# Lärbarhetsteori











Frågor för teoretisk analys av inlärning

- Hur svårt är ett givet inlärningsproblem?
- Hur många träningsexempel krävs?
- Hur många felklassningar kan vi förvänta oss under och efter inlärningen?
- Hur stor är risken att inlärningen misslyckas?



Förutsättningar:

- Begreppsinlärning
- Träning och test från samma fördelning $\mathcal{D}$

Vilken sorts fel kan uppstå?

- Resultatet av inlärningen kan vara dåligt
  Hypotesen ger fel klassning ibland
- Inlärningen själv kan gå dåligt
  Hittar ibland ingen vettig hypotes



Begränsning av hur dåliga hypoteser får bli

True Error — sannolikheten att en viss hypotes gör fel

$$\text{error}_{\mathcal{D}}(h) \equiv P_{x \in \mathcal{D}}\left[h(x) \neq c(x)\right]$$

En hypotes $h$ sägs vara ganska rätt (*approximately correct*) när

$$\text{error}_{\mathcal{D}}(h) < \epsilon$$

Begränsning av hur ofta inlärningen får misslyckas

Risken att inlärningen inte hittar en tillräckligt bra hypotes

$$P_L\left[\text{error}_{\mathcal{D}}(h) \geq \epsilon\right]$$

Algoritmen $L$ sägs troligen (*probably*) hitta en lösning när

$$P_L\left[\text{error}_{\mathcal{D}}(h) \geq \epsilon\right] < \delta$$







## PAC-learning

Probably Approximately Correct

Givet

| | |
|---|---|
| $C$ | Begreppet som ska läras |
| $\epsilon$ | krav på felet |
| $\delta$ | krav på risken |
| $n$ | exemplens storlek |

PAC-learnable: Tiden att hitta en lösning växer polynomiellt i size$(C)$, $n$, $\frac{1}{\epsilon}$ och $\frac{1}{\delta}$



Analys av en Consistent learner

- Förutsättning: inga fel i träningsexemplen
- Exemplen dras från fördelningen $\mathcal{D}$
- Svaret är alltid konsistent med träningsexemplen
- ”Farliga hypoteser”:

$$\text{error}_{\mathcal{D}}(h) \geq \epsilon$$

Vi vill inte att inlärningen
ger en farlig hypotes som svar!

Hur stor är risken att en farlig hypotes är konsistent med alla träningsexempel?



- Sannolikheten att en hypotes $h$ motsägs av ett exempel

$$\text{error}_{\mathcal{D}}(h)$$

- Sannolikheten att $h$ inte motsägs

$$1 - \text{error}_{\mathcal{D}}(h)$$

- Risken att en *farlig hypotes* ($\text{error}_{\mathcal{D}}(h) \geq \epsilon$) inte motsägs av ett slumpmässigt valt träningsexempel

$$\leq (1 - \epsilon)$$



- Risken att en *farlig hypotes* inte motsägs av ett slumpmässigt valt träningsexempel

$$\leq (1 - \epsilon)$$

- Risken att en *farlig hypotes* inte motsägs av $m$ slumpmässigt dragna träningsexempel

$$\leq (1 - \epsilon)^m$$

- Hur stor är risken att någon farlig hypotes i $H$ råkar stämma med alla exempel:

$$\leq |H| \cdot (1 - \epsilon)^m$$

$$\leq |H| \cdot e^{-\epsilon m}$$

Hur många träningsexempel behövs?

Hur många exempel $m$ krävs för att risken att få en farlig hypotes säkert ska vara mindre än $\delta$

$$\delta \geq |H| \cdot e^{-\epsilon m}$$

$$e^{\epsilon m} \geq \frac{|H|}{\delta}$$

$$m \geq \frac{1}{\epsilon}\left[\ln|H| + \ln\frac{1}{\delta}\right]$$



Viktigt samband:

$$m \geq \frac{1}{\epsilon}\left[\ln|H| + \ln\frac{1}{\delta}\right]$$

Är detta PAC-learnable?

