Úvod do strojového učení, učení z příkladů

Jiří Kléma

Katedra kybernetiky,
FEL, ČVUT v Praze

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

Plán přednášky

- Motivační příklad
  - jak se učí lidé?
- Definice a role strojového učení
  - proč, využití.
- Základní termíny
  - pojem, hypotéza, generalizační schopnost,
  - zaujetí, klasifikační přesnost.
- Příklady algoritmů
  - základní rozdíly.
- Shrnutí
  - kategorizace metod,
  - souvislost se statistickými modely.

Motivační příklad [Kubát, UI1]

:: Cíl: naučit mimozemšťana pojmu **pták**.

Příklad 1: kos (+)
- létá, má křídla
- žlutý zobák
- zpívá
- snáší vajíčka

Příklad 2: pes (-)
- běhá, nemá křídla

Zobecnění 3:
- pták je vše co má křídla a zobák

Příklad 4: vrána (+)
- létá, má křídla
- má zobák

Zobecnění 1:
- pták je vše co má křídla

Příklad 3: moucha (-)
- létá, má křídla
- nemá žlutý zobák

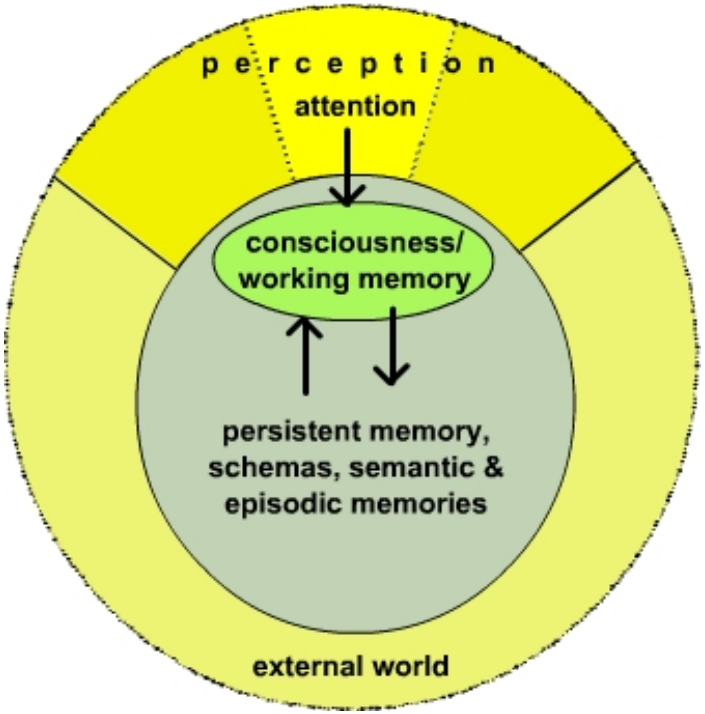
Zobecnění 2:
- pták je vše co má křídla a žlutý zobák
- (snáší vajíčka se skořápkou)
- (má peří)

Motivační příklad [Kubát, UI1]

- Učení se pojmu (konceptu),
- učení s učitelem
  - inteligentně předkládal příklady s vysvětlením tak, aby žák co nejrychleji pochopil a korigoval svůj popis,
- žák hledá popis pojmu (hypotézu)
  - popis by měl pokrýt všechny myslitelné pozitivní instance daného pojmu,
  - a žádnou negativní instanci,
- využívá k tomu generalizace a specializace
  - pozitivní příklad = generalizace (zobecnění),
  - negativní příklad = specializace (vymezení, omezení),
- šlo o učení z příkladů, doplněné vysvětlováním
  - učitel vybíral pouze důležité charakteristiky a doplňoval je vysvětlením relevantním pro danou chvíli,
- inkrementální učení.

Jak se učí lidé?

- Učení probíhá v kontextu dalších kognitivních funkcí
  - pozornost, vnímání, paměť, uvažování, ...
- tato integrace je u umělých systémů velmi obtížná, interakce s okolím pouze zprostředkovaná,
- samo učení není přímo pozorovatelné – jde o proces změn ve vnitřní reprezentaci znalostí,



©Cognitive perspectives, Michigan State University.

Definice

- Učení
  - široký a relativně obtížně definovatelný pojem
  - (podobně jako inteligence)
  - Učení je proces založený na získávání a využívání zkušeností vedoucí k dlouhodobým změnám možného chování. (Wikipedia)
- Strojové učení jako vědní obor
  - Jde o obor zaměřující se na induktivní a další algoritmy, o kterých můžeme prohlásit, že se "učí". (ML Journal, 1998, Glossary of Terms, by Kohavi and Provost)
  - Jde o obor zabývající se otázkou, jak konstruovat počítačové programy, které se automaticky zdokonalují nad danou třídou úloh na základě zkušenosti. (Tom Mitchell: Machine Learning, McGraw-Hill, 1997)
- Strojové učení jako schopnost
  - Jde o schopnost zařízení zlepšovat svoji výkonnost na základě výkonnosti dřívější.

Proč strojové učení?

- Obecné důvody
  - bez učení není inteligence,
  - pokrok ve vývoji algoritmů a teorii,
  - rostoucí objemy (on-line) dat,
  - dostupné výpočetní prostředky.
- Využití
  - dolování dat
    - využití historických dat ke zlepšení příštích rozhodnutí,
    - příklad: lékařské záznamy $\longrightarrow$ znalost; dále marketing, finance, technická diagnostika, logistika, ...
  - SW aplikace, jež nelze programovat (zcela) manuálně
    - autonomní řízení, rozpoznávání řeči,
  - adaptivní aplikace
    - příklad: udržování a využívání uživatelského profilu.

