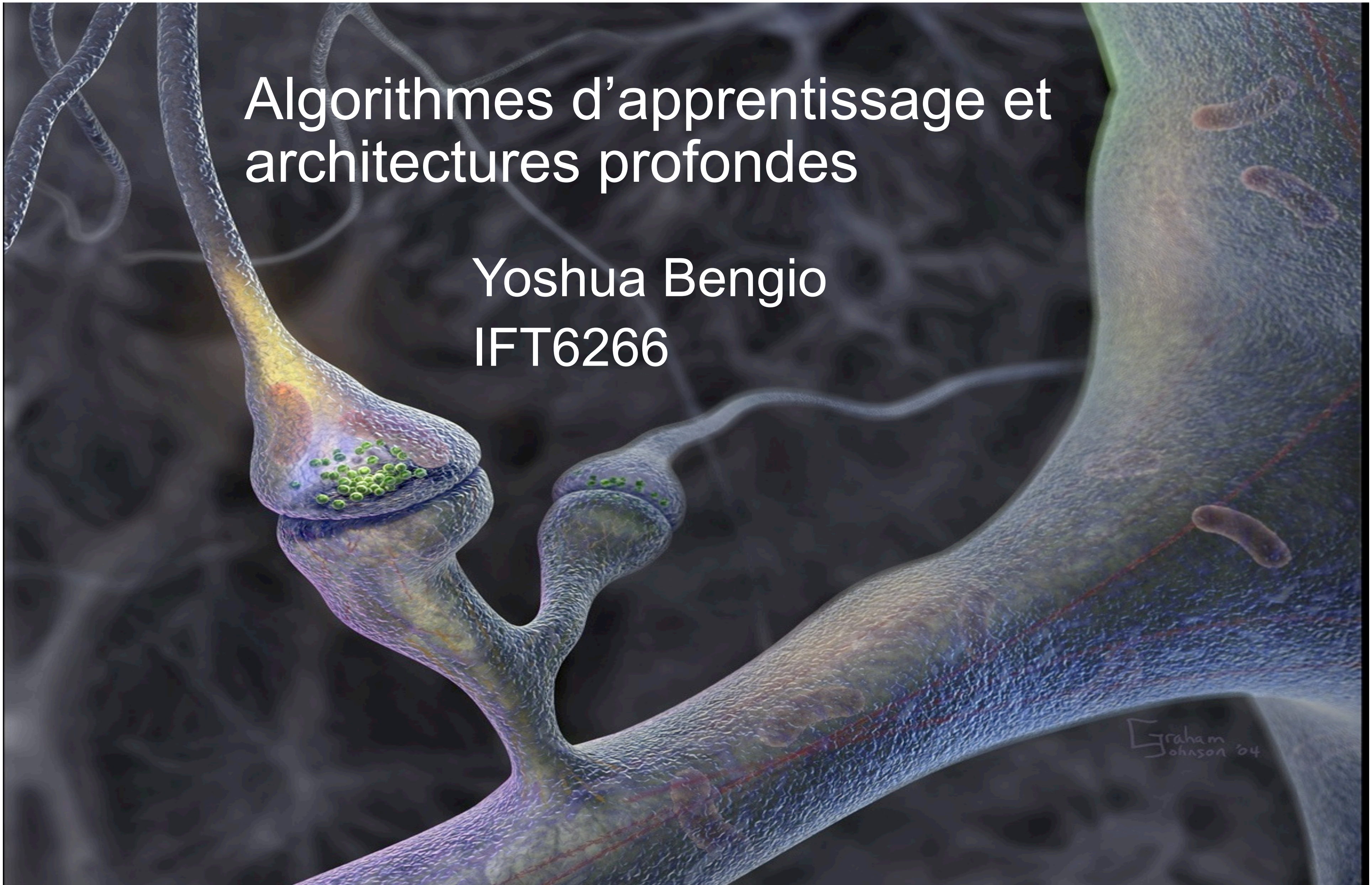
Algorithmes d’apprentissage et architectures profondes

Yoshua Bengio

IFT6266



Université de Montréal

Département d'informatique et de recherche opérationnelle

1. Machine Learning
2. Local generalization
3. The curse of highly-variable functions
4. Inspiration from the brain:
    1. Distributed representations
    2. Deep architectures

