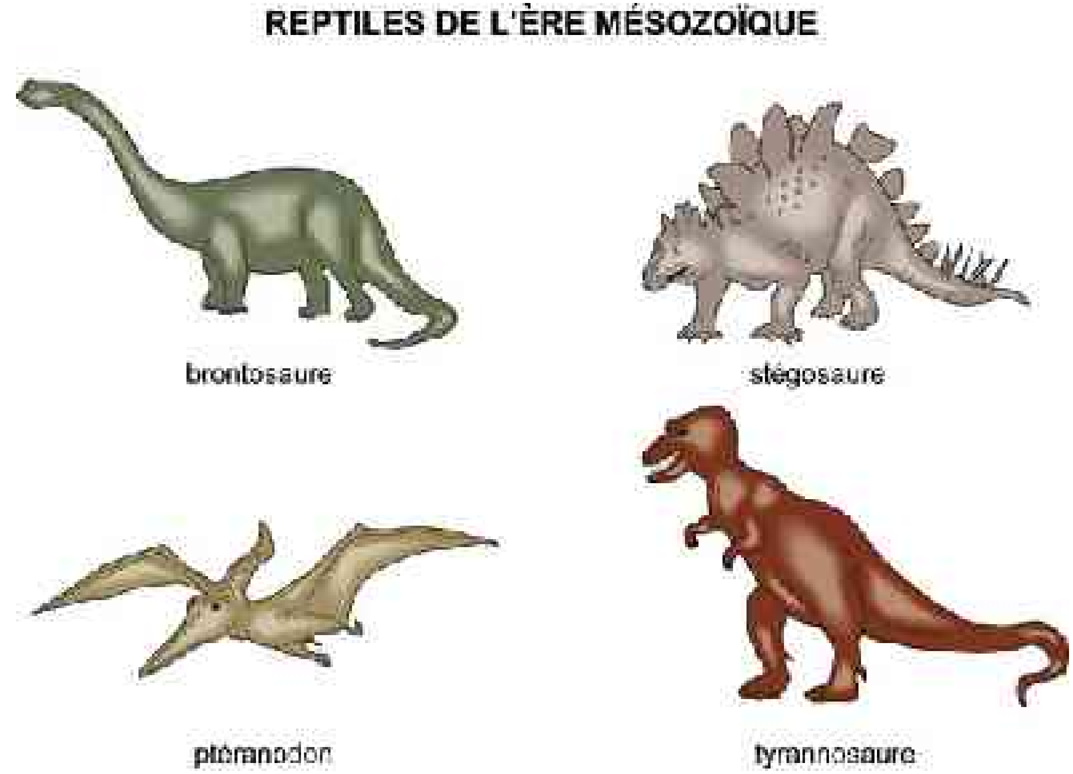
Inductive-deductive systems : learning mathematical theories

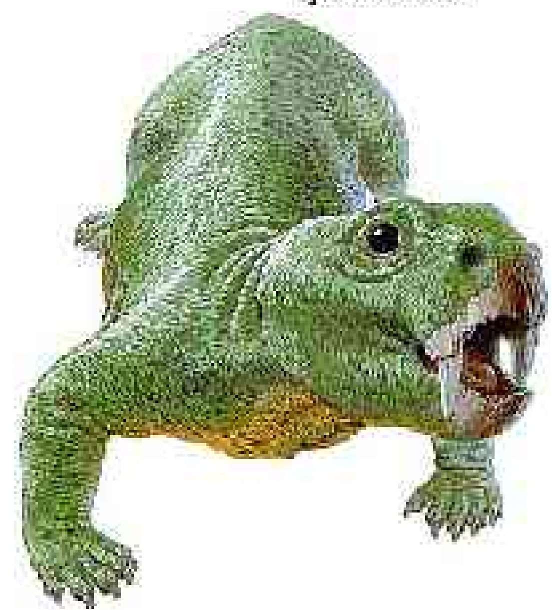
- Summary of statistical learning theory
- Inductive-deductive systems
- The VC-dimension of mathematical theories
- Consequences

Summary of statistical learning

What is learning ?



?



What is learning theory ?