Rámec pro učení

- Formální definice pracuje s entitami
  - třída úloh T (vymezení pozornosti)
  - výkonnost V (určení kvality učeného popisu)
  - zkušenost Z (forma, v jaké je reprezentována)
- Návrat k motivačnímu příkladu
  - T: pojem pták
  - V: klasifikační přesnost (poměr správně označených instancí)
  - Z: trénovací instance popsané symbolickými či numerickými atributy

Rámec pro učení

- Formální definice pracuje s entitami
  - třída úloh T (vymezení pozornosti)
  - výkonnost V (určení kvality učeného popisu)
  - zkušenost Z (forma, v jaké je reprezentována)
- Návrat k motivačnímu příkladu
  - T: pojem pták
  - V: klasifikační přesnost (poměr správně označených instancí)
  - Z: trénovací instance popsané symbolickými či numerickými atributy

ID	létá	má_křídla	snáší_vejce	skořápka	zobák	barva_zobáku	pták
kos	dobře	ano	ano	ano	ano	žlutá	ano

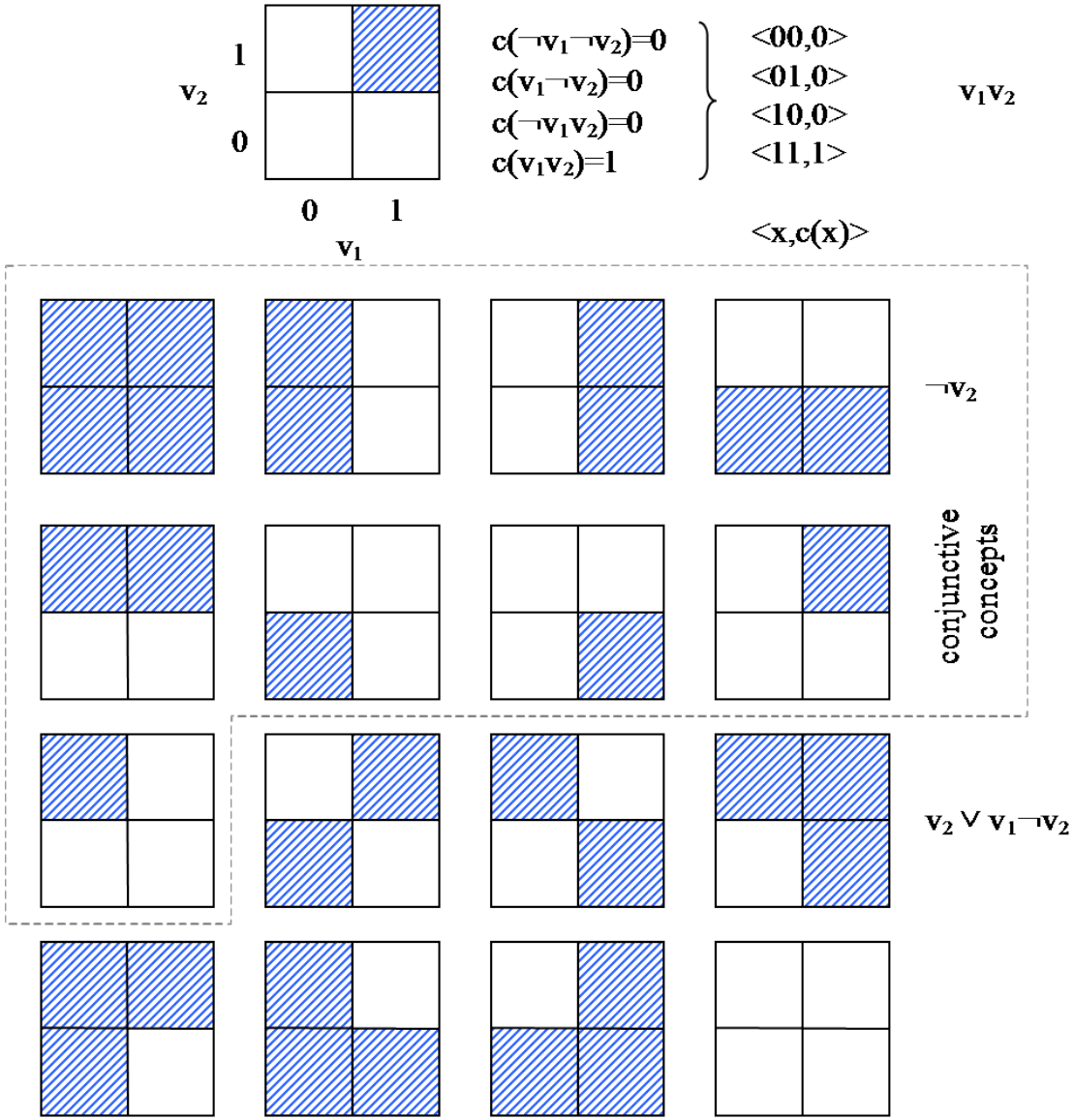
Induktivní učení konceptů

- Indukce
  - na základě vlastností individuí usuzujeme na vlastnosti jejich množin (od popisu individuí k popisu pojmu),
  - (logická) generalizace + (statistické) ověření na datech,
- Koncept (pojem)
  - X – množina všech instancí, univerzum,
  - koncept C je podmnožinou X, tj. $C \subseteq X$,
  - tj., bool. fce nad X: $\forall x \in X$, $C(x) \rightarrow \{0, 1\}$,
  - v našem případě: X jsou živočichové, C jsou ptáci
- Trénovací příklady
  - podmnožina univerza, $E \subseteq X$,
  - pozitivní a negativní příklady, $\forall x \in E : C(x) \Rightarrow E^+(x)$, $\neg C(x) \Rightarrow E^-(x)$,
  - v našem případě: kos, pes, moucha, vrána.
- Hypotéza
  - pokus o jednoznačný popis pojmu ve zvolené reprezentaci,
  - $\forall x \in X$, kde x je vektor vlastností individua, rozhodne, zda $x \in C$.

