

# 3EL05

## Chapitre 4 Complexité et apprentissage automatique

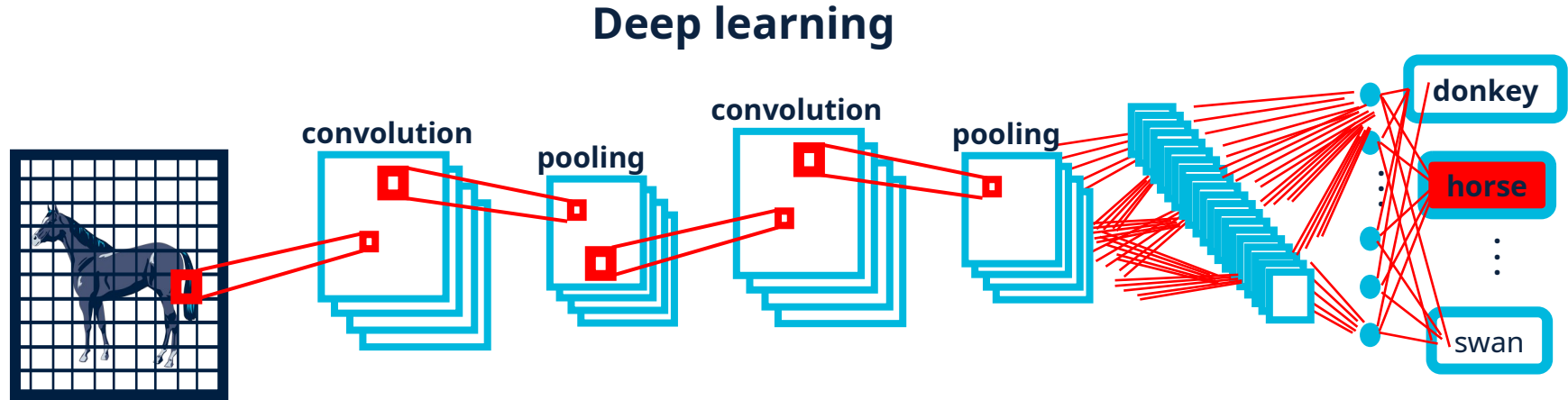
# Apprentissage avec connaissances



# Contenu

- Analogies
- Clustering
- Longueur de description minimale
- Détection d'anomalies
- Apprentissage avec connaissances

