

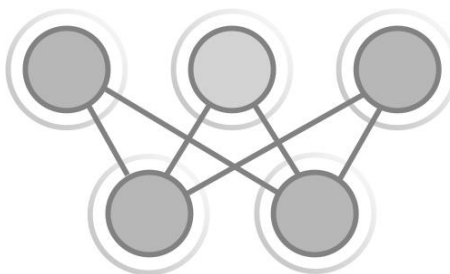
Prof.dr.sc. Bojana Dalbello Bašić

Fakultet elektrotehnike i računarstva
Zavod za elektroniku, mikroelektroniku, računalne i inteligentne sustave

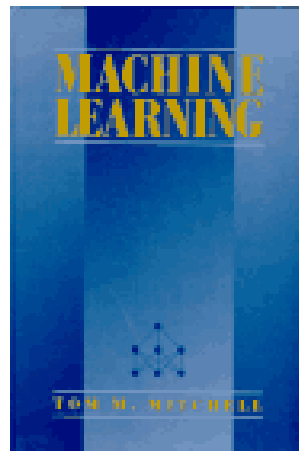
www.zemris.fer.hr/~bojana
bojana.dalbello@fer.hr

Strojno učenje

Učenje koncepata



- *Chapter 2*
Concept Learning and
the General-to-specific Ordering



2. UČENJE KONCEPATA I UREĐAJ HIPOTEZA OD OPĆENITE PREMA POJEDINAČNOJ

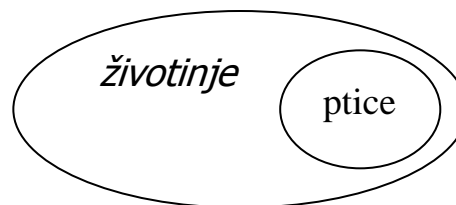
Učenje koncepata

- uobličavanje definicije **općenite klase/kategorije** ako je dan skup **pozitivnih i negativnih primjera** te kategorije.
- Kako?
Pretraživanjem prostora potencijalnih hipoteza za onom koja najbolje odgovara primjerima za učenje → koristeći strukturu uređaja hipoteza od općenite prema pojedinačnoj

- Što je to koncept?
- Primjer: općeniti koncepti ili kategorije:
 - “ptica”,
 - “auto”,
 - “situacije u kojima bih trebao učiti više da bi prošao ispit” itd.

**Učenje → uobličavanje općenite
kategorije na temelju primjera**

- **koncept** → podskup objekata (događaja) nekog većeg skupa (*primjer. skup ptica \subset skup životinja*)



ili alternativno

- **Koncept (klasa)** → funkcija koja poprima boolove vrijednosti definirana na nekom nadskupu. (*primjer: funkcija definirana na skupu svih životinja... **HIPOTEZA***)

Učenje koncepata je izvođenje funkcije boolovih vrijednosti na temelju skupa pozitivnih i negativnih primjera.

ZADATAK UČENJA KONCEPATA

- Učenje koncepta:
- “Dan kada moj prijatelj Aldo uživa u omiljenom vodenom sportu”.
- Skup primjera za učenje u obliku tablice atributa:

| <i>Primjer - dan</i> | <i>Naoblaka</i> | <i>Temperat ura</i> | <i>Vlažnost</i> | <i>Vjetar</i> | <i>Voda</i> | <i>Prognoza</i> | <i>dan za sport</i> |
|--------------------------|-----------------|-------------------------|-----------------|---------------|-------------|-----------------|-------------------------|
| 1 | sunčano | vruće | normalna | jak | topla | isto | DA |
| 2 | sunčano | vruće | visoka | jak | topla | isto | DA |
| 3 | kišno | hladno | visoka | jak | topla | promjena | NE |
| 4 | sunčano | vruće | visoka | jak | hladna | promjena | DA |

- **Zadatak:** predvidjeti vrijednost atributa **dan_za_sport** za proizvoljan dan, na temelju vrijednosti ostalih njegovih atributa (naoblaka, temperatura...)
- **Kako ćemo predstaviti hipotezu?**

ZADATAK UČENJA KONCEPATA

- Predstavljanje hipoteze:
Hipoteza = konjunkcija uvjeta na vrijednosti atributa
- U obliku vektora:

$$\text{Hipoteza } \mathbf{h} = \begin{bmatrix} \textit{Naoblaka} \\ \textit{Temperatura zraka} \\ \textit{Vlažnost} \\ \textit{Vjetar} \\ \textit{Voda} \\ \textit{Prognoza} \end{bmatrix}$$

- Za svaki atribut hipoteza može sadržavati:
 - ? - znači da je svaka vrijednost prihvatljiva
 - određena vrijednost atributa, npr. *Vlažnost*: visoka
 - \emptyset - nijedna vrijednost nije prihvatljiva

ZADATAK UČENJA KONCEPATA

- Primjer hipoteze h : “*Aldo uživa u sportu samo na hladne dane s visokom vlagom*” (bez obzira na druge vrijednosti atributa):

- $h = (?, \text{ hladno}, \text{ visoka}, ?, ?, ?)$

Primjeri:

- $(?, ?, ?, ?, ?, ?)$ – najopćenitija hipoteza
 - $(\emptyset, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset)$ – najspecifičnija hipoteza – niti jedan dan nije pozitivan primjer
- Ako su za neki primjer x svi uvjeti hipoteze h zadovoljeni, primjer se klasificira kao pozitivan primjer

$$h(x) = 1$$

Zadaća učenja konceptata (notacija):

- **X skup primjera** = skup svih dana opisan sa 6 atributa
- **c ciljna funkcija ili ciljni koncept** koji se uči.
c je funkcija, $c: X \rightarrow \{0, 1\}$.

$c(x) = 1$ ako **Dan_za_sport** = Da

i $c(x) = 0$ ako **Dan_za_sport** = Ne, $x \in X$.

- **H skup svih potencijalnih hipoteza**.
Hipoteze - funkcije boolovih vrijednosti, $h: X \rightarrow \{0, 1\}$.
(Dizajner određuje reprezentaciju hipoteze)

(?, hladno, visoka, ?, ?, ?)

h je opisana s konjunkcijom uvjeta

ZADATAK UČENJA KONCEPATA

- **D** skup primjera za učenje $(x, c(x))$.
 - pozitivni primjeri ($c(x) = 1$), negativni primjeri ($c(x) = 0$)

Primjer:

((sunčano, vruće, normalno, jak, topla, ista), 1)

Učenje koncepta *dan_za_sport* je učenje skupa dana za koji funkcija *dan_za_sport* = Da.

Taj je skup opisan sa konjunkcijom uvjeta nad atributnim vrijednostima

ZADATAK UČENJA KONCEPATA

- **DANO:**
- **Primjeri X:** mogući dani, svaki opisan skupom atributa:
 - Nebo (*sunčano, oblačno, kišno*)
 - Temperatura zraka (*vruće, hladno*)
 - Vlaga (*normalna, visoka*)
 - Vjetar (*jak, slab*)
 - Voda (*topla, hladna*)
 - Prognoza (*ista, promjena*)

ZADATAK UČENJA KONCEPATA

TRAŽI SE:

Hipoteza $h \in H$, takva da je $h(x) = c(x)$, za $\forall x \in X$.