Induktivní učení konceptů

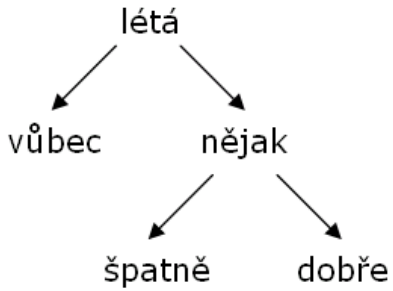
- Hypotéza konzistentní na množině příkladů
  - správně hodnotí příklady (shoda s cílovým pojmem),
  - konzistentní(h,E) $\equiv (\forall \langle x, c(x) \rangle \in E) : h(x) = c(x)$.
- Prostor verze
  - prostor všech konzistentních hypotéz.
- Algoritmy učení se mohou lišit v:
  - cílech, reprezentaci, v dostupných trénovacích datech i ve strategii učení.
- V čem se ale neliší je to, že
  - učení vždy probíhá na základě prohledávání prostoru hypotéz,
  - hledáme korektní generalizaci, resp. generalizaci s nejmenší chybou.
- Velikosti prohledávaných prostorů
  - máme-li n binárních atributů, existuje $|X| = 2^n$ instancí,
  - každá instance může buď patřit nebo nepatřit ke konceptu $|V_0| = 2^{2^n}$,
  - úplné prohledávání je NP úlohou,
  - tj. není reálné vyhodnocovat a uchovávat všechny (korektní) hypotézy.

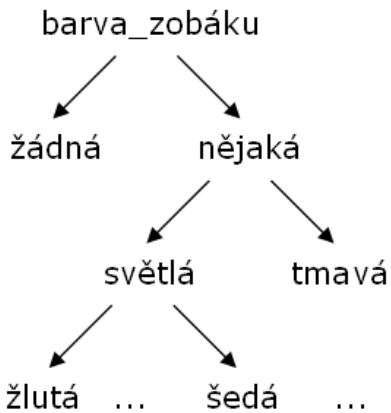
Prostor hypotéz – příklad pro $\{0,1\}^2$



Příklady a hypotézy

- Pracujeme se dvěma formálními jazyky
  - L_E jazyk definující instance,



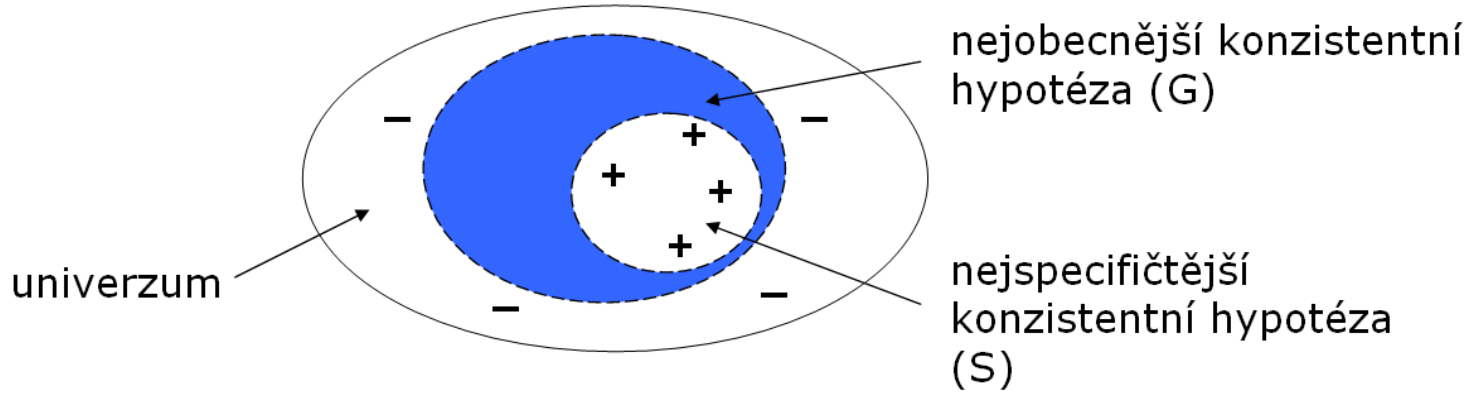


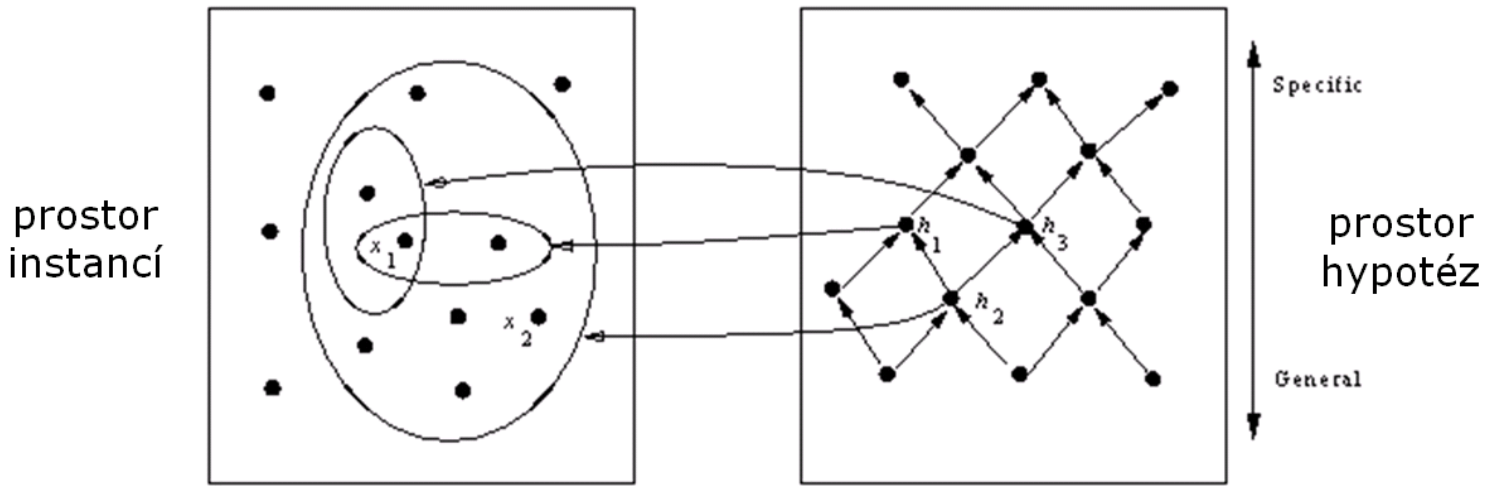
  - L_H jazyk definující hypotézy
    - např. připouštíme disjunktivně-konjunktivní hypotézy typu:
      if (má_křídla and zobák=žlutý) or (snáší_vejce and skořápka) **then** pták
    - nebo pouze konjunktivní hypotézy: pak ovšem nelze popsat pojmy typu:
      if (zobák=žlutý) or (zobák=šedý) **then** pták
- Terminologie
  - příklad e = (PopisObjektu, OznačeníTřídy),
  - pozitivní příklad e^+ – popis pozitivní instance pojmu C,
  - negativní příklad e^- – popis negativní instance pojmu C,
  - množina trénovacích příkladů $E : E = E^+ \cup E^-$,
  - konzistence: h(PopisObjektu)=OznačeníTřídy.