1. INTELLIGENCE

Nécessite l’acquisition de connaissances

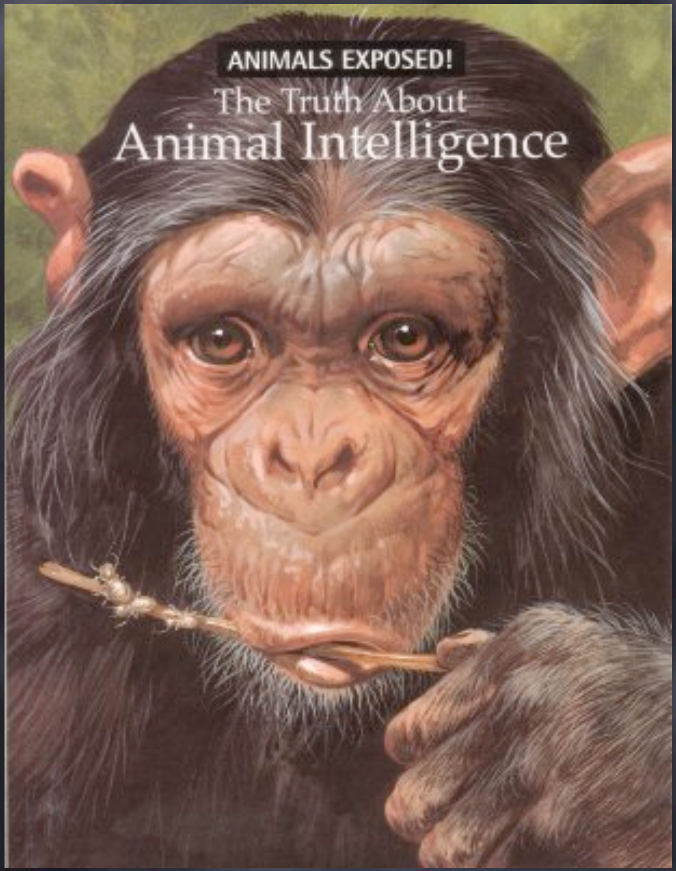
= APPRENTISSAGE

Intelligence naturelle

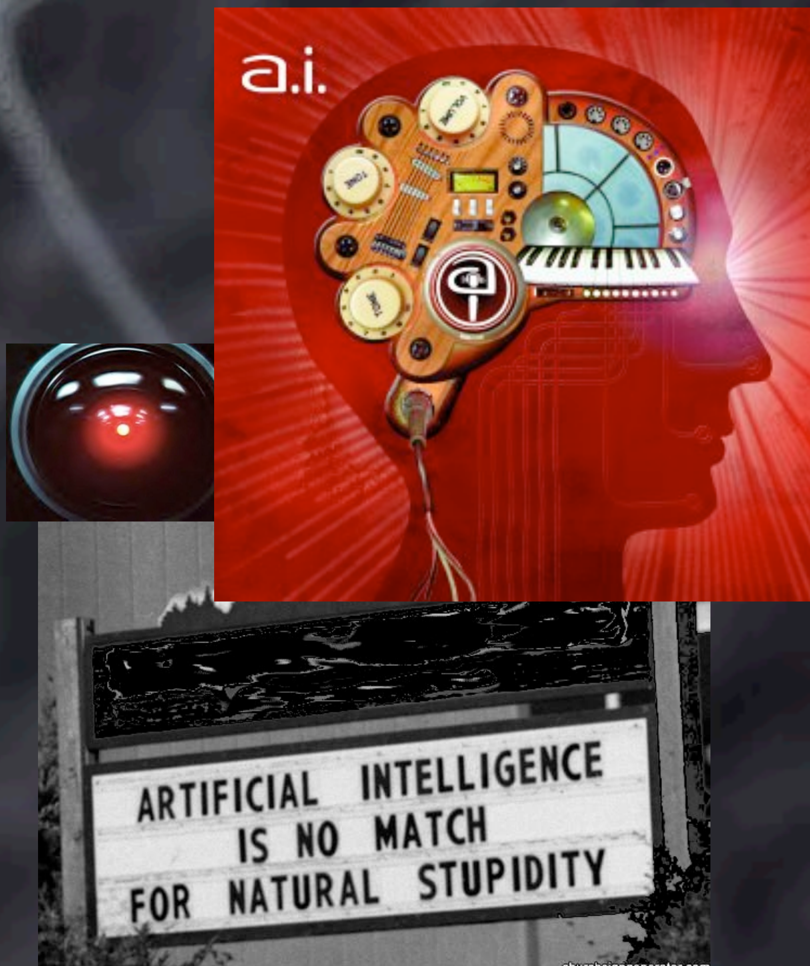
- S'adapter à de nouvelles situations pour survivre / se reproduire.

- Schémas comportementaux variés et adaptés

- Évolution + apprentissage individuel



Intelligence artificielle



- Encore loin du but!
- Pourquoi?
- Trop pressés d’y arriver plutôt que de comprendre?

Où prendre les connaissances?

- Intelligence:

  connaissances implicites ou explicites

- Faillite de l'IA symbolique-explicite

A prendre les connaissances

- Chaque exemple, chaque expérience contribue à notre modèle du monde

- Innées vs acquises

- Apprentissage de tâches non prévues par l'évolution!

Statistical Learning

- New multi-disciplinary field

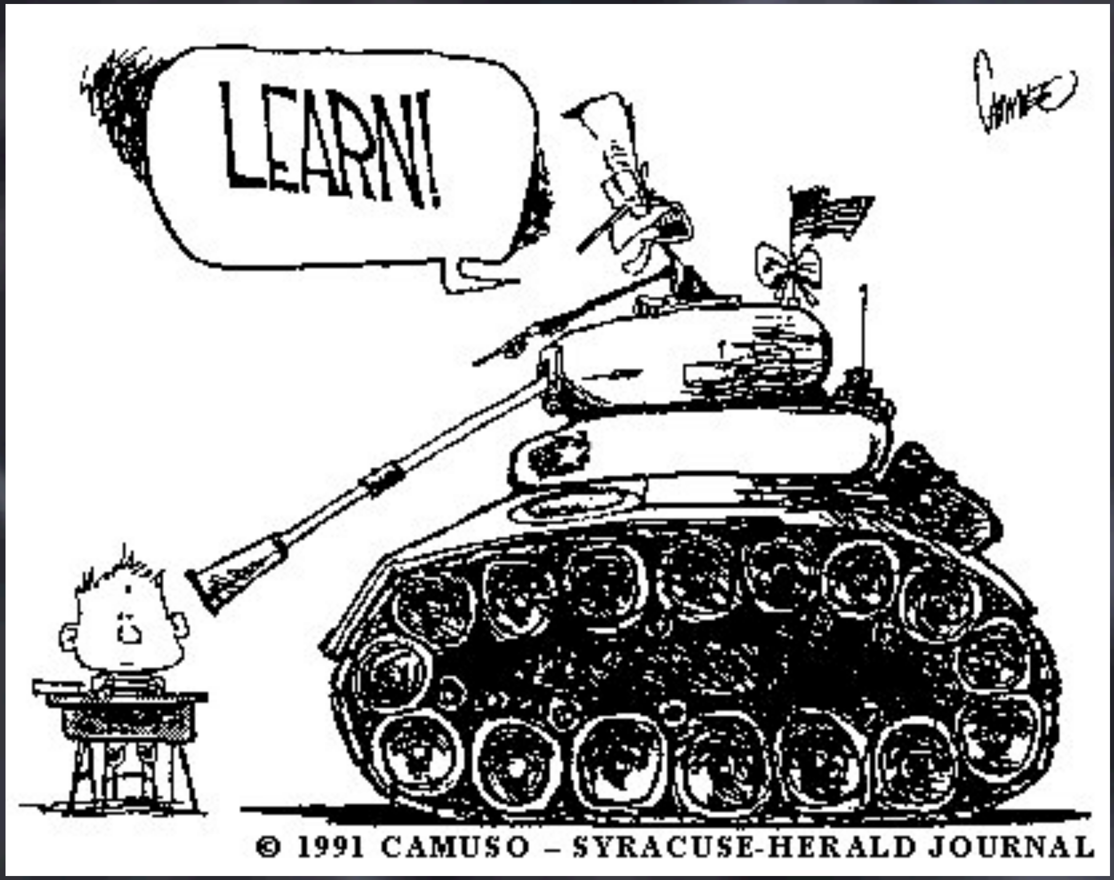
- Numerous applications

What is Learning?