Examples :

Wings,	small,	warm-blooded,	*bird*
Wings,	small,	warm-blooded,	*bird*
┐ wings,	big,	cold-blooded,	*dinosaur*
┐ wings,	small,	cold-blooded,	*dinosaur*

Goal :

mapping f : {W, ┐ W} x[0,1] x{c,w} → {b,d}

Prior knowledge :

f Є F

Goal : *generalization*

==> No wings, medium, cold-blooded, dinosaur

What is statistical learning theory ?

if the sample is not too bad,

P (error) decreases as

number of examples ↑

prior knowledge ↑

with probability $1-\delta$

$$P(error) = O(\quad (V-\log(\delta))/ n \quad)$$

- n = number of examples
- V big si family of functions « big »

Inductive-deductive systems

Inductive / deductive systems

Deductive systems :

{ A, A ==> B } hold

then B holds

(==> expert systems)

Inductive systems :

{ A(2), A(4), ┐A(7), ┐A(5), A(6) } hold

then {A(2n), ┐A(2n+1)} hold

(==> inductive logic programming)

Inductive-deductive systems :

{ A(2), A(4), ┐A(7), ┐A(5), A(6), A(n)↔B(n) } hold

then B(2n) and ┐B(2n+1)

Our framework

Hypothesis :
e1,e2,...,en « examples » of a theory
(i.i.d., law P)

Algorithm :
e1, e2, ... en → theory T

efficiency :
Ln = P (e(n+1) Є T)
(Ln= random variable in [0,1])

VC-dimension

F a family of functions with values in {0,1}

shattering coefficients :

S(F,n) = sup | { (f(x1),f(x2),...,f(xn)) ; f Є F } |

S(F,n) upper bounded by 2^n

equality if x1,...,xn is « shattered »

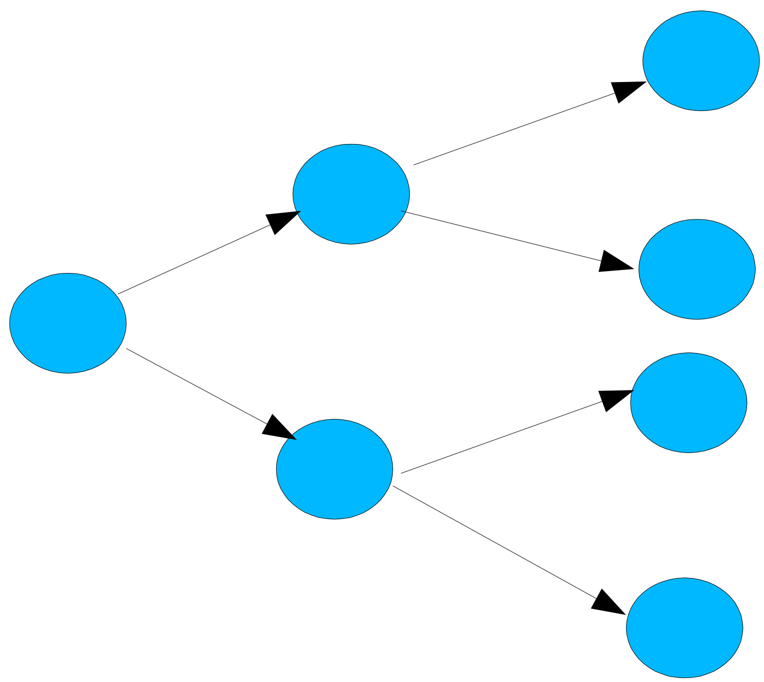
VC-dim(F) = sup { n; S(F,n) = 2 ^ n }

==> *quantifies the difficulty of learning on F*

VC-dimension : consequences

- *VC-dim finite : fast convergence (faster if V small)*
- *VC-dim infinite :*
  for any x, n, there exists P such that $E\ Ln > 1/4$
  - *no infinite set shattered and a few technical properties :*
    $Ln \to 0$ almost surely nearly $O(1/n)$
    by choosing the simplest
  - *at least one infinite set shattered*
    $Ln \to 0$ still possible but arbitrarily slowly

VC-dimension of theories



→ *infinite VC-dimension*

Theories with finite description length

X shattered ==> |F| >= 2^|X|

→ *infinite VC-dimension*
but no infinite set shattered

==> *if the number of examples is sufficient, then the optimal theory is reached*
==> *the convergence is « fast »*

Consequences

T= {e1,...,en} — *space complexity :-(* *convergence rate :-(*

T=pruning { e1, ... , en } — *space complexity :-(* *convergence rate :-(*

T=shortest $\vdash$ *{ e1, ... , en }* — *space complexity :-)* *convergence rate :-)*

Consequences for algorithmic approximations

T=shortest $\vdash$ *{ e1, ... , en } that does not prove* $\perp$

= non-Turing computable !

==> idem within k steps (k(n) quickly increasing)

T converges also (same rate, same limit) !

Some questions : « finitely discribable » theories

- *Finitely « discribable » theories :*
  Ln → 0 quickly and the optimal theory is reached.
  ==> extends to recursive theories ? (not sure)

- *Non-finitely « discribable » theories :*
  Ln → 0 still possible but arbitrarily slowly and the optimal theory is never reached.

Some questions : theories with finite VC-dimension

Complete theories have VC-dim 0.

Essentially undecidable theories have infinite VC-dim

Are there <u>interesting</u> classes of theories such that the VC-dim is positive and finite ?

- *more easily computable classes of theories ?*
- *problem-specific classes of theories ?*