CANDIDATE-ELIMINATION alg.

- Efektivně reprezentuje prostor verze
  - neukládá všechny konzistentní hypotézy,
  - hypotézy jsou setříděny od nejobecnější (G) k nejspecifičtější (S),
  - ostatní hypotézy nejsou uloženy.
- Musíme definovat relaci *obecnější* nebo *stejně obecná hypotéza*
  - nechť h_i a h_k jsou booleovské fce definované nad prostorem instancí X,
  - pak h_j je obecnější nebo stejně obecná jako h_k $(h_j \geq h_k)$ právě tehdy když: $(\forall x \in X)[(h_k(x) = 1) \Rightarrow (h_j(x) = 1)]$
- Nejobecnější hypotéza
  - zapíšeme jako $\langle ?, ?, ?, ?, ?, ? \rangle$,
  - v našem příkladě: všichni živočichové jsou ptáci.
- Nejspecifičtější hypotéza
  - zapíšeme jako $\langle \emptyset, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset \rangle$,
  - v našem příkladě: žádný živočich není pták.

Obecnost hypotéz





CANDIDATE-ELIMINATION alg.

:: Pro všechny trénovací příklady x_i proveď:

- Je-li x_i pozitivní příklad
  - odstraň z G všechny hypotézy nekonzistentní s x_i,
  - pro všechny hypotézy s z S nekonzistentní s x_i
    - odstraň s z S,
    - přidej do S všechny min. generalizace h daného s takové, že h je konzistentní s x_i a existuje $g \in G$, které je obecnější než h,
    - odstraň z S všechny hypotézy obecnější než jakákoli jiná hypotéza z S.
- Je-li x_i negativní příklad
  - odstraň z S všechny hypotézy nekonzistentní s x_i,
  - pro všechny hypotézy g z G nekonzistentní s x_i
    - odstraň g z G,
    - přidej do G všechny min. specializace h daného g takové, že h je konzistentní s x_i a existuje $s \in S$, které je specifičtější než h,
    - odstraň z G všechny hypotézy méně obecné než jakákoli jiná hypotéza z G.

CANDIDATE-ELIMINATION př.

ID	létá	má_křídla	snáší_vejce	skořápka	zobák	barva_zobáku	pták
kos	dobře	ano	ano	ano	ano	žlutá	ano
pes	vůbec	ne	ne	ne	ne	žádna	ne
moucha	dobře	ano	ano	ne	ne	žádná	ne
vrána	dobře	ano	ano	ano	ano	šedá	ano

$G_0 : \{\langle ?, ?, ?, ?, ?, ? \rangle\}$, $S_0 : \{\langle \emptyset, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset \rangle\}$

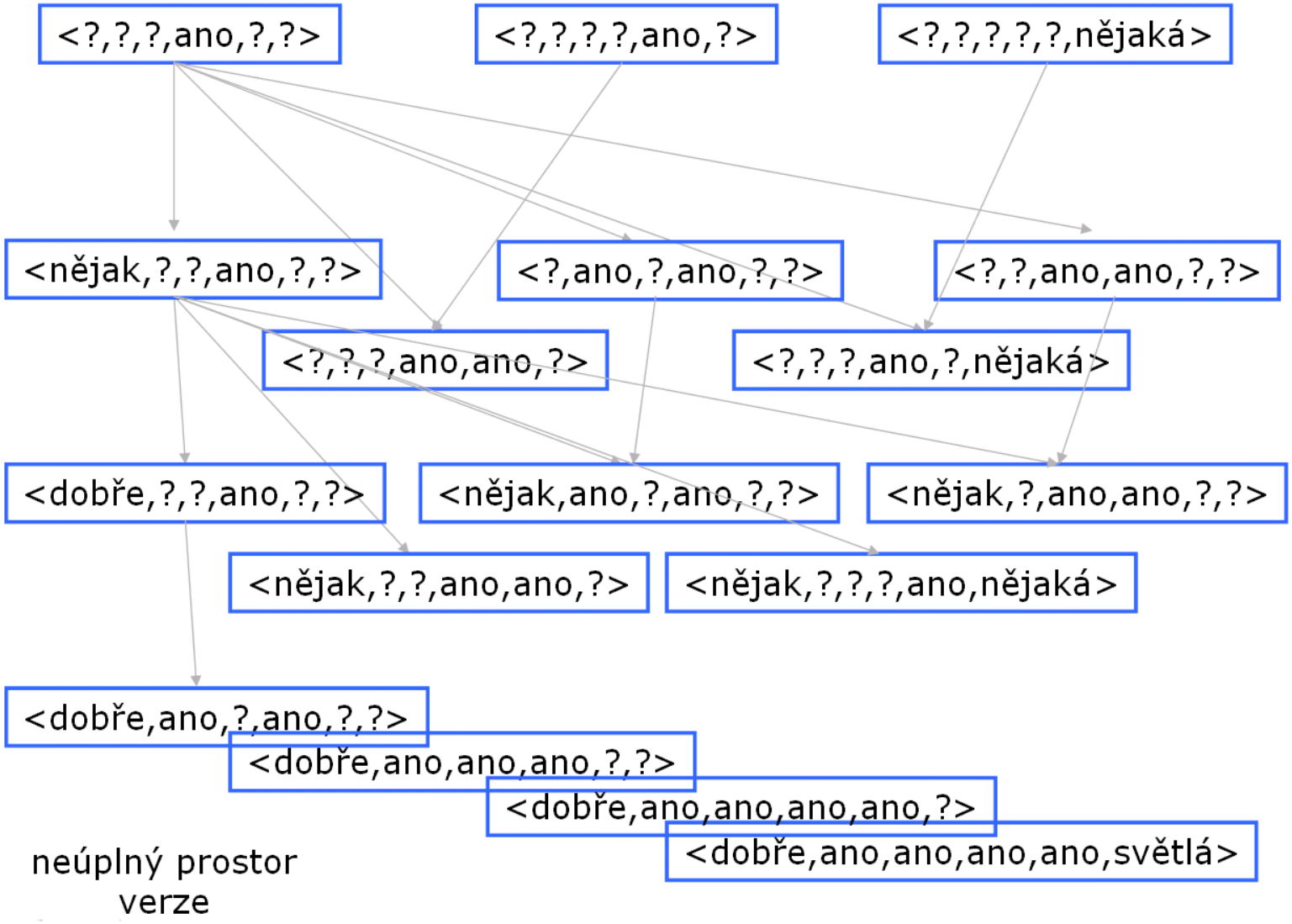
$G_1 : \{\langle ?, ?, ?, ?, ?, ? \rangle\}$, $S_1 : \{\langle dobre, ano, ano, ano, ano, zluta \rangle\}$