Learn underlying and previously unknown structure, from examples

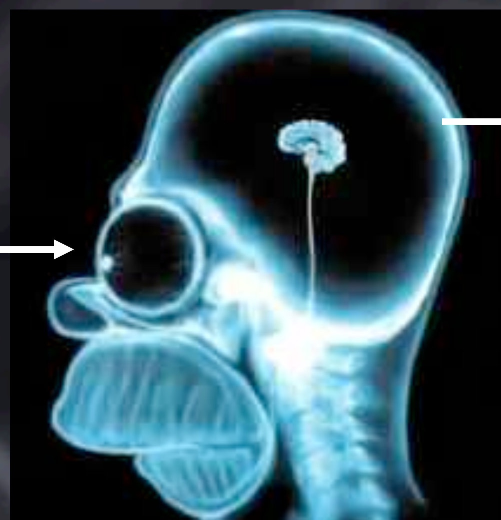
= CAPTURE THE VARIATIONS

Supervised Learning



Supervised Training Example

Entrée X



Sortie f(X)

six

Cible Y

deux!

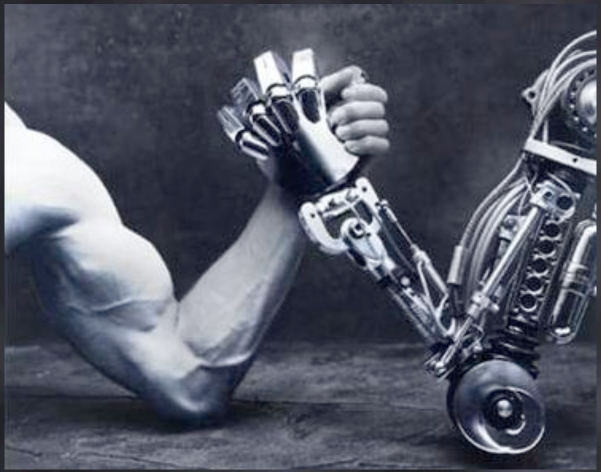
Learn by heart

or

Generalize

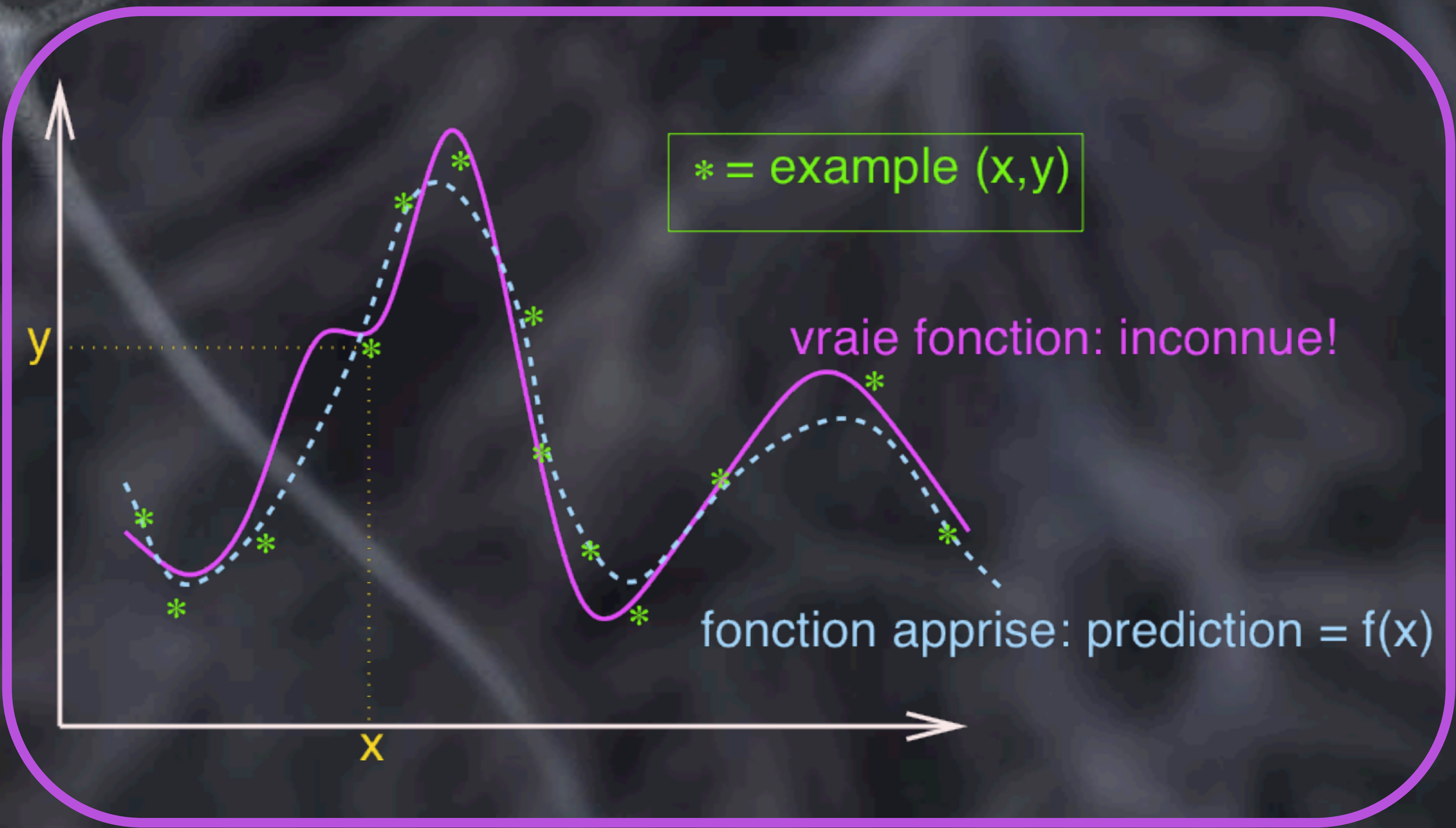
- Easy for machines
- Hard for humans (!!???)