- **Ako je dan skup primjera za učenje D , problem učenja koncepta je aproksimacija ciljne funkcije c .**

Hipoteza induktivnog učenja

- Uočite: iako je zadaća učenja određivanje hipoteze h identične ciljnoj funkciji c na cijelom X , **jedina dostupna informacija jest vrijednost funkcije c na skupu primjera za učenje D !**
- Nema potpune informacije - naša pretpostavka jest da najbolja hipoteza s obzirom na neviđene primjere jest **upravo ona koja najbolje opisuje primjere za učenje!**

- Hipoteza induktivnog učenja

HIPOTEZA INDUKTIVNOG UČENJA

Bilo koja hipoteza koja dobro aproksimira ciljnu funkciju na dovoljno velikom skupu primjera aproksimirat će ciljnu funkciju dovoljno dobro i nad novim primjerima.

UČENJE KONCEPATA KAO PRETRAŽIVANJE

- Izbor reprezentacije hipoteze određuje prostor svih hipoteza.
- *Primjer : Dan_za_sport*
 - *Nebo* (sunčano, oblačno, kišno)
 - *Temperatura zraka* (vruće i hladno)
 - *Vlaga* (normalna, visoka)
 - *Vjetar* (jak, slab)
 - *Voda* (topla, hladna)
 - *Prognoza* (ista, promjena)
- *Različitih primjera ima , elementi skupa X*
$$3*2*2*2*2*2 = 96$$

UČENJE KONCEPATA KAO PRETRAŽIVANJE

- *Prostor svih hipoteza H*
- *Sintaktički različitih hipoteza u H ima*
[redacted]
- *Semantički različitih hipoteza u H ima*
[redacted]

(Hipoteza koja sadrži jedan ili više \emptyset predstavlja prazan skup primjera)

- Primjer koncepta ***dan_za_sport*** je jednostavan, H je konačan. Obično, je H vrlo velik ili beskonačan.
- Zanimaju nas algoritmi koji mogu efikasno pretraživati vrlo veliki ili beskonačni skup hipoteza
→ oslanjanje na UREĐAJ!

UČENJE KONCEPATA KAO PRETRAŽIVANJE

Uređaj hipoteza od općenitog prema specifičnom

- Temelj algoritama za učenje koncepata → **parcijalni uređaj prostora hipoteza**
- Svaka hipoteza odgovara nekom podskupu primjera iz X koje pozitivno klasificira:
 $h1 = (\text{sunčano}, ?, ?, \text{jak}, ?, ?)$
 $h2 = (\text{sunčano}, ?, ?, ?, ?, ?)$
- Pogledajmo skup primjera klasificiran pozitivno kod jedne ili druge hipoteze!
 $h2$ je općenitija od $h1$

UČENJE KONCEPATA KAO PRETRAŽIVANJE

Definicija

Za neki $x \in \mathbf{X}$ i hipotezu $h \in \mathbf{H}$ kažemo da x **zadovoljava** h akko $h(x)=1$.

Definicija

Neka su h_j i h_k funkcije boolovih vrijednosti definirane na X . Tada je h_j **općenitija_od_ili_jednaka** h_k (piše se $h_j \geq_g h_k$) akko

$$(\forall x \in \mathbf{X}) [(h_k(x) = 1) \Rightarrow (h_j(x) = 1)].$$

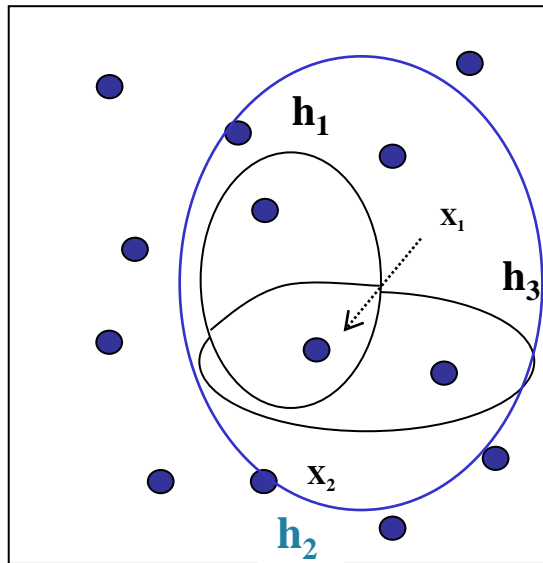
UČENJE KONCEPATA KAO PRETRAŽIVANJE

- Možemo razmatrati i relaciju ***strogo_općenitija_od***
$$(h_j >_g h_k) \Leftrightarrow (h_j \geq_g h_k) \wedge (h_k \neg \geq_g h_j)$$
- Inverzna relacija naziva se ***specifičnija_od***.
- \geq_g je **relacija parcijalnog uređaja** na **H**
(refleksivna, antisimetrična i tranzitivna).
- Neki je skup potpuno uređen ako nema neusporedivih elemenata (što ovdje nije slučaj!).

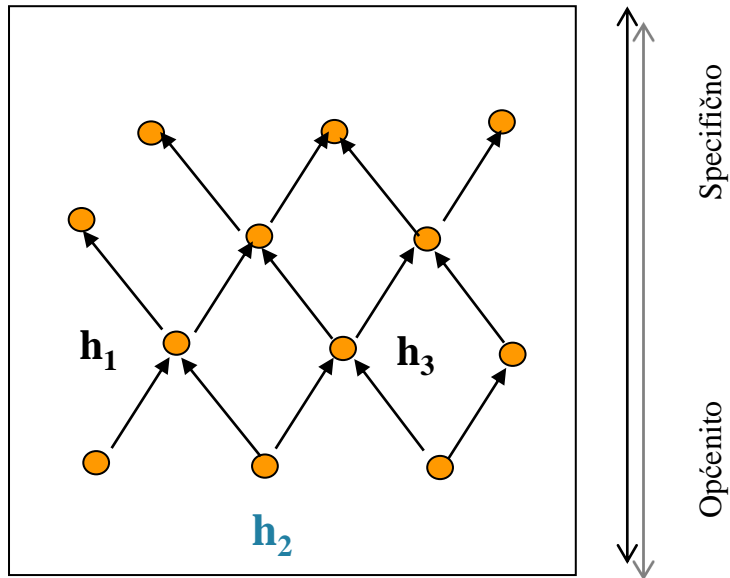
Zašto je \geq_g - vrlo važna relacija?