$G_2 : \{\langle nejak, ?, ?, ?, ?, ? \rangle, \langle ?, ano, ?, ?, ?, ? \rangle, \langle ?, ?, ano, ?, ?, ? \rangle, \langle ?, ?, ?, ano, ?, ? \rangle, \langle ?, ?, ?, ?, ano, ? \rangle,$
$\langle ?, ?, ?, ?, ?, nejaka \rangle\}, S_2 = S_1$

$G_3 : \{\langle nejak, ?, ?, ano, ?, ? \rangle, \langle nejak, ?, ?, ?, ano, ? \rangle, \langle nejak, ?, ?, ?, ?, nejaka \rangle, \langle ?, ano, ?, ano, ?, ? \rangle,$
$\langle ?, ano, ?, ?, ano, ? \rangle, \langle ?, ano, ?, ?, ?, nejaka \rangle,$
$\langle ?, ?, ?, ano, ?, ? \rangle, \langle ?, ?, ?, ?, ano, ? \rangle, \langle ?, ?, ?, ?, ?, nejaka \rangle\}, S_3 = S_2$

$G_4 : \{\langle ?, ?, ?, ano, ?, ? \rangle, \langle ?, ?, ?, ?, ano, ? \rangle, \langle ?, ?, ?, ?, ?, nejaka \rangle\},$
$S_4 : \{\langle dobre, ano, ano, ano, ano, svetla \rangle\}$

Prostor konzistentních hypotéz



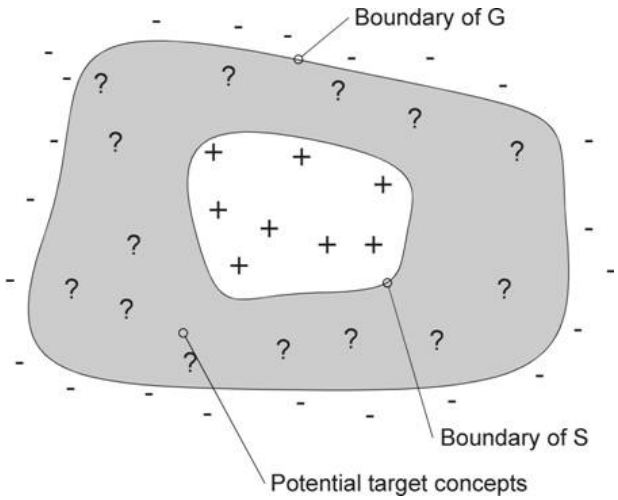
neúplný prostor verze

CANDIDATE-ELIMINATION alg.

- Aktuální prostor verze nezávisí na pořadí příkladů
- Konverguje ke správné hypotéze pokud
  1. trénovací příklady jsou bezchybné,
  2. správná hypotéza je obsažena v prostoru hypotéz (reprezentace).
- Nalezení správné hypotézy
  - G=S.
- Chyby v trénovacích příkladech
  - pouze detekuje tím, že konverguje k prázdnému prostoru verze.
- Není (přímo) použitelný v řadě praktických úloh
  - numerické veličiny – diskretizace,
  - obecná funkce namísto pojmu (bool. fce)
    - kategoriální funkce – rozklad na více pojmů,
    - reálná funkce – zcela jiné postupy.
- Jde více o ilustrativní než prakticky používaný algoritmus.

Zaujetí (inductive bias)

- Jak si "pojistit" to, že pojem je v prostoru hypotéz?
- Lze pracovat s úplným hypotézovým prostorem?
  - obsahuje všechny myslitelné hypotézy.
- Možné kontraindikace
  - exponenciální velikost prostoru verze,
  - schopnost zobecnění pro neznámé instance,
  - nároky na počet trénovacích příkladů.