- Mathematically: fondamentally difficult

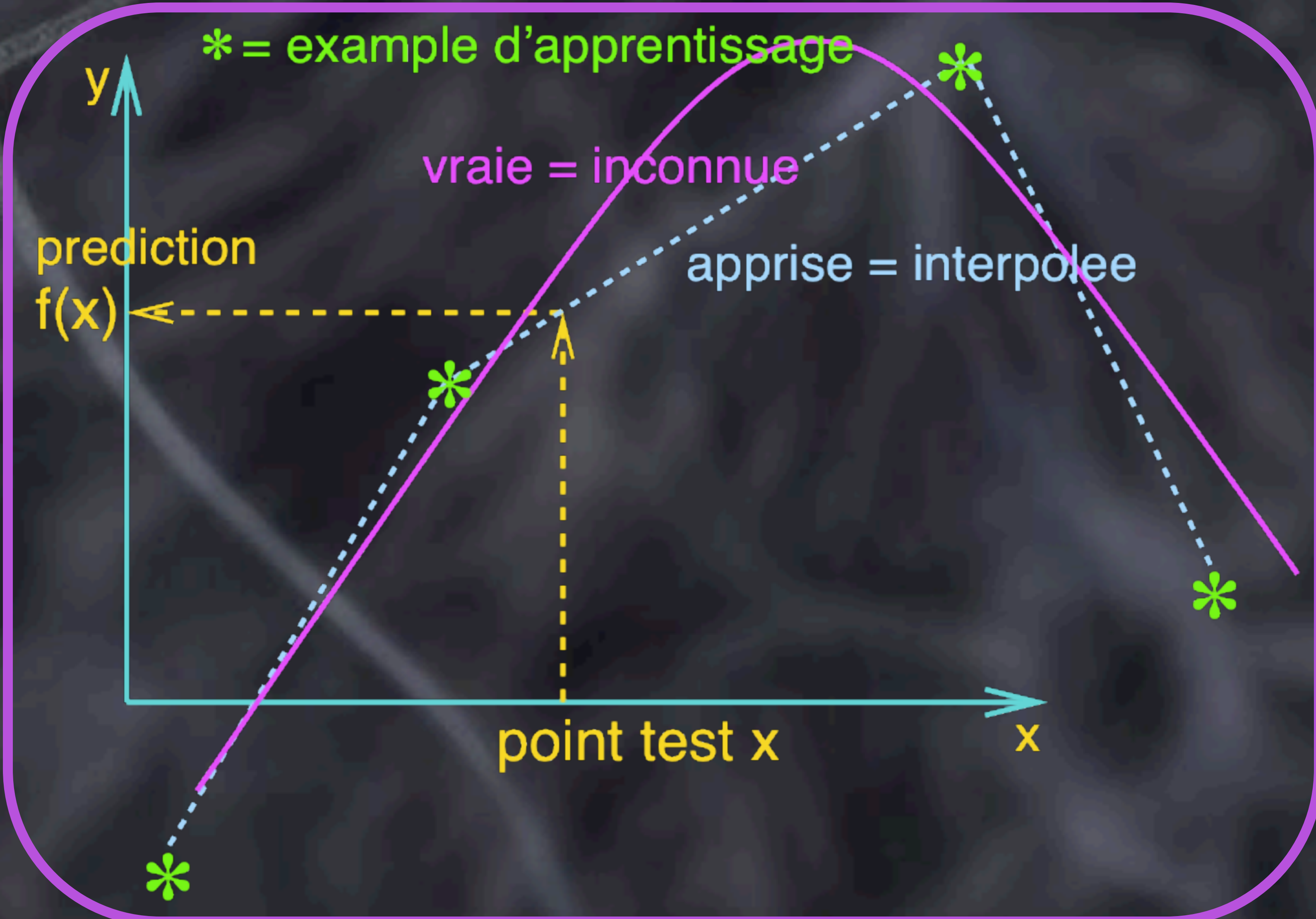


Easier for humans!!!!

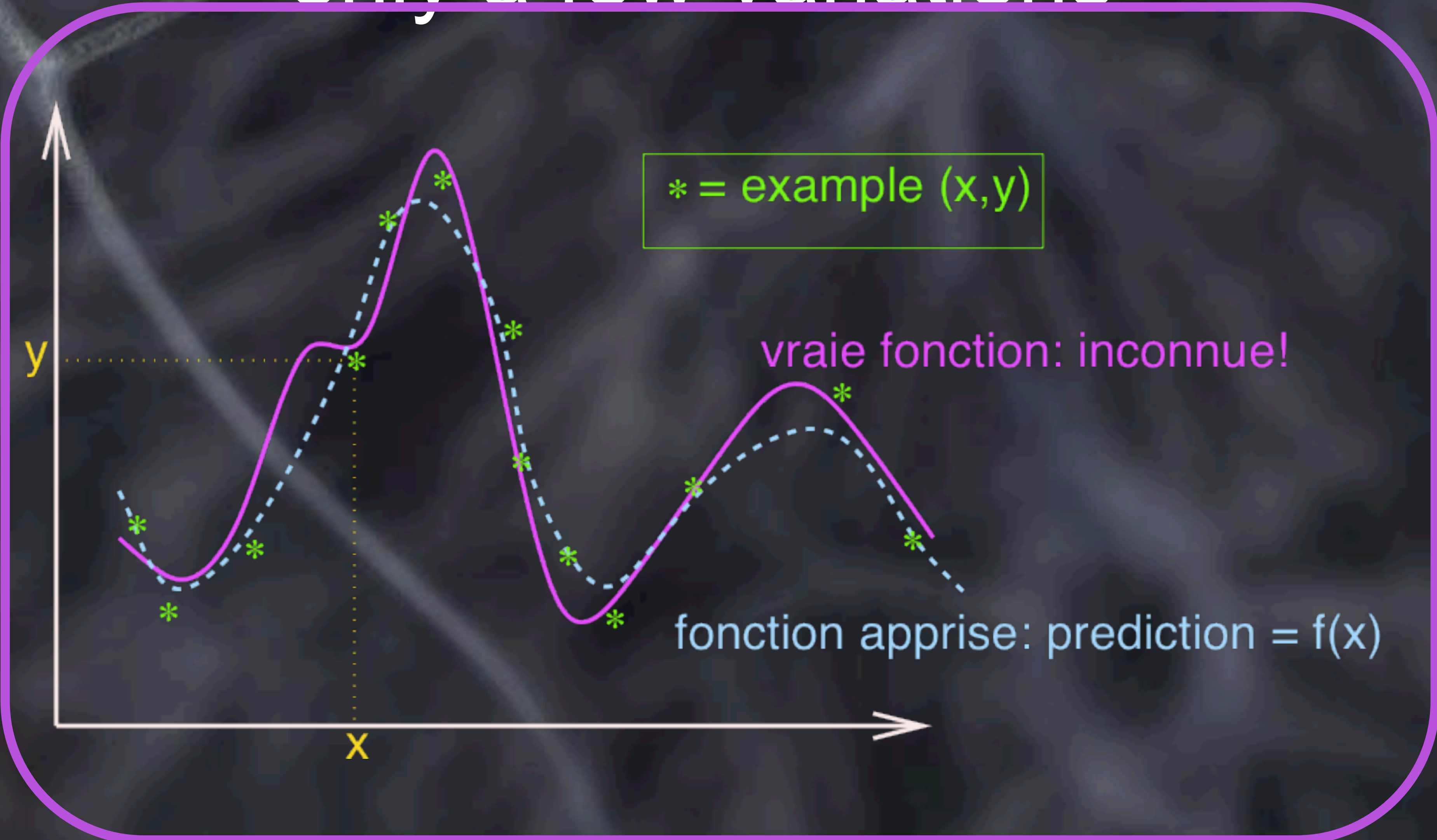
Supervised Learning and Local Generalization