UČENJE KONCEPATA KAO PRETRAŽIVANJE

Primjeri **X**



Hipoteze **H**



$x_1 = (\text{sunčano}, \text{vruće}, \text{visoka}, \text{jak}, \text{hladna}, \text{ista})$

$x_2 = (\text{sunčano}, \text{vruće}, \text{visoka}, \text{slab}, \text{topla}, \text{ista})$

$h_1 = (\text{sunčano}, ?, ?, \text{jak}, ?, ?)$

$h_2 = (\text{sunčano}, ?, ?, ?, ?, ?)$

$h_3 = (\text{sunčano}, ?, ?, ?, \text{hladna}, ?)$

$$h_1(x_1) = 1, h_2(x_1) = 1, h_3(x_1) = 1$$

$$h_1(x_2) = 0, h_2(x_2) = 1, h_3(x_2) = 0$$

PRONAĐI-S ALGORITAM: NALAŽENJE MAKSIMALNO SPECIFIČNE HIPOTEZE

- Jedan način je krenuti od najspecifičnije hipoteze i generalizirati svaki put kada hipoteza ne pokrije pozitivni primjer za učenje

Pronađi-S algoritam

1. Inicijaliziraj ***h*** na najspecifičniju hipotezu u ***H***
2. Za svaki pozitivni primjer za učenje ***x***
Za svaki atribut ***a_i*** u ***h***
AKO je uvjet ***a_i*** zadovoljen sa ***x***
TADA ne čini ništa
INAČE zamijeni ***a_i*** u ***h*** sa slijedećim općenitijim uvjetom koji je zadovoljen sa ***x***
3. Predoči na izlazu ***h***

PRONAĐI-S ALGORITAM: NALAŽENJE MAKSIMALNO SPECIFIČNE HIPOTEZE

- Primjer

| <i>Primjer - dan</i> | <i>Naoblaka</i> | <i>Temper atura</i> | <i>Vlažnos t</i> | <i>Vjetar</i> | <i>Voda</i> | <i>Prognoza</i> | <i>Sport</i> |
|---------------------------------|------------------------|--------------------------------|-----------------------------|----------------------|--------------------|------------------------|---------------------|
| 1 | sunčano | vruće | normalna | jak | topla | isto | DA |
| 2 | sunčano | vruće | visoka | jak | topla | isto | DA |
| 3 | kišno | hladno | visoka | jak | topla | promjena | NE |
| 4 | sunčano | vruće | visoka | jak | hladna | promjena | DA |

$$h \leftarrow (\emptyset, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset)$$

Niti jedan uvjet hipoteze nije zadovoljen -> generalizacija

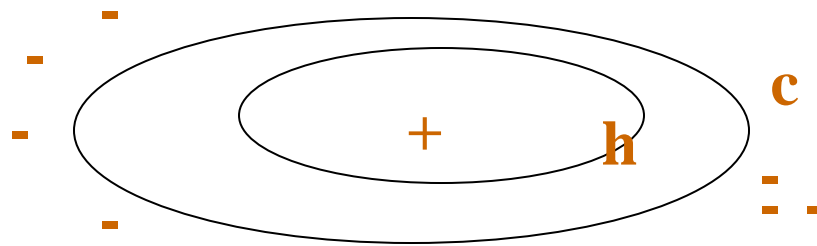
$$h \leftarrow (\text{sunčano}, \text{vruće}, \text{normalna}, \text{jak}, \text{topla}, \text{isto})$$

PRONAĐI-S ALGORITAM: NALAŽENJE MAKSIMALNO SPECIFIČNE HIPOTEZE

- Nakon 2. pozitivnog primjera:
 $h \leftarrow (\text{sunčano, vruće, ?, jak, topla, isto})$
- Treći primjer se ignorira jer je negativan
(h ga korektno klasificira kao -)
- *Komentar.* Ne treba revizija negativnih primjera
Zašto?

PRONAĐI-S ALGORITAM: NALAŽENJE MAKSIMALNO SPECIFIČNE HIPOTEZE

- Tekuća hipoteza je uvijek najspecifičnija koja pokriva + primjere.
- Pretpostavka: **c** se nalazi u H i nema pogrešaka u podacima.
- Ciljni koncept **c** je konzistentan je s + primjerima i vrijedi $c \geq_g h$,
no kako **c** ne pokriva negativne primjere, ne pokriva ih ni **h**.



- Četvrti primjer

$h \leftarrow (\text{sunčano, vruće, ?, jak, ?, ?})$

PRONAĐI-S ALGORITAM: NALAŽENJE MAKSIMALNO SPECIFIČNE HIPOTEZE

- Ukratko:
- **PRONAĐI-S** algoritam pretražuje po stablu parcijalnog uređaja od najspecifičnije prema općenitijim hipotezama.
- U svakom koraku hipoteza se generalizira samo onoliko koliko je nužno da pokrije pozitivne primjere.
- Dakle, **PRONAĐI-S** nalazi najspecifičniju hipotezu konzistentnu s pozitivnim primjerima

PRONAĐI-S ALGORITAM: NALAŽENJE MAKSIMALNO SPECIFIČNE HIPOTEZE

- Svojstvo **PRONAĐI-S**:
 - prostor hipoteza **H** opisan je kao konjunkcija uvjeta postavljenih na attribute.
 - **algoritam garantira nalaženje najspecifičnije hipoteze konzistentne sa pozitivnim primjerima za učenje.**
 - završna hipoteza konzistentna je i sa negativnim primjerima za učenje uz uvjet:
 - **ciljni koncept **c** je u **H****
 - **primjeri za učenje su ispravni**

PRONAĐI-S ALGORITAM: NALAŽENJE MAKSIMALNO SPECIFIČNE HIPOTEZE

Problemi u vezi PRONAĐI-S algoritmom

- Da li je učenik konvergirao prema pravom konceptu? Da li je pronađena hipoteza jedina hipoteza konzistentna s podacima (tj. ciljnim konceptom)?
- Zašto preferirati najspecifičniju hipotezu?
- Što ako primjeri za učenje sadrže šum ili pogrešku?
- Što ako ima više maksimalno specifičnih konzistentnih hipoteza? (To nije slučaj u navedenom primjeru “dan_za_sport”, no postoje slučajevi gdje ima više maksimalno specifičnih hipoteza kao i primjera prostora hipoteza gdje nema maksimalno specifične hipoteze. Algoritam treba biti u mogućnosti vraćati se unatrag i pretraživati različite druge grane prostora hipoteza s parcijalnim uređajem)

PRONAĐI-S ALGORITAM: NALAŽENJE MAKSIMALNO SPECIFIČNE HIPOTEZE

■ 1. Zadatak

- Razmotrite zadatak učenja koncepta u kojem je skup primjera jednak skupu realnih brojeva, tj. $X = \mathbb{R}$ i svaka hipoteza je u obliku $a < x < b$, gdje su a i b realne konstante, a x primjer, tj.

$$H = \{ (a, b) \mid a, b \in \mathbb{R} \}.$$

- Na primjer, hipoteza $3.1 < x < 6.2$ klasificira sve realne brojeve iz otvorenog intervala $(3.1, 6.2)$ kao pozitivne, a ostale kao negativne.
- a) Objasni zašto ne postoji maksimalno specifična hipoteza koja bi bila konzistentna sa bilo kojim skupom pozitivnih primjera?
- b) Sa kakvom modifikacijom predstavljanja hipoteze se to može postići?

