

Begrepp och Hypoteser 00000000 Inläring genom sökning 0000000000000000 Objektiv inläring 00

Begräppsinläring

Begrepp och Hypoteser 00000000 Inläring genom sökning 0000000000000000 Objektiv inläring 00

- 1 Begrepp och Hypoteser
 - Definitioner
 - Exempel
 - Hypoteser
- 2 Inläring genom sökning
 - Find-S
 - List-then-Eliminate
 - Candidate Elimination
- 3 Objektiv inläring
 - Bias
 - Unbiased Learner

Begrepp och Hypoteser 00000000 Inläring genom sökning 0000000000000000 Objektiv inläring 00

- 1 Begrepp och Hypoteser
 - Definitioner
 - Exempel
 - Hypoteser
- 2 Inläring genom sökning
 - Find-S
 - List-then-Eliminate
 - Candidate Elimination
- 3 Objektiv inläring
 - Bias
 - Unbiased Learner

Begrepp och Hypoteser 00000000 Inläring genom sökning 0000000000000000 Objektiv inläring 00

Begräppsinläring

Begräppsinläring (*Concept Learning*)

Inläring av en **boolsk funktion** från exempel

Kategorier

- "Vackert väder"
- "Hundar"
- "Motorfordon"
- "Lagbrott"

Delmängd av någon överordnad mängd X

Begrepp och Hypoteser 00000000 Inläring genom sökning 0000000000000000 Objektiv inläring 00

Definitioner

Beteckningar

c Begreppet som ska läras

$$c(x) \rightarrow 0/1, \quad x \in X$$

h Hypotes, Resultatet av inläringen ("gissad c ")

$$h(x) \rightarrow 0/1, \quad x \in X$$

H Hypotesrummet, Alla tänkbara hypoteser

$$h \in H$$

D Mängden tillgängliga tränings exempel

$$D \subseteq X$$

Begrepp och Hypoteser 00000000 Inläring genom sökning 0000000000000000 Objektiv inläring 00

Definitioner

Beteckningar

Två sorters tränings exempel

Positiva exempel:

$$x : c(x) = 1, \quad x \in D$$

Negativa exempel:

$$x : c(x) = 0, \quad x \in D$$

Exempel på ett *begrepp*
 "Vackert Väder"

Låt "väderinstanserna" x_i vara sammansatta av fyra **attribut**:

- $x_1 = \langle \text{Soligt, Varmt, Blåsig, Torrt} \rangle$
- $x_2 = \langle \text{Molnigt, Varmt, Stilla, Torrt} \rangle$
- $x_3 = \dots$

Generellt: $\text{Himmel} \times \text{Temp} \times \text{Vind} \times \text{Fukt}$

Antag att attributen bara kan anta speciella diskreta värden:

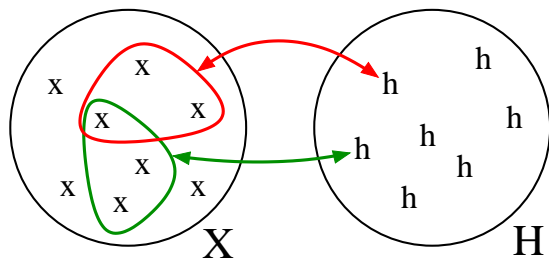
- Himmel $\in \{ \text{Soligt, Molnigt, Regnigt} \}$
- Temp $\in \{ \text{Varmt, Kallt} \}$
- Vind $\in \{ \text{Blåsig, Stilla} \}$
- Fukt $\in \{ \text{Torrt, Fuktigt} \}$

Antal möjliga väder: $|X| = 3 \cdot 2 \cdot 2 \cdot 2 = 24$

Typiska träningsexempel

- $x_1 = \langle \text{Soligt, Varmt, Blåsig, Torrt} \rangle \rightarrow \text{Vackert}$
- $x_2 = \langle \text{Soligt, Varmt, Blåsig, Fuktigt} \rangle \rightarrow \text{Vackert}$
- $x_3 = \langle \text{Regnigt, Kallt, Blåsig, Fuktigt} \rangle \rightarrow \text{Usch}$
- $x_4 = \langle \text{Soligt, Varmt, Stilla, Fuktigt} \rangle \rightarrow \text{Vackert}$

Hur ser hypotesmängden H ut?



Varje hypotes h motsvaras av en **delmängd** av X

Hur många hypoteser finns det att välja bland?
 Hur många delmängder av X finns det?

$$|H| = 2^{|X|}$$

$$|H| = 2^{24} = 16777216$$

Man måste göra begränsande antaganden!