Locally capture the variations

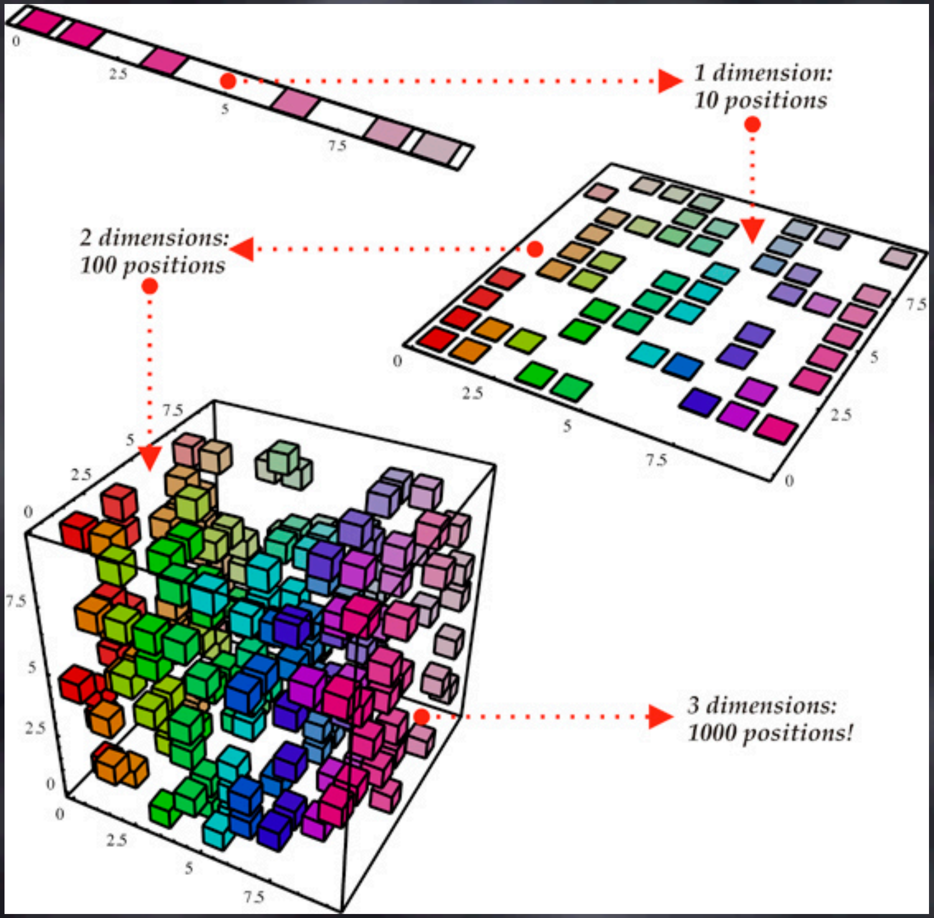


Easy when there are only a few variations



Curse of dimensionnality

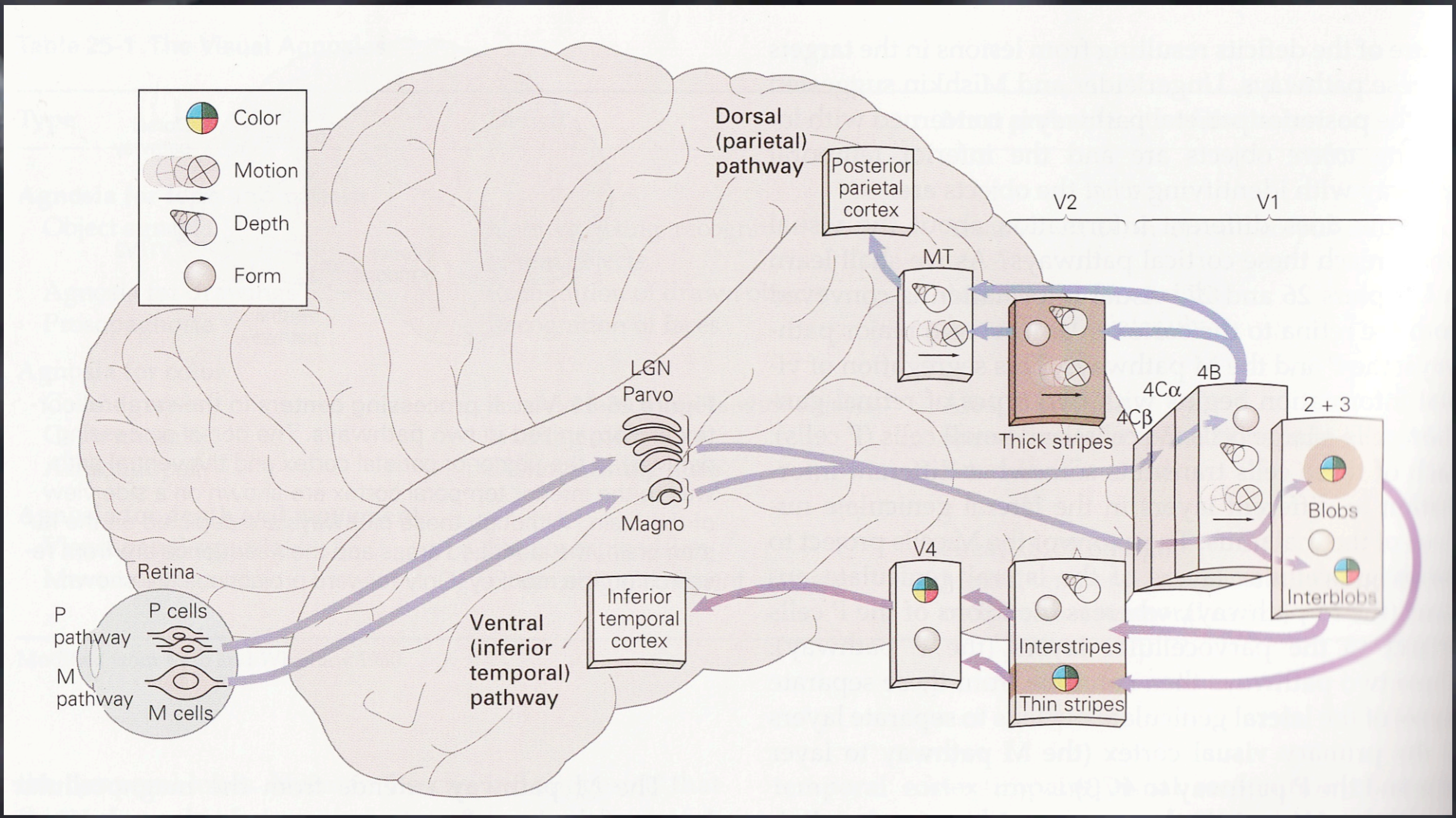
To generalize locally, need examples representative of each possible variation.