PRONAĐI-S ALGORITAM: NALAŽENJE MAKSIMALNO SPECIFIČNE HIPOTEZE

- **2. Zadatak**
- Implementiraj PRONAĐI-S algoritam.
- Najprije provjeri da implementacija proizvodi trag identičan onom iz odjeljka 2.4.
- Zatim koristi program za slučajno generiranje primjera za učenje koji su potrebni da nauče točan ciljni koncept (iz istog primjera).
- Implementiraj generator primjera za učenje koji generira slučajne primjere, a zatim ih klasificira u skladu s ciljnim konceptom
(sunčano, vruće, ?, ?, ?, ?)
- Razmotri učenje svog PRONAĐI-S algoritma na slučajno generiranim primjerima i promatraj broj primjera potrebnih da bi program točno naučio ciljnu hipotezu.
- Možete li predvidjeti broj primjera za učenje?
- Napravi eksperiment barem 20 puta i navedi prosječan broj primjera za učenje. Kako taj broj ovisi o broju “?” u ciljnom konceptu? Kako ovisi o broju atributa koji se koriste za opis primjera i hipoteza?

ALGORITAM ELIMINACIJE-KANDIDATA

- PRONAĐI-S algoritam nalazi samo jednu od mogućih hipoteza koja je konzistentna s podacima.
- Algoritam ELIMINACIJE-KANDIDATA (engl. *CANDIDATE-ELIMINATION ALGORITHM*) daje opis skupa svih hipoteza konzistentnih s primjerima za učenje (i dalje važna relacija ***općenitija_od_ili_jednaka***)
- Oba spomenuta algoritma – isti nedostatak – loše ponašanje ako su podaci sa šumom.

ALGORITAM ELIMINACIJE-KANDIDATA

- Algoritam ELIMINACIJE-KANDIDATA nalazi sve opisive hipoteze konzistentne s primjerima za učenje

Definicija

Hipoteza **h** je konzistentna sa skupom primjera za učenje D akko $h(x)=c(x)$ za svaki primjer $(x,c(x))$ iz D .

Uočite bitnu razliku između definicije konzistentnosti i definicije zadovoljivosti!

ALGORITAM ELIMINACIJE-KANDIDATA

- Algoritam **ELIMINACIJE-KANDIDATA** → nalazi sve hipoteze konzistentne s primjerima za učenje - **PROSTOR INAČICA** (engl. *VERSION SPACE*)

Definicija

Prostor inačica, označen s $VS_{H,D}$, u odnosu na prostor hipoteza H i skup primjera za učenje D , jest podskup skupa hipoteza H , konzistentan s primjerima za učenje.

$$VS_{H,D} \equiv \{h \in H \mid \text{Konzistentna}(h, D)\}$$

LISTAJ-ZATIM-ELIMINIRAJ algoritam

Listaj-zatim-eliminiraj algoritam

1. *Prostor inačica* \leftarrow lista koja sadrži sve hipoteze iz H
2. Za svaki primjer za učenje $(x, c(x))$ iz D ukloni iz *Prostora inačica* sve hipoteze za koje je $h(x) \neq c(x)$
3. Predoči na izlazu hipoteze iz *Prostora inačica*

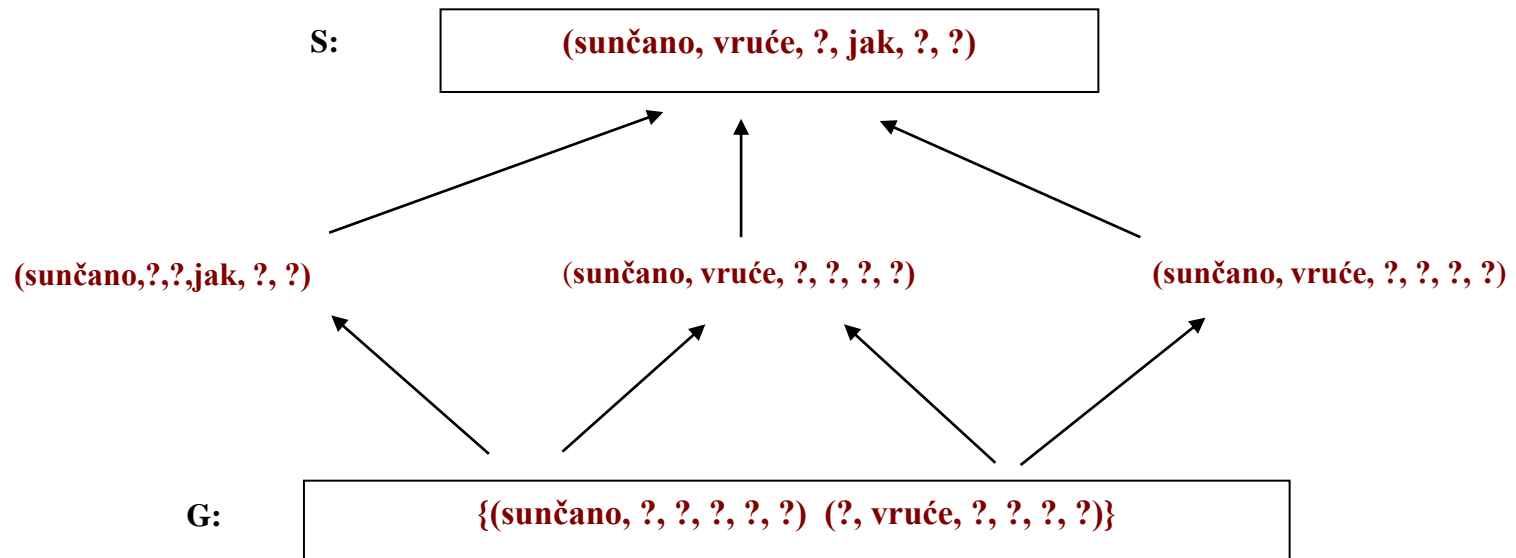
- Prostor inačica se reducira na
 - jednu hipotezu
 - cijeli skup hipoteza

} konzistentnih s podacima za učenje

(ako nema dovoljno podataka VS se ne može reducirati na jednu hipotezu)
- **nedostatak**: samo ako je H konačan,
- **prednost**: garantira nalaženje **SVIH** hipoteza konzistentnih s D

KOMPAKTNIJA REPREZENTACIJA *PROSTORA INAČICA*

- *Prostor inačica* predstavljen sa najopćenitijima (G) i najspecifičnijim hipotezama (S).
- Ta dva skupa hipoteza određuju granice *prostora inačica* unutar parcijalnog uređaja prostora hipoteza.
- VS za primjer učenja koncepta **dan_za_sport**



KOMPAKTNIJA REPREZENTACIJA *PROSTORA INAČICA*

- PRONAĐI-S algoritam daje samo hipotezu:
(sunčano, vruće, ?, jak, ?, ?),
koja je samo jedna od šest hipoteza koja je konzistentna sa primjerima za učenje.
- Strelice u VS označuju relaciju **općenitija_od**
- Algoritam ELIMINACIJA-KANDIDATA pohranjuje:
 - skup **S** (najspecifičnije hipoteze)
 - skup **G** (najopćenitije hipoteze).
 - iz ova dva skupa moguće je generirati sve ostale članove VS na temelju parcijalnog uređaja

KOMPAKTNIIJA REPREZENTACIJA *PROSTORA INAČICA*

Definicija

Najopćenitija granica G , u odnosu na prostor hipoteza H i skup primjera za učenje D , jest skup maksimalno općenitih članova skupa hipoteza H , konzistentnih s primjerima za učenje D .