Begränsande antagande
 Begreppen är alltid konjunktioner av attributvärden

Exempel på begrepp c av denna typ

- Soligt & Varmt
- Kallt & Vindstilla & Torrt

Hur många hypoteser finns det nu?

Himmel	Temp	Vind	Fukt
Soligt			
Molnigt	Varmt	Blåsig	Torrt
Regnigt	Kallt	Stilla	Fuktigt
*	*	*	*

$$4 \cdot 3 \cdot 3 \cdot 3 = 108$$

Begrepp och Hypoteser ○○○○○○○○ Inlärnin g genom sökning ○○○○○○○○○○ Objektiv inlärnin g ○○

- 1 Begrepp och Hypoteser
 - Definitioner
 - Exempel
 - Hypoteser
- 2 Inlärnin g genom sökning
 - Find-S
 - List-then-Eliminate
 - Candidate Elimination
- 3 Objektiv inlärnin g
 - Bias
 - Unbiased Learner

Begrepp och Hypoteser ○○○○○○○○ Inlärnin g genom sökning ●○○○○○○○○○ Objektiv inlärnin g ○○

Find-S

Inlärnin g \equiv sök efter en hypotes som stämmer med exemplen

Utnyttja strukturen hos H för att söka snabbare

Begrepp och Hypoteser ○○○○○○○○ Inlärnin g genom sökning ○○○○○○○○○○ Objektiv inlärnin g ○○

Find-S

Vissa hypoteser är **generellare** än andra

Partiell ordning mellan hypoteserna

Begrepp och Hypoteser ○○○○○○○○ Inlärnin g genom sökning ○○○○○○○○○○ Objektiv inlärnin g ○○

Find-S

Generella hypoteser

Speciella hypoteser

Generellast i vårt exempel: "Alla väder är vackra"

Speciellast i vårt exempel: "Inga väder är vackra" (!)

Begrepp och Hypoteser ○○○○○○○○ Inlärnin g genom sökning ○○○○○○○○○○ Objektiv inlärnin g ○○

Find-S

Begrepp och Hypoteser ○○○○○○○○ Inlärnin g genom sökning ○○○○○○○○○○ Objektiv inlärnin g ○○

Find-S

Find-S algoritmen

Starta med den/de mest speciella hypoteserna och generalisera succesivt vid behov.

$\hat{h} \leftarrow$ mest speciella hypotesen i H

for $e \leftarrow$ nästa exempel:

if positivt exempel:

generalisera \hat{h} så att även e täcks in

Ger den mest speciella hypotesen som är **konsistent** med alla exemplen.

Konkret exempel: "Vackert väder" med antagandet att begreppet är en konjunktion av attributvärden.

Starthypotes: Aktuell hypotes: $\langle \emptyset, \emptyset, \emptyset, \emptyset \rangle$ (Maximalt pessimistisk) $\langle \text{Soligt, Varmt, Blåsig, Torrt} \rangle$ $\langle \text{Soligt, Varmt, Blåsig, } * \rangle$ $\langle \text{Soligt, Varmt, } *, * \rangle$

Träningsexempel:

- $x_1 = \langle \text{Soligt, Varmt, Blåsig, Torrt} \rangle \rightarrow \text{Vackert}$
- $x_2 = \langle \text{Soligt, Varmt, Blåsig, Fuktigt} \rangle \rightarrow \text{Vackert}$
- $x_3 = \langle \text{Regnigt, Kallt, Blåsig, Fuktigt} \rangle \rightarrow \text{Usch}$
- $x_4 = \langle \text{Soligt, Varmt, Stilla, Fuktigt} \rangle \rightarrow \text{Vackert}$

Slutlig hypotes: "Vackert väder" $\equiv \text{Soligt} \wedge \text{Varmt}$

Problem med Find-S

- Omöjligt att veta ifall endast en unik hypotes återstår.
- Varför skulle vi föredra den mest specifika hypotesen?
- Vi kan inte upptäcka inkonsistenta indata eftersom alla negativa exempel ignoreras.
- Vad händer ifall det finns fler lika specifika hypoteser?

Version Space (VS)

Alla hypoteser som är konsistenta med de exempel man hittills har sett.

- $VS \subseteq H$
- $|VS| = 1$ En unik lösning
- $VS = \emptyset$ Inkonsistenta exempel

List-then-Eliminate algoritmen

Direkt representation av Version Space (VS)

$VS \leftarrow H$

for $e \leftarrow$ nästa exempel:

tag bort alla hypoteser ur VS som inte är konsistenta med e

Problem: H är oftast för stor!

Candidate Elimination

- Effektiv representation av Version Space
- Utnyttjar den partiella ordningen mellan hypoteserna