Learning Brains

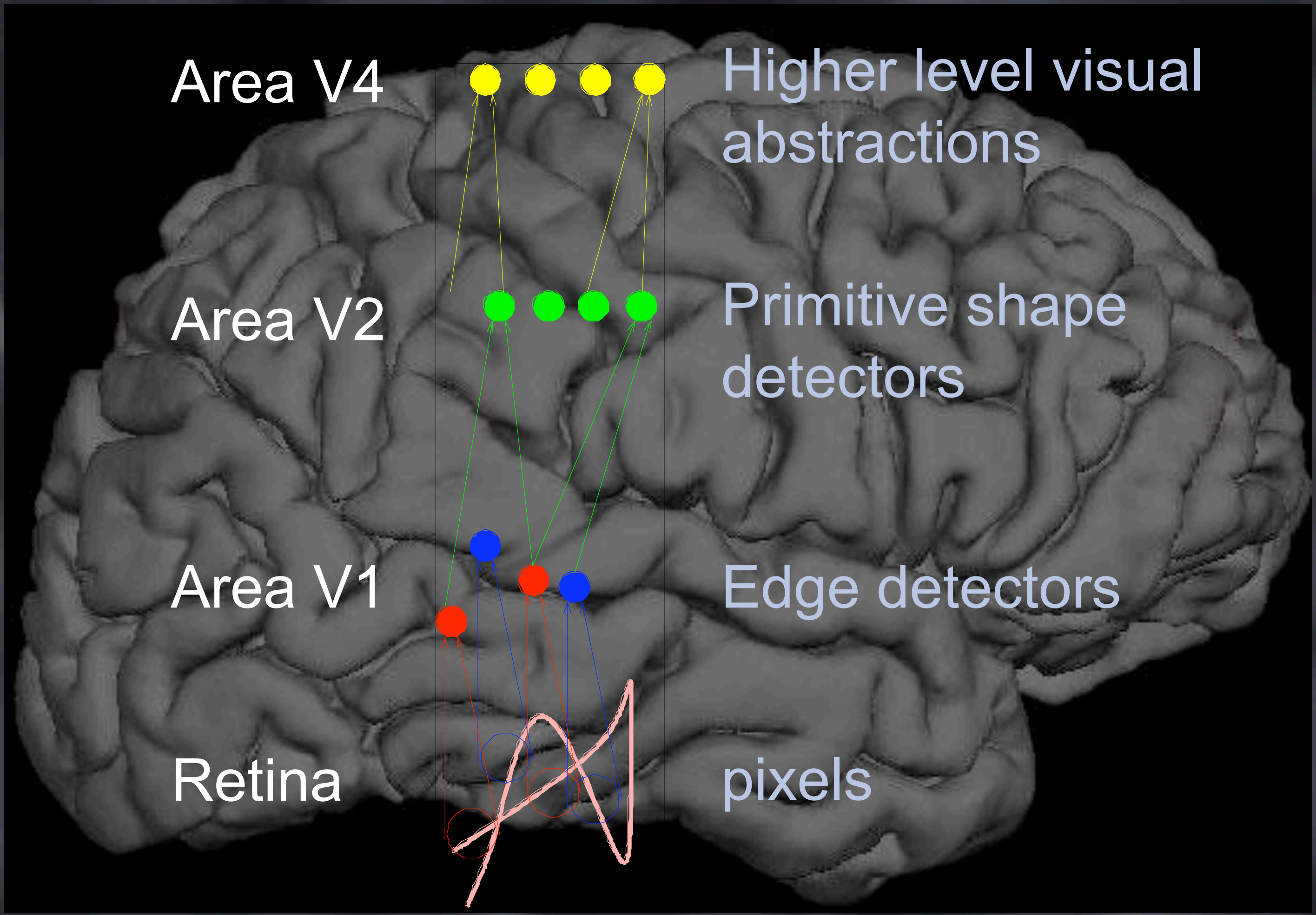
- 10^{11} neurons, 10^{14} synapses
- Complex neural network
- Learning: modify synapses