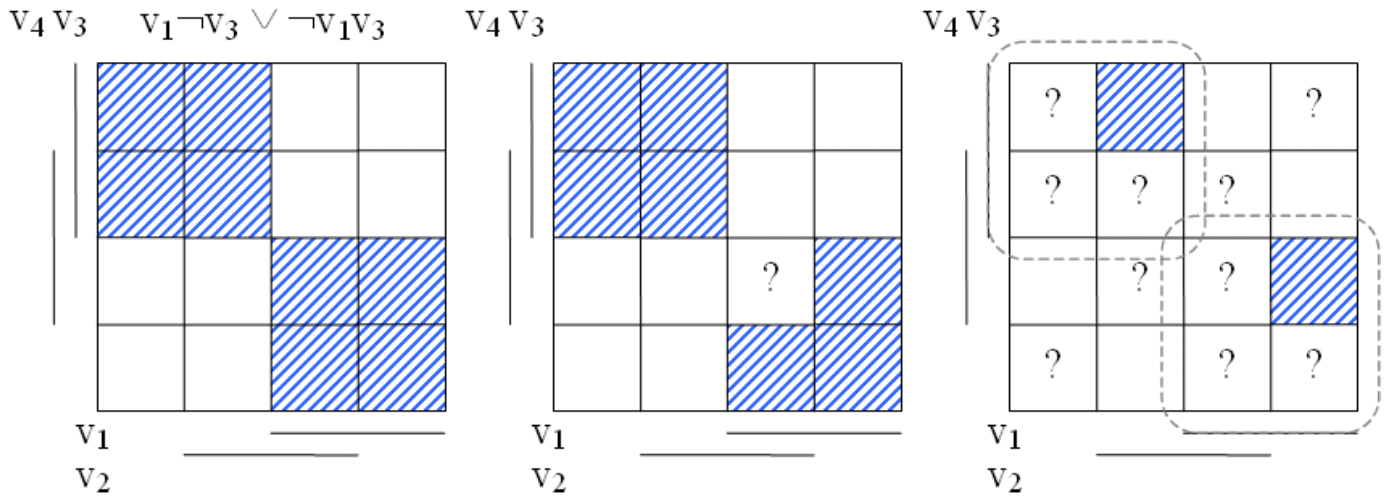


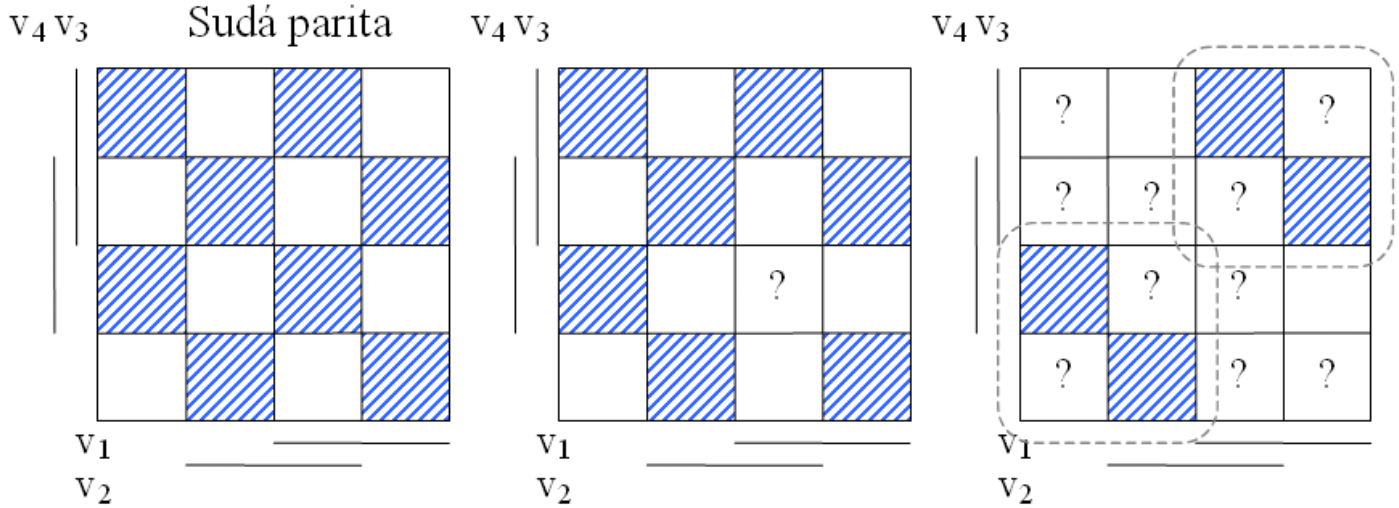
- Ad zobecnění u CE algoritmu:
  - G: negovaná disjunkce negativních trénovacích příkladů,
  - S: disjunkce pozitivních trénovacích příkladů,
  - důsledek: jednoznačně klasifikujeme pouze trénovací příklady!
    $E^+ : (x_1, x_2, x_3), E^- : (x_4, x_5) \rightarrow G : \{\neg(x_4 \vee x_5)\}, S : \{(x_1 \vee x_2 \vee x_3)\}$
- Nutné použít **zaujatý** algoritmus
  - pracuje s omezeným prostorem hypotéz (např. pouze konjunktivní hypotézy),
  - nebo je schopen preference.

Zaujetí (inductive bias)

- Reálně používané algoritmy
  - **Occamova břitva** (Occam's razor)
    - "Plurality should not be assumed without necessity." (William of Occam, 14. století)
    - Moderní interpretace: máme-li dvě hypotézy, z nichž obě vysvětlují nějaké naše pozorování, preferuj tu jednodušší.
    - Často používané, ačkoli také neméně často diskutované:
    - Webb, G.: Further Experimental Evidence against the Utility of Occam's razor.
- V čem se potom různé algoritmy strojového učení vlastně liší:
  - v tom jak hypotézy reprezentují,
  - logicky následně v tom, které v dané reprezentaci považují za jednodušší.
- Příklad učícího se algoritmu nad booleovskými koncepty:
  - reprezentace disjunkce konjunkcí, preferujeme kratší hypotézy,
  - dobrá generalizační schopnost – koncept, který lze vhodně vyjádřit zvolenou reprezentací, např. $\neg v_1 v_3 \vee v_1 \neg v_3$,
  - špatná generalizační schopnost – např. sudá parita, nelze jednoduše vyjádřit.

