The BATmobile: Towards a Bayesian Automated Taxi**, IJCAI 1995
 - TAXI příklad,
 - <http://www.cs.berkeley.edu/~russell/papers/ijcai95>.