Visual System



Sequence of transformations / abstraction levels

Deep Architecture in the Brain



Distributed Representations

Many neurons active simultaneously in the brain: around 1%

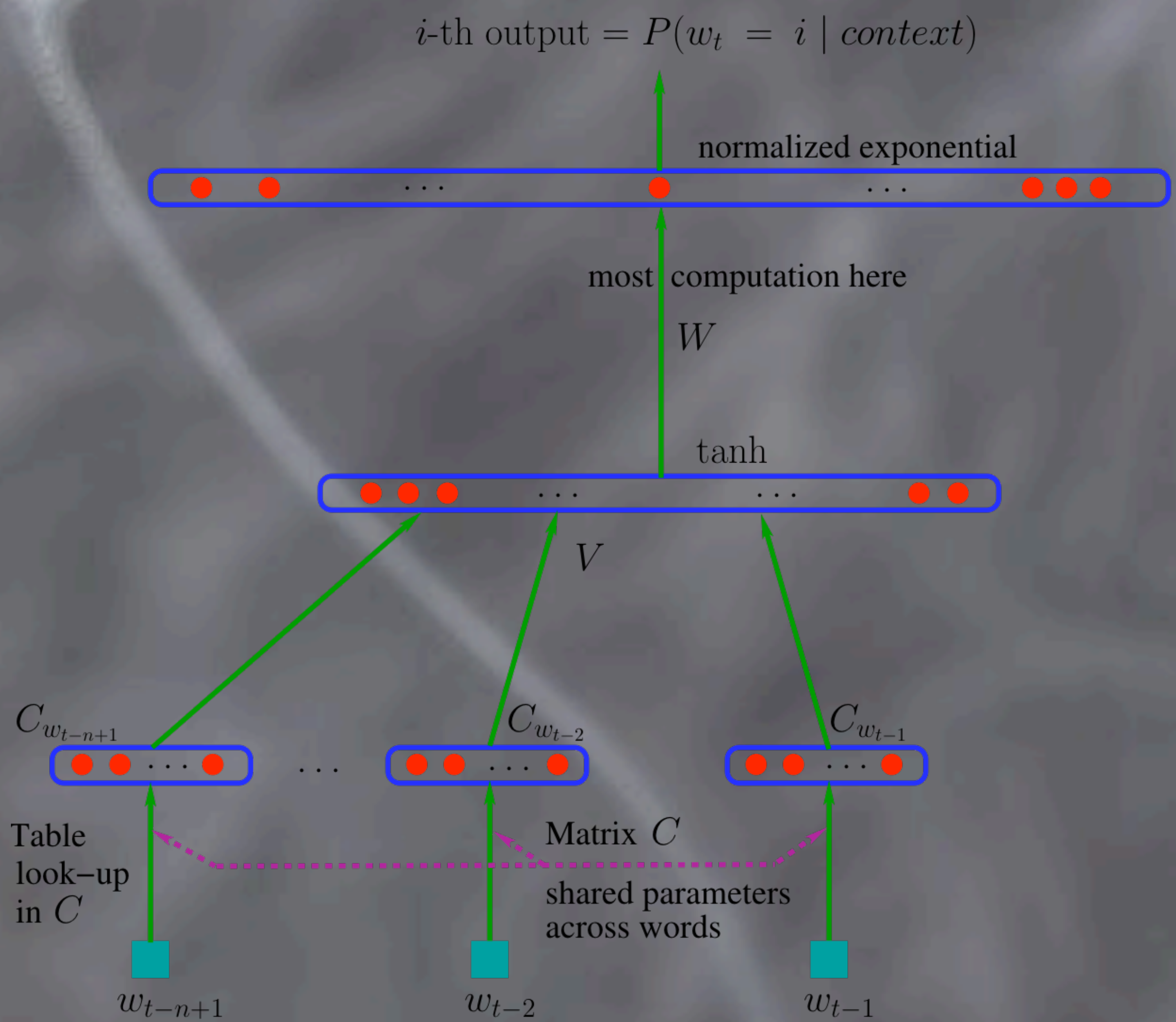
The input is represented by the activation of a set of features that are not mutually exclusive.

Can be exponentially more efficient than local representations

Neurally Inspired Language Models

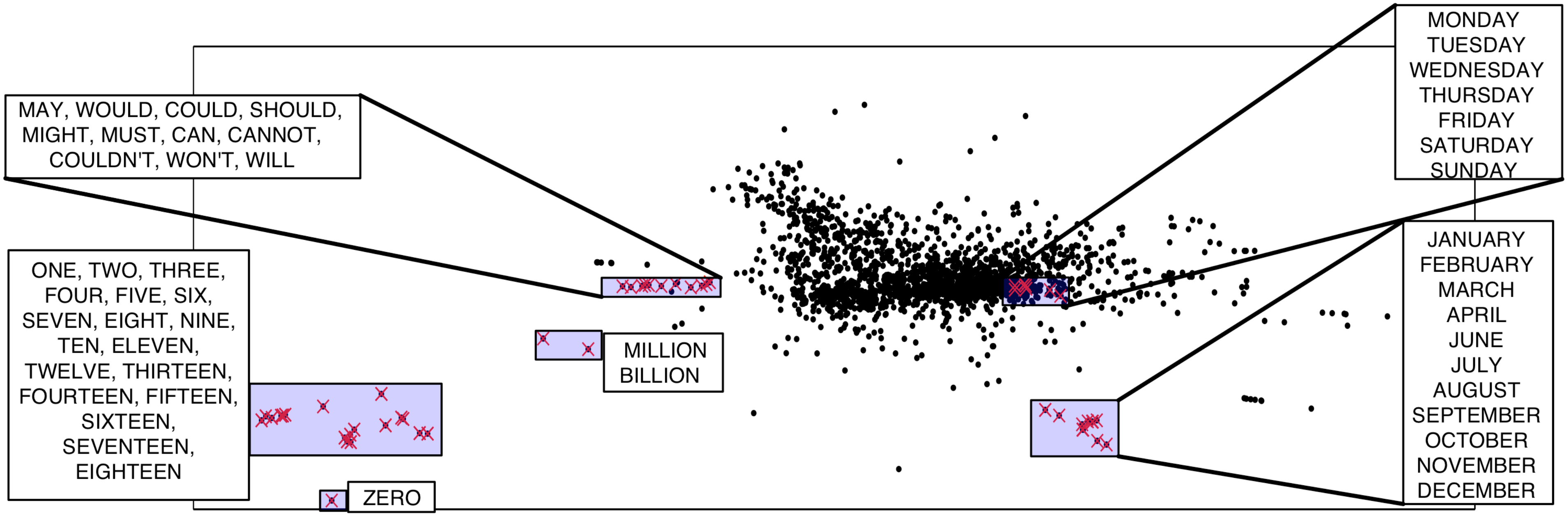
- Classical statistical models of word sequences: local representations
- Input = sequence of symbols, each element of sequence = 1 of N possible words
- Distributed representations: learn to embed the words in a continuous-valued low-dimensional semantic space

Neural Probabilistic Language Models



Successes of this architecture and its descendents: beats localist state-of-the-art in NLP in most tasks (language model, chunking, semantic role labeling, POS)