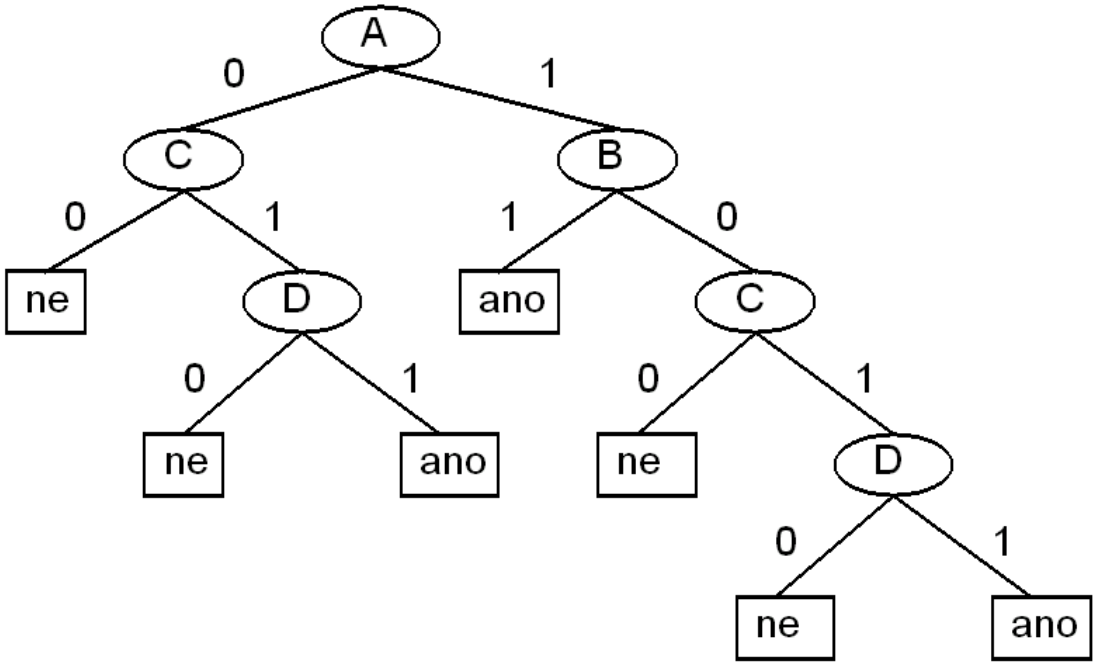
Příklad 1: $\neg v_1 v_3 \vee v_1 \neg v_3$, Příklad 2: sudá parita





Úlohy nevhodné pro stromy

- stromová reprezentace hypotéz,
- disjunkce konjunkcí, větev = konjunkce,
- induktivní zaujetí – Occamova břitva – preferuj menší stromy
- př: binární atributy A, B, C, D, cílový koncept nechť je definován takto:
  if A=B=1 or C=D=1 **then** $c^*(x_i) = 1$ (yes) **otherwise** $c^*(x_i) = 0$ (no)

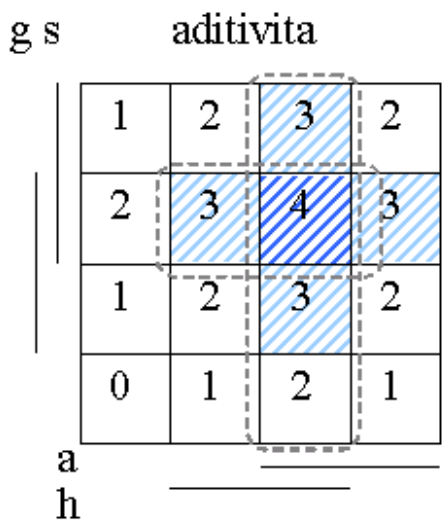


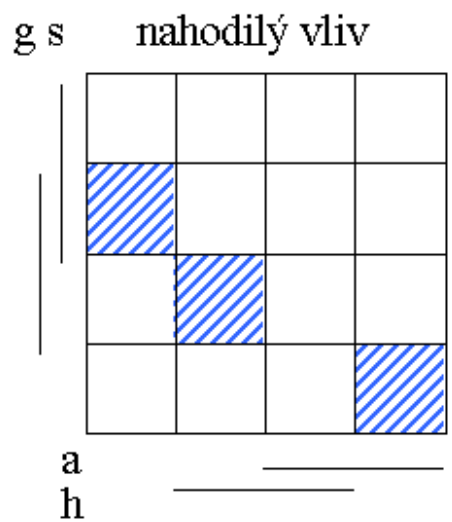
Jsou některé třídy konceptů v PRAXI častější?

- často diskutovaná, ale obtížně obecně řešitelná otázka,
- argument pro: běžné jevy mají často podobný charakter,
- příklad: diagnostika **cirhózy**
  - veličiny: abusus alkoholu, hepatitida, genetické předpoklady, nezdravá strava,
  - objevení nemoci: vlivy veličin mají přirozeně aditivní charakter, kombinace negativních jevů (např. alkoholu a genetických předpokladů) nikdy nemá pozitivní vliv,
  - projevy v konceptových mapách ... jakou třídu konceptů lze očekávat?
  - platí pro běžné DNF koncepty Occamova břitva (penalizuje koncepty typu parita)?

$AH \vee GSH$

nemocní měli hepatitidu a buď nadužívají alkohol nebo mají genetické předpoklady a nevhodně se stravují