Definicija

Najspecifičnija granica S , u odnosu na prostor hipoteza H i skup primjera za učenje D , jest skup maksimalno specifičnih (minimalno općenitih) članova skupa hipoteza H , konzistentnih s primjerima za učenje D .

KOMPAKTNIJA REPREZENTACIJA *PROSTORA INAČICA*

Teorem 2.1. - *Predstavljanje prostora inačica*

Neka je X proizvoljni skup primjera i neka je H skup hipoteza s boolovim vrijednostima definiran na X . Neka je $c: X \rightarrow \{0, 1\}$ proizvoljni ciljni koncept i neka je D proizvoljni skup primjera za učenje $\{(x, c(x))\}$.

Za sve X, H, c i D takve da su S i G dobro definirani,
$$VS_{H,D} \equiv \{h \in H \mid (\exists s \in S) (\exists g \in G) (g \geq_g h \geq_g s)\}$$

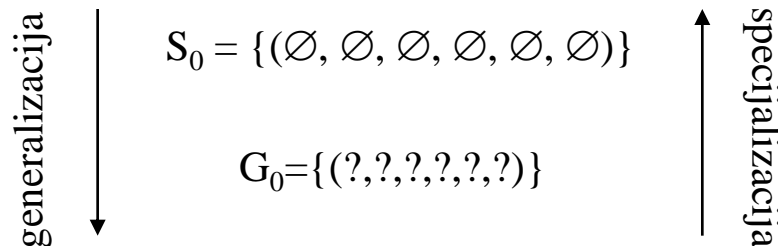
ALGORITAM ELIMINACIJE KANDIDATA

- *Prostor inačica* se inicijalizira na cijeli H – tj. inicijalno je određen sa

$$S_0 = \{(\emptyset, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset)\}$$

$$G_0 = \{(? , ? , ? , ? , ? , ?)\}$$

- Predočavanjem primjera za učenje, rubni se skupovi S i G specijaliziraju, odnosno generaliziraju tako da se eliminiraju sve hipoteze koje su nekonzistentne s novim primjerima za učenje. Na kraju tog postupka, prostor inačica sadrži samo one hipoteze **koje su konzistentne sa svim primjerima za učenje**



ALGORITAM ELIMINACIJE KANDIDATA

- Inicijaliziraj **G** na skup maksimalno općenitih hipoteza u **H**
- Inicijaliziraj **S** na skup maksimalno specifičnih hipoteza u **H**
- Za svaki primjer za učenje $\mathbf{d} \in D$ učini
- Ako je **d** pozitivni primjer
 - Ukloni iz **G** sve hipoteze nekonzistentne s **d**
 - Za svaku hipotezu $\mathbf{s} \in \mathbf{S}$ koja nije konzistentna s **d**
 - Ukloni $\mathbf{s} \in \mathbf{S}$
 - Dodaj u **S** sve minimalne generalizacije **h** od **s** takve da
 - **h** je konzistentna s **d**, i neki član iz **G** je općenitiji od **h** (osigurava konzistentnost s prethodnim negativnim primjerima)
 - Ukloni iz **S** sve hipoteze koje su općenitije od drugih hipoteza u **S**

ALGORITAM ELIMINACIJE KANDIDATA

- Ako je **d** negativni primjer
 - Ukloni iz **S** sve hipoteze nekonzistentne s **d**
 - Za svaku hipotezu **g** \in **G** koja nije konzistentna s **d**
 - Ukloni **g** \in **G**
 - Dodaj u **G** sve minimalne specijalizacije **h** od **g** takve da
 - **h** je konzistentna s **d**, i neki članovi iz **S** su specifičniji od **h** (osigurava konzistentnost s prethodnim **pozitivnim** primjerima)
 - Ukloni iz **G** sve hipoteze koje su specifičnije od drugih hipoteza u **G**
- Uoči dualnost kako pozitivni i negativni primjeri utječu na **S** i **G**!

ALGORITAM ELIMINACIJE KANDIDATA

- Operacije:
 - računanje minimalne generalizacije i specijalizacije dane hipoteze
 - identificiranje neminimalne i nemaksimalne hipoteze

Algoritam ELIMINACIJE_KANDIDATA može biti primijenjen na bilo koju zadaću učenja koncepta za koju su gornje operacije dobro definirane!

- *Ako tijekom učenja S ili G postanu prazni skupovi tada to znači da ciljni koncept nije moguće predstaviti u tom prostoru hipoteza.*

| Primjer - dan | Naoblaka | Temperat ura | Vlažnost | Vjetar | Voda | Prognoza | Dan za sport |
|------------------|----------|-----------------|----------|--------|--------|----------|-----------------|
| 1 | sunčano | vruće | normalna | jak | topla | isto | DA |
| 2 | sunčano | vruće | visoka | jak | topla | isto | DA |
| 3 | kišno | hladno | visoka | jak | topla | promjena | NE |
| 4 | sunčano | vruće | visoka | jak | hladna | promjena | DA |

- Prva dva primjera generaliziraju granicu **S0**, dok je **G0 = G1 = G2**

$S_0 \{(\emptyset, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset)\}$



$S_1 \{(\text{sunčano}, \text{vruće}, \text{normalna}, \text{jak}, \text{topla}, \text{isto})\}$



$S_2 (\text{sunčano}, \text{vruće}, ?, \text{jak}, \text{topla}, \text{isto})$



$G_0 \ G_1 \ G_2 : \{(? , ? , ? , ? , ? , ?)\}$

(Primijeti da prva dva koraka podsjećaju na PRONAĐI-S algoritam, pozitivni primjeri generaliziraju S)

Ako je **d** negativni primjer

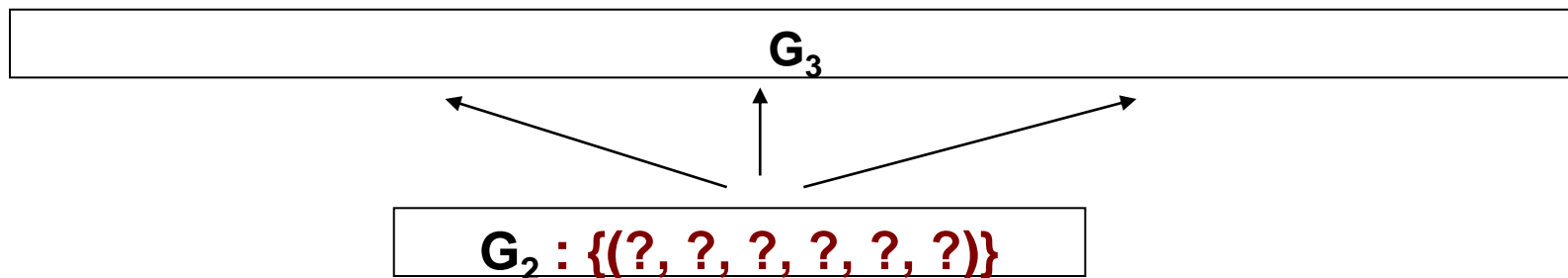
Ukloni iz **S** sve hipoteze nekonzistentne s **d**