Embedding Symbols



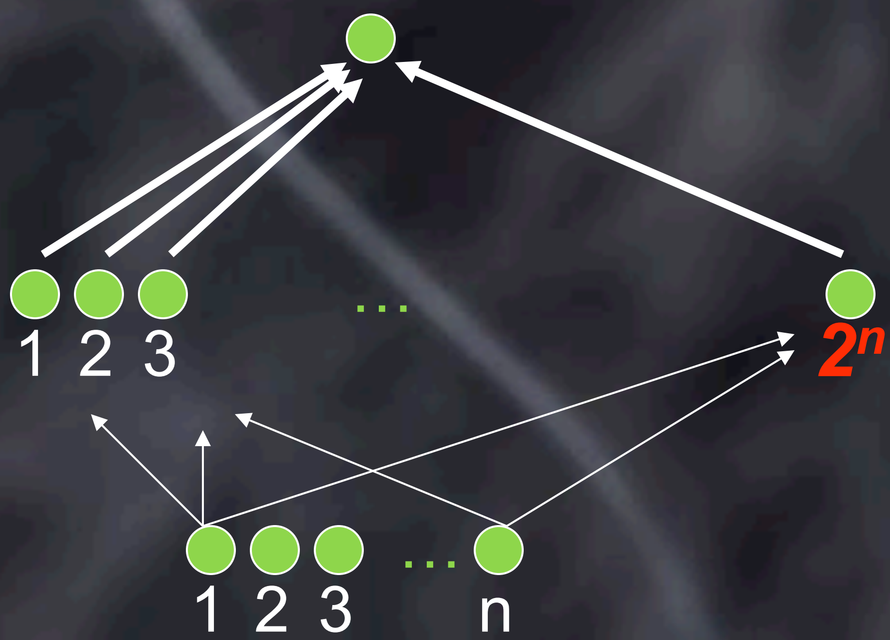
Nearby Words in Semantic Space

France	Jesus	XBOX	Reddish	Scratched
Spain	Christ	Playstation	Yellowish	Smashed
Italy	God	Dreamcast	Greenish	Ripped
Russia	Resurrection	PS###	Brownish	Brushed
Poland	Prayer	SNES	Bluish	Hurled
England	Yahweh	WH	Creamy	Grabbed
Denmark	Josephus	NES	Whitish	Tossed
Germany	Moses	Nintendo	Blackish	Squeezed
Portugal	Sin	Gamecube	Silvery	Blasted
Sweden	Heaven	PSP	Greyish	Tangled
Austria	Salvation	Amiga	Paler	Slashed

Insufficient Depth

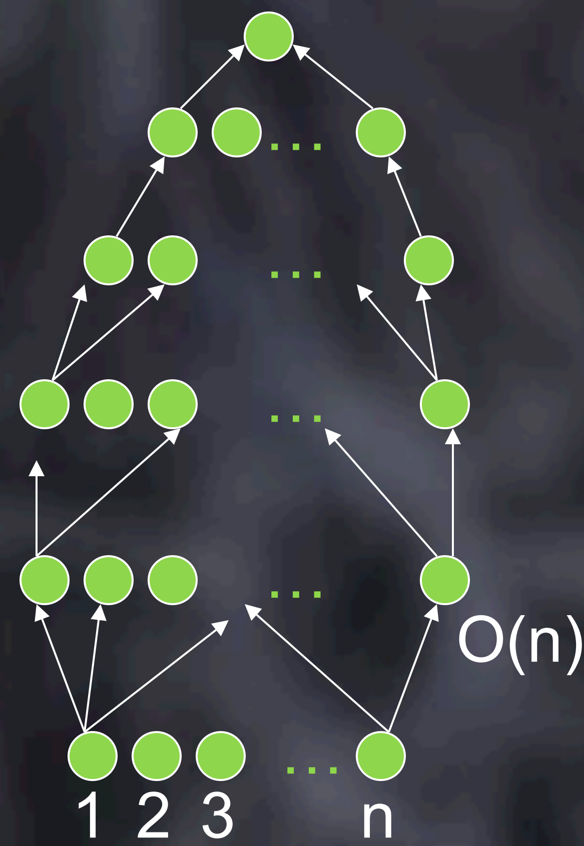
Insufficient depth =

May require exponential-size architecture



Sufficient depth =

Compact representation



Breakthrough!

Before 2006

Failure of deep architectures

After 2006

Train one level after the other, unsupervised, extracting abstractions of gradually higher level

Success of deep distributed neural networks

Since 2006

- Record broken on MNIST handwritten characer recognition benchmark
- A component of winning NetFlix entry
- State-of-the-art beaten in language modeling
- NSF et DARPA are interested…

Deep Architectures Work Well

- Beating shallow neural networks on vision and NLP tasks
- Beating SVMs on visions tasks from pixels (and handling dataset sizes that SVMs cannot handle in NLP)
- Reaching state-of-the-art performance in NLP
- Beating deep neural nets without unsupervised component
- Learn visual features similar to V1 and V2 neurons

Deep Motivations

- Brains have a deep architecture
- Humans organize their ideas hierarchically, through composition of simpler ideas
- Unsufficiently deep architectures can be exponentially inefficient
- Distributed (possibly sparse) representations are necessary to achieve non-local generalization, exponentially more efficient than 1-of-N enumeration latent variable values
- Multiple levels of latent variables allow combinatorial sharing of statistical strength

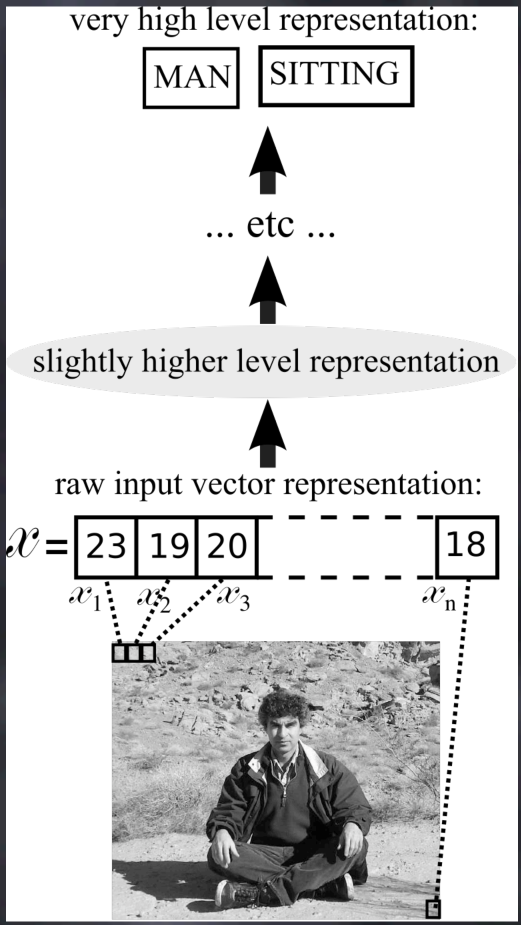
Neuro-cognitive inspiration

- Brains use a distributed representation
- Brains use a deep architecture
- Brains heavily use unsupervised learning
- Brains learn simpler tasks first
- Human brains developed with society / culture / education

Deep Architecture in our Mind

- Humans organize their ideas and concepts hierarchically
- Humans first learn simpler concepts and then compose them to represent more abstract ones
- Engineers break-up solutions into multiple levels of abstraction and processing

It would be nice to learn / discover these concepts
(knowledge engineering failed because of poor introspection?)



Sharing Components in a Deep Architecture

Polynomial expressed with shared components: advantage of depth may grow exponentially