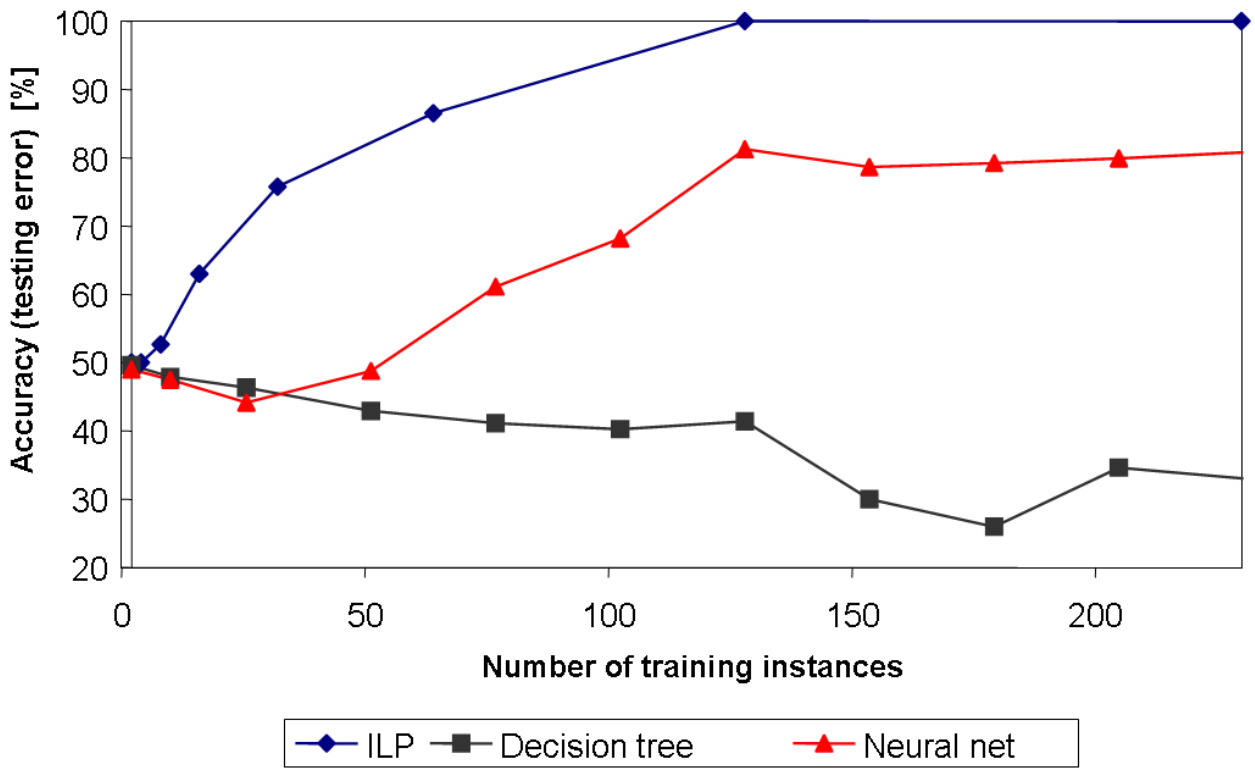


$\overline{AH}GS \vee \overline{A}HG\overline{S} \vee A\overline{HGS}$

nepravděpodobný a obtížně vysvětlitelný koncept

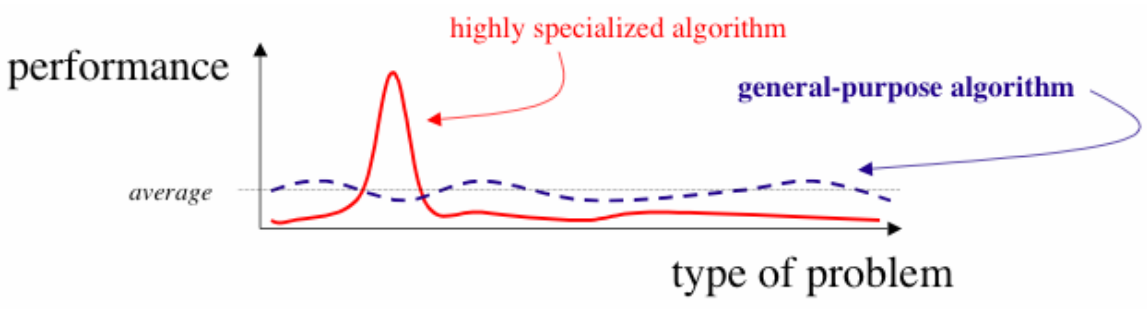
Křivka učení

- způsob, jak vyhodnotit schopnost učit se,
- znovu sudá parita, $X = \{01\}^8$ (256 instancí),
- tři odlišné algoritmy aplikované s odlišným výsledkem.



Širší souvislosti

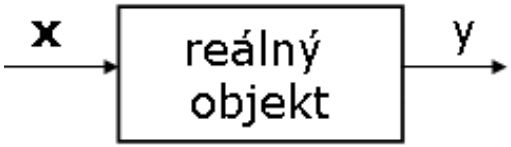
- **"No free lunch"** teorém
  - obědy zadarmo ve skutečnosti neexistují,
  - Robert Heinlein, román The Moon Is a Harsh Mistress, 1966,
  - osoba ani společnost nemůže získat hodnotu bez protihodnoty.
- Prohledávání a optimalizace
  - obecná a univerzální optimalizační strategie je teoreticky vyloučená, nicméně platí, že jedna strategie může být lepší než jiná pokud je zaměřená na konkrétní problém.



- Bude někdy nalezen jediný vždy nejlepší algoritmus učení?
  - všechny klasifikátory mají stejnou generalizační schopnost pokud je porovnáme přes všechny možné koncepty.

Souvislost se statistikou

- Statistické datové modely
  - klasický příklad: lineární či logistická regrese,
  - vyhraněnější orientace – numerické modely,
  - stejný vstup – příklady, instance = data,
  - stejný cíl – generalizace
    - model pro predikci či deskripci,
  - vycházejí z předpokladů
    - data jsou generována daným statistickým modelem,
    - je třeba optimalizovat jeho parametry a testovat platnost,
  - odlišný způsob validace modelů
    - testy dobré shody a residuální analýza vs. prediktivní přesnost,
  - jistý překryv: statistické metody učení.



Závěr

- Učení jako prohledávání stavového prostoru
  - efektivní CE algoritmus
- Důležitost zaujetí u každého algoritmu
- Role integrace s dalšími kognitivními funkcemi
- Zaměření na učení z příkladů, mimo to i
  - učení zapamatováním,
  - učení se z instrukcí,
  - učení se pozorováním a objevováním,
  - učení se z analogie,
  - učení na základě vysvětlení,
- Zaměření na učení s učitelem
  - učení bez učitele (shluková analýza),
  - příště – **posilované učení** – učení se z interakce s prostředím.

Doporučené doplňky – zdroje přednášky

:: Četba

- Tom Mitchell: **Machine Learning**, McGraw-Hill, 1997.
  - ucelený přehled metod strojového učení,
  - slidy: http://www.cs.cmu.edu/%7Etom/mlbook-chapter-slides.html
- Mařík a kol. **Umělá inteligence 1**
  - kapitola Strojové učení: základní pojmy, rozhodovací stromy, AQ algoritmus, učení bez učitele.
- Mařík a kol. **Umělá inteligence 4**
  - kapitola Strojové učení v dobývání znalostí: konkrétní příklady algorimů a jejich použití při dolování dat.