Za svaku hipotezu **g** \in **G** koja nije konzistentna s **d**

Ukloni **g** \in **G**.....

| | | | | | | | |
|---|---------|--------|----------|-----|--------|----------|----|
| 1 | sunčano | vruće | normalna | jak | topla | isto | DA |
| 2 | sunčano | vruće | visoka | jak | topla | isto | DA |
| 3 | kišno | hladno | visoka | jak | topla | promjena | NE |
| 4 | sunčano | vruće | visoka | jak | hladna | promjena | DA |

S₂ = S₃ {(sunčano, vruće, ?, jak, topla, isto)}



Ako je **d** negativni primjer

Ukloni iz **S** sve hipoteze nekonzistentne s **d**

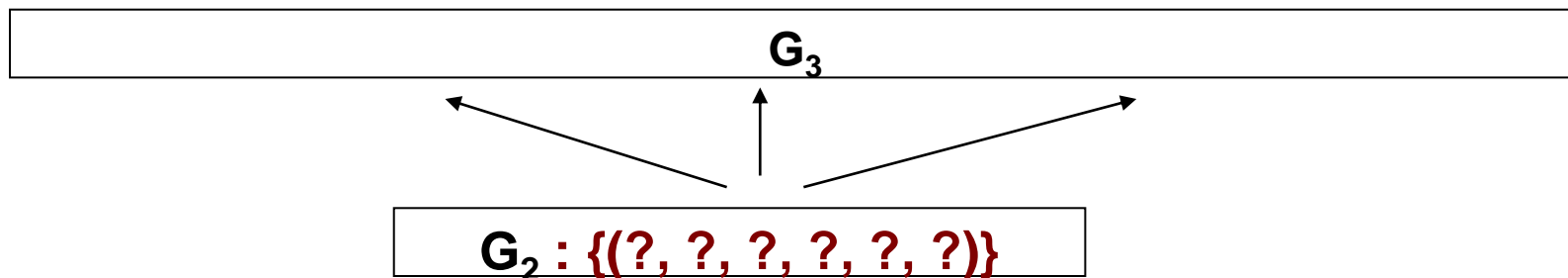
Za svaku hipotezu **g** \in **G** koja nije konzistentna s **d**

Ukloni **g** \in **G**

Dodaj u **G** sve minimalne specijalizacije **h** od **g** takve da **h** je konzistentna s **d**, i neki članovi iz **S** su specifičniji od **h** (osigurava konzistentnost s prethodnim **pozitivnim** primjerima)

| | | | | | | | |
|---|---------|--------|--------|-----|--------|----------|----|
| 3 | kišno | hladno | visoka | jak | topla | promjena | NE |
| 4 | sunčano | vruće | visoka | jak | hladna | promjena | DA |

S₂ S₃ {(sunčano, vruće, ?, jak, topla, isto)}



Prijelaz **G₂ \Rightarrow G₃** minimalne specijalizacije.

PRIMJER

| Primjer - dan | Naoblaka | Temperat ura | Vlažnost | Vjetar | Voda | Prognoza | Dan za sport |
|------------------|----------|-----------------|----------|--------|--------|----------|-----------------|
| 1 | sunčano | vruće | normalna | jak | topla | isto | DA |
| 2 | sunčano | vruće | visoka | jak | topla | isto | DA |
| 3 | kišno | hladno | visoka | jak | topla | promjena | NE |
| 4 | sunčano | vruće | visoka | jak | hladna | promjena | DA |

$S_2 S_3 \{(\text{sunčano, vruće, ?, jak, topla, isto})\}$

$G_3 \{(\text{sunčano, ?, ?, ?, ?, ?}) \dots \dots \dots \}$

$G_2 : \{(\text{?, ?, ?, ?, ?, ?})\}$

Uoči: negativni primjeri specijaliziraju G i to tako da se negativni primjer korektno klasificira kao negativan,



PRIMJER

| Primjer - dan | Naoblaka | Temperat ura | Vlažnost | Vjetar | Voda | Prognoza | Dan za sport |
|------------------|----------|-----------------|----------|--------|--------|----------|-----------------|
| 1 | sunčano | vruće | normalna | jak | topla | isto | DA |
| 2 | sunčano | vruće | visoka | jak | topla | isto | DA |
| 3 | kišno | hladno | visoka | jak | topla | promjena | NE |
| 4 | sunčano | vruće | visoka | jak | hladna | promjena | DA |

$S_2 S_3 \{(\text{sunčano, vruće, ?, jak, topla, isto})\}$

$G_3 \{(\text{sunčano, ?, ?, ?, ?, ?}) (\text{?, vruće, ?, ?, ?, ?}) \dots \dots \dots\}$

$G_2 : \{(\text{?, ?, ?, ?, ?, ?})\}$

Prijelaz $G_2 \Rightarrow G_3$ minimalne specijalizacije.



PRIMJER

| Primjer - dan | Naoblaka | Temperat ura | Vlažnost | Vjetar | Voda | Prognoza | Dan za sport |
|------------------|----------|-----------------|----------|--------|--------|----------|-----------------|
| 1 | sunčano | vruće | normalna | jak | topla | isto | DA |
| 2 | sunčano | vruće | visoka | jak | topla | isto | DA |
| 3 | kišno | hladno | visoka | jak | topla | promjena | NE |
| 4 | sunčano | vruće | visoka | jak | hladna | promjena | DA |

$S_2 S_3 \{(\text{sunčano, vruće, ?, jak, topla, isto})\}$

$G_3 \{(\text{sunčano, ?, ?, ?, ?, ?}) (\text{?, vruće, ?, ?, ?, ?}) , (\text{?, ?, normalna, ?, ?, ?}) ..$

$G_2 : \{(\text{?, ?, ?, ?, ?, ?})\}$

PRIMJER

| Primjer - dan | Naoblaka | Temperat ura | Vlažnost | Vjetar | Voda | Prognoza | Dan za sport |
|------------------|----------|-----------------|----------|--------|--------|----------|-----------------|
| 1 | sunčano | vruće | normalna | jak | topla | isto | DA |
| 2 | sunčano | vruće | visoka | jak | topla | isto | DA |
| 3 | kišno | hladno | visoka | jak | topla | promjena | NE |
| 4 | sunčano | vruće | visoka | jak | hladna | promjena | DA |

$S_2 S_3 \{(\text{sunčano, vruće, ?, jak, topla, isto})\}$

$G_3 \{(\text{sunčano, ?, ?, ?, ?, ?}) (\text{?, vruće, ?, ?, ?, ?}) , (\text{?, ?, normalna, ?, ?, ?}) ..$

$G_2 : \{(\text{?, ?, ?, ?, ?, ?})\}$

h je konzistentna s d , i neki članovi iz S su specifičniji od h



Hipoteza

$h = (?, ?, \text{normalna}, ?, ?, ?)$

nije općenitija od tekuće granice

$S2 \{(\text{sunčano}, \text{vruće}, ?, \text{jak}, \text{topla}, \text{isto})\}.$

- U algoritmu:

...

h je konzistentna s d , i neki članovi iz S su specifičniji od h

....