Potentiellt problem: $|H|$ kan vara för stor



## Inlärning av konjunktioner

Exempel: Soligt $\wedge$ $\neg$Blåsigt $\wedge$ Blött
$n$ attribut $\Rightarrow 3^n$ möjliga begrepp $\Rightarrow |H| = 3^n$

$$m \geq \frac{1}{\epsilon}\left[n \ln 3 + \ln\frac{1}{\delta}\right]$$

- Linjärt i $\frac{1}{\epsilon}$
- Linjärt i $n$
- Logaritmiskt i $\frac{1}{\delta}$

Verkar vara PAC-learnable!
Dessutom krävs att *tiden för varje exempel* är polynomisk.
Find-S: Ok



## Inlärning utan Bias

Alla tänkbara delmängder av $X$ är hypoteser

$$|X| = 2^n$$

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

$$m \geq \frac{1}{\epsilon}\left[2^n \ln 2 + \ln\frac{1}{\delta}\right]$$

Ej PAC-learnable!
Skattningen är dock en övre gräns (överskattning)
Vi har inte visat att $m$ verkligen växer exponentiellt med $n$
I detta fall *är* det sant







Problem med $|H|$

- Alltför pessimistiska skattningar
- Oanvändbar när $|H| = \infty$

**Vapnik — Chervonenkis observation:**

Det intressanta är *inte antalet* hypoteser,
utan hur de kan forma olika delmängder i $X$

**Splittring**

En ändlig mängd $S$ splittras av hypoteserna $H$ om varje delmängd ur $S$ beskrivs av något $h \in H$

Storleken på $S$ ger ett mått på hur uttrycksfull $H$ är

**VC-dimension**

VC($H$)
Storleken på den största delmängden av $X$ som splittras av $H$



Exempel:

$H$ Intervall på reela talaxeln
$X$ Reella tal

- Kan 2 punkter splittras?
- Kan 3 punkter splittras?

Slutsats: $\mathrm{VC}(H) = 2$



Exempel:

$H$ Separerande hyperplan
$X$ Punkter i $\Re^r$

- När $r = 1$
$$\mathrm{VC}(H) = 2$$
- När $r = 2$
$$\mathrm{VC}(H) = 3$$
- Generellt
$$\mathrm{VC}(H) = r + 1$$



## Antal träningsexempel

Tidigare skattning

$$m \geq \frac{1}{\epsilon}\left[\ln |H| + \ln \frac{1}{\delta}\right]$$

Ny skattning

$$m \geq \frac{1}{\epsilon}\left[4 \log_2 \frac{2}{\delta} + 8\mathrm{VC}(H) \cdot \log_2 \frac{13}{\epsilon}\right]$$

Mycket bättre (mindre)







**Alternativt prestandamått för inlärningsalgoritmer:**

Hur många fel görs under inlärningen

## Find-S

- Lär sig bara vid felklassning
- Generaliserar i värsta fall bara ett attribut varje gång
- Första gången väljs en speciellt hypotes
- Maximalt $n + 1$ ändringar

Man gör maximalt $n + 1$ fel



## Candidate Elimination

- Vi måste tvinga ur algoritmen ett förslag
- Antag att vi använder majoritetsröstning bland hypoteserna som är kvar i *Version Space*
- Fel bara när minst hälften av VS har fel
- Varje gång vi gör fel försvinner minst hälften av VS

Maximalt $\log_2 |H|$ fel



**Optimal inlärning**

- Bästa algoritmen
- Värsta fallet
  Svåraste begreppet, värsta sekvensen av exemplen

Antal fel för att lära begreppet $C$:

$$\mathrm{Opt}(C)$$

Teoretiska gränser

$$\mathrm{VC}(C) \leq \mathrm{Opt}(C) \leq \log_2 |C|$$