PRIMJER

| Primjer - dan | Naoblaka | Temperat ura | Vlažnost | Vjetar | Voda | Prognoza | Dan za sport |
|------------------|----------|-----------------|----------|--------|--------|----------|-----------------|
| 1 | sunčano | vruće | normalna | jak | topla | isto | DA |
| 2 | sunčano | vruće | visoka | jak | topla | isto | DA |
| 3 | kišno | hladno | visoka | jak | topla | promjena | NE |
| 4 | sunčano | vruće | visoka | jak | hladna | promjena | DA |

$S_2 S_3 \{(\text{sunčano, vruće, ?, jak, topla, isto})\}$

$G_3 \{(\text{sunčano, ?, ?, ?, ?, ?}) (\text{?, vruće, ?, ?, ?, ?}) , (\text{?, ?, visoka, ?, ?, ?}) \dots\}$

$G_2 : \{(\text{?, ?, ?, ?, ?, ?})\}$

PRIMJER

| Primjer - dan | Naoblaka | Temperat ura | Vlažnost | Vjetar | Voda | Prognoza | Dan za sport |
|------------------|----------|-----------------|----------|--------|--------|----------|-----------------|
| 1 | sunčano | vruće | normalna | jak | topla | isto | DA |
| 2 | sunčano | vruće | visoka | jak | topla | isto | DA |
| 3 | kišno | hladno | visoka | jak | topla | promjena | NE |
| 4 | sunčano | vruće | visoka | jak | hladna | promjena | DA |

$S_2 S_3 \{(\text{sunčano, vruće, ?, jak, topla, isto})\}$

$G_3 \{(\text{sunčano, ?, ?, ?, ?, ?}) (\text{?, vruće, ?, ?, ?, ?}) , (\text{?, ?, visoka, ?, ?, ?}) \dots\}$

$G_2 : \{(\text{?, ?, ?, ?, ?, ?})\}$

h je konzistentna s d , i neki članovi iz S su specifičniji od h



PRIMJER

| Primjer - dan | Naoblaka | Temperat ura | Vlažnost | Vjetar | Voda | Prognoza | Dan za sport |
|------------------|----------|-----------------|----------|--------|--------|----------|-----------------|
| 1 | sunčano | vruće | normalna | jak | topla | isto | DA |
| 2 | sunčano | vruće | visoka | jak | topla | isto | DA |
| 3 | kišno | hladno | visoka | jak | topla | promjena | NE |
| 4 | sunčano | vruće | visoka | jak | hladna | promjena | DA |

$S_2 S_3 \{(\text{sunčano, vruće, ?, jak, topla, isto})\}$

$G_3 \{(\text{sunčano, ?, ?, ?, ?, ?}) (\text{?, vruće, ?, ?, ?, ?}) (\text{?, ?, ?, jak, ?, ?}) \dots\}$

$G_2 : \{(\text{?, ?, ?, ?, ?, ?})\}$

Izbacujemo (?, ?, ?, jak, ?, ?) jer...?.

1. Zašto $G3$ ne sadrži $\{(? , ? , ? , \text{jak} , ? , ?)\}$?
2. Zašto $G3$ ne sadrži $\{(? , ? , ? , ? , \text{topla}, ?)\}$?

1. i 2. zbog nekonzistentcije – ovakve hipoteze klasificirale bi tekući primjer kao pozitivan, a ne kao negativan

.....

Ako je d pozitivni primjer

Ukloni iz G sve hipoteze nekonzistentne s d

Za svaku hipotezu $g \in G$ koja nije konzistentna s d

Ukloni $g \in G$

Dodaj u G sve minimalne specijalizacije h od g takve da h je konzistentna s d , i neki članovi iz S su specifičniji od h (osigurava konzistentnost s prethodnim pozitivnim primjerima)....

PRIMJER

| Primjer - dan | Naoblaka | Temperat ura | Vlažnost | Vjetar | Voda | Prognoza | Dan za sport |
|------------------|----------|-----------------|----------|--------|--------|----------|-----------------|
| 1 | sunčano | vruće | normalna | jak | topla | isto | DA |
| 2 | sunčano | vruće | visoka | jak | topla | isto | DA |
| 3 | kišno | hladno | visoka | jak | topla | promjena | NE |
| 4 | sunčano | vruće | visoka | jak | hladna | promjena | DA |

$S_2 S_3 \{(\text{sunčano, vruće, ?, jak, topla, isto})\}$

$G_3 \{(\text{sunčano, ?, ?, ?, ?, ?}) (\text{?, vruće, ?, ?, ?, ?}) (\text{?, ?, ?, ?, ?, isto})\}$

$G_2 : \{(\text{?, ?, ?, ?, ?, ?})\}$

Prijelaz $G_2 \Rightarrow G_3$ minimalne specijalizacije.

- Interpretacija **S** i **G**:
- **S** granica *prostora_inačica* sumira sve prethodne **pozitivne** primjere koji se koriste za određivanje da li je neka dana hipoteza konzistentna s tim primjerima. Bilo koja hipoteza općenitija od **S** će tada pokrivati pozitivne primjere koje pokriva i **S**
- Na dualan način **G** sumira informacije svih prethodnih **negativnih** primjera. Bilo koja hipoteza specifičnija od **G** bit će konzistentna s prethodnim negativnim primjerima

| Primjer - dan | Naoblaka | Temperat ura | Vlažnost | Vjetar | Voda | Prognoza | Dan za sport |
|------------------|----------|-----------------|----------|--------|--------|----------|-----------------|
| 1 | sunčano | vruće | normalna | jak | topla | isto | DA |
| 2 | sunčano | vruće | visoka | jak | topla | isto | DA |
| 3 | kišno | hladno | visoka | jak | topla | promjena | NE |
| 4 | sunčano | vruće | visoka | jak | hladna | promjena | DA |

- Četvrti primjer:
- Jedan član **G3** (?, ?, ?, ?, ?, **isto**) mora biti uklonjen zato što nije konzistentan s četvrtim primjerom
 - Ako je **d** pozitivni primjer
 - Ukloni iz **G** sve hipoteze nekonzistentne s **d**
- (uoči da je uklanjanje neizbježno jer specijalizacijom hipoteze iz G ne postizemo konzistentnost)
- Nadalje, četvrti primjer generalizira donju granicu sa **S3** na **S4**

Četvrti primjer:

$S_4 \{(\text{sunčano, vruće, visoka, jak, hladna, promjena}), DA\}$.

$S_3 \{(\text{sunčano, vruće, ?, jak, topla, isto})\}$



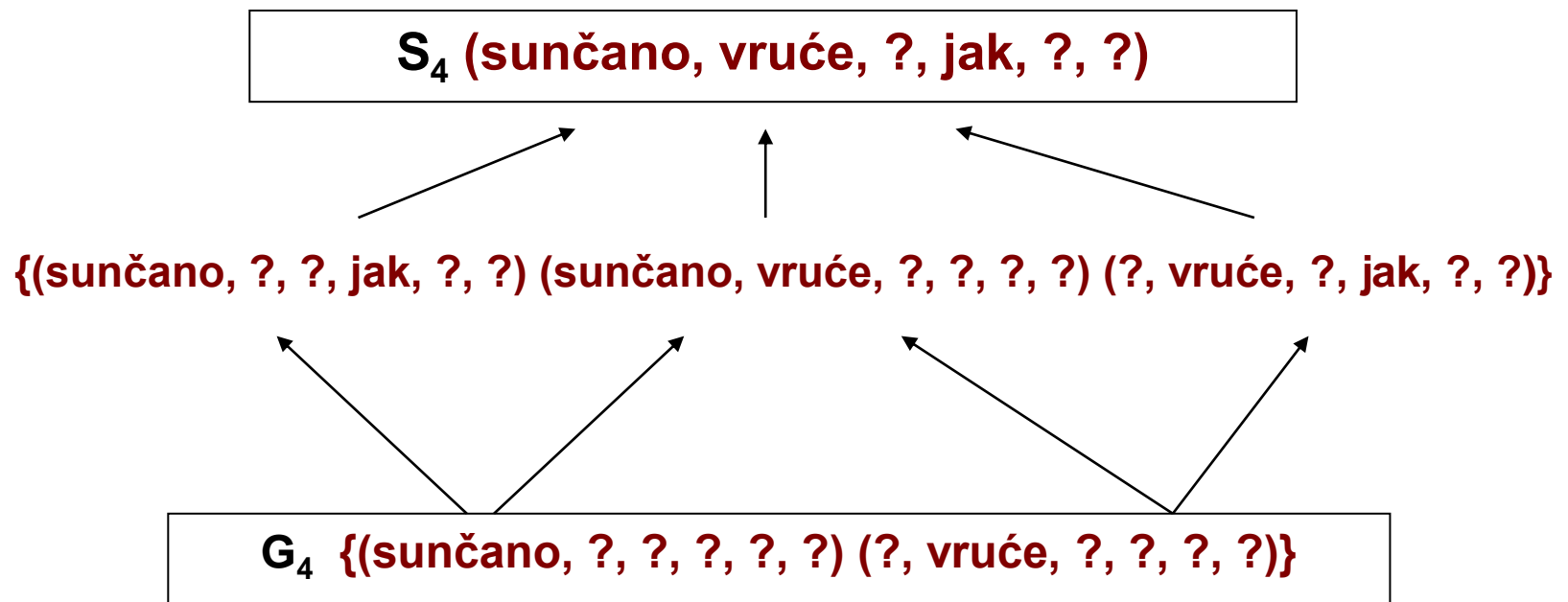
$S_4 \{(\text{sunčano, vruće, ?, jak, ?, ?})\}$

$G_4 \{(\text{sunčano, ?, ?, ?, ?, ?}) (?, vruće, ?, ?, ?, ?)\}$



$G_3 \{(\text{sunčano, ?, ?, ?, ?, ?}) (?, vruće, ?, ?, ?, ?) (?, ?, ?, ?, ?, isto)\}$

- Konačni oblik prostora_inačica:



- Da li naučeni *prostor inačica* zavisi od redoslijeda predočavanja primjera?

PRIMJEDBE O *PROSTORU INAČICA* I ALGORITMU ELIMINACIJA_KANDIDATA

Da li algoritam ELIMINACIJA_KANDIDATA konvergira prema korektnoj hipotezi?

- Prostor inačica uz E_K algoritam konvergira prema pravoj hipotezi ako
 - Nema pogrešaka u primjerima za učenje
 - Postoji neka hipoteza u H koja korektno opisuje ciljni koncept

Ciljni koncept je točno naučen kada S i G konvergiraju prema jednoj identičnoj hipotezi!

PRIMJEDBE O *PROSTORU INAČICA* I ALGORITMU ELIMINACIJA_KANDIDATA

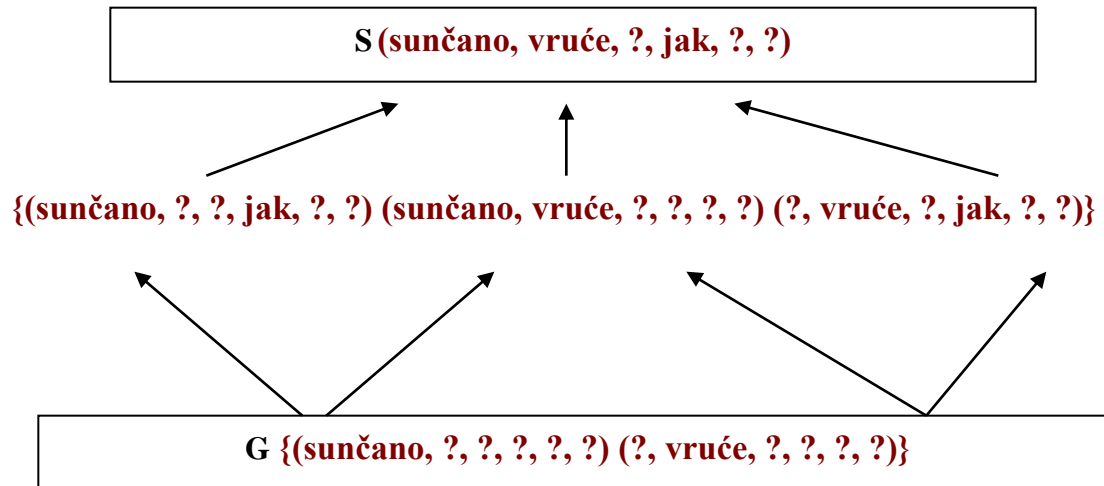
- Što se dešava kada se pozitivan primjer pogrešno deklarira kao negativan?
- Ispravni ciljni koncept se uklanja iz prostora inačica. (Uklanjaju se sve hipoteze nekonzistentne s primjerima za učenje).

Ako ima dovoljno primjera za učenje eventualno će S i G konvergirati prema praznom skupu – NEMA HIPOTEZE KONZISTENTNE S PRIMJERIMA

- Slično ako se ciljni koncept ne može opisati u odabranoj reprezentaciji hipoteze. (konjunkcija -> disjunkcija atributa)

KOJI SLJEDEĆI PRIMJER ZA UČENJE UČENIK TREBA ZATRAŽITI?

- Do sada – primjere za učenje daje učitelj
- Što ako učenik može voditi eksperiment i sam izabrati primjer za učenje i zatim dobiti informaciju iz vanjskog svijeta (učitelja, prirode) o korektnoj klasifikaciji izabranog primjera? → u tom slučaju koristimo termin **upit** (engl. *query*)



- Što bi bilo dobro pitanje u ovom slučaju?
- Što je dobra strategija upita uopće?