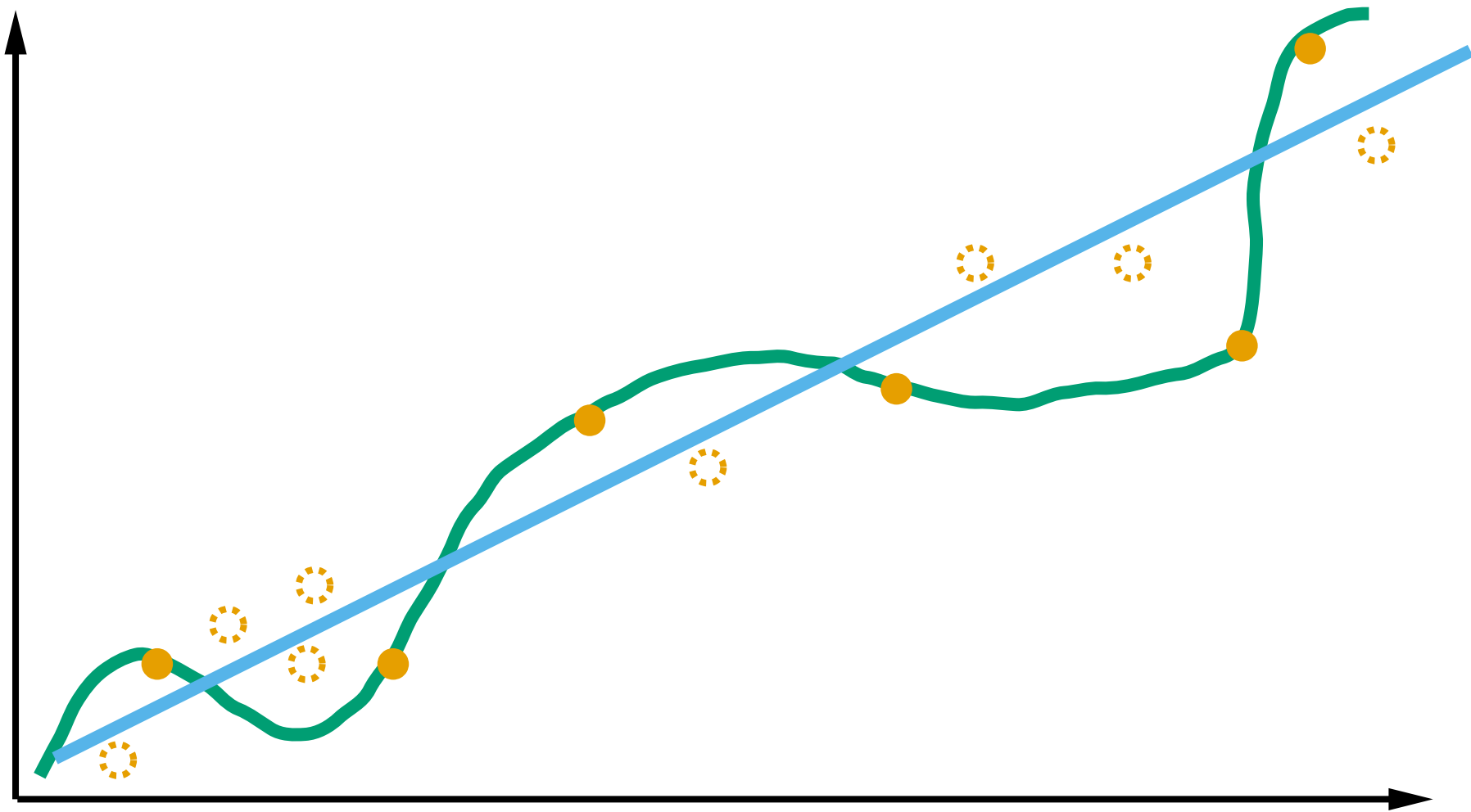
# Apprentissage automatique



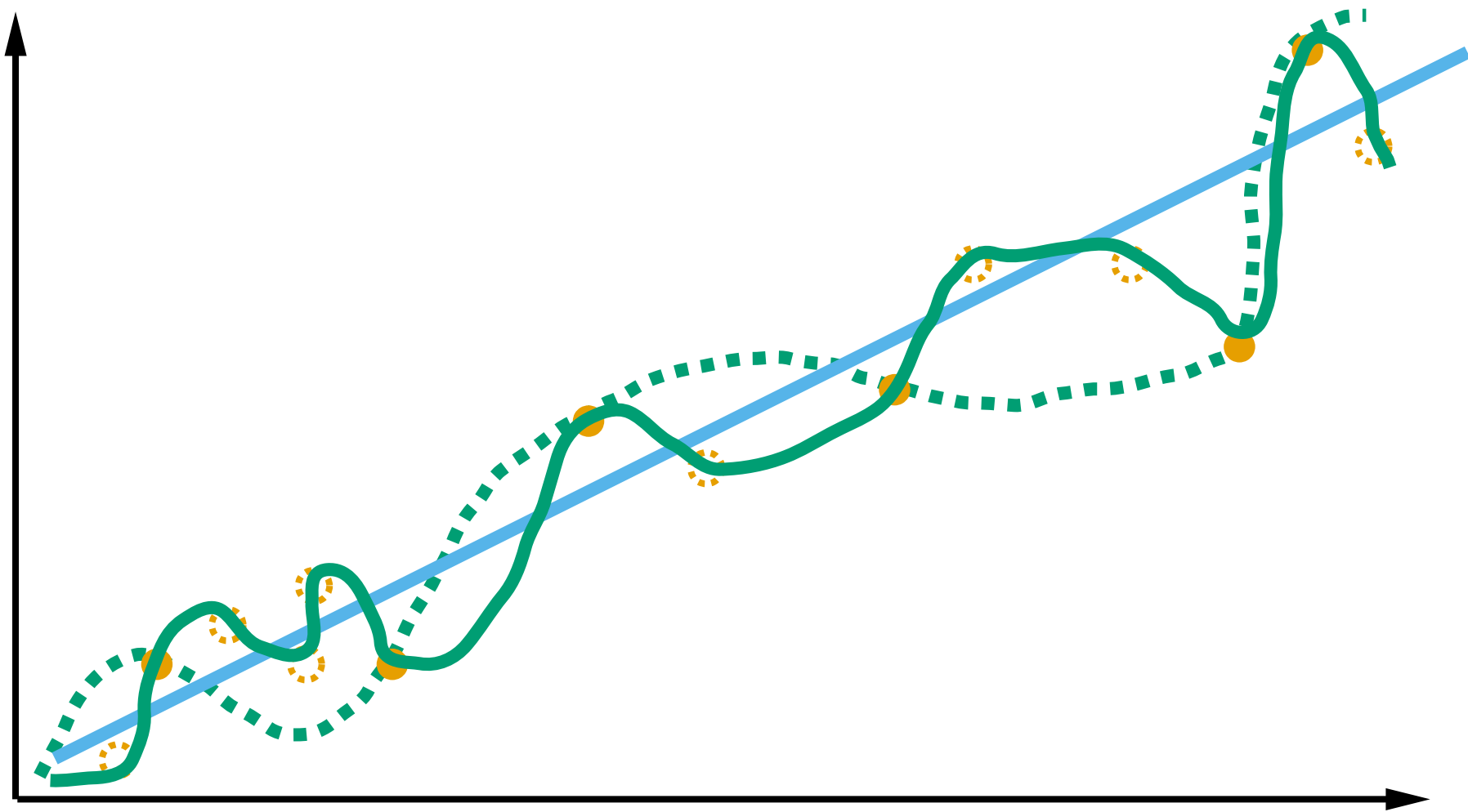
Apprentissage automatique :

- Collecter des observations
- Exprimer les probabilités des observations avec une fonction paramétrique
- Trouver les paramètres qui maximisent la probabilité des observations

# Longueur de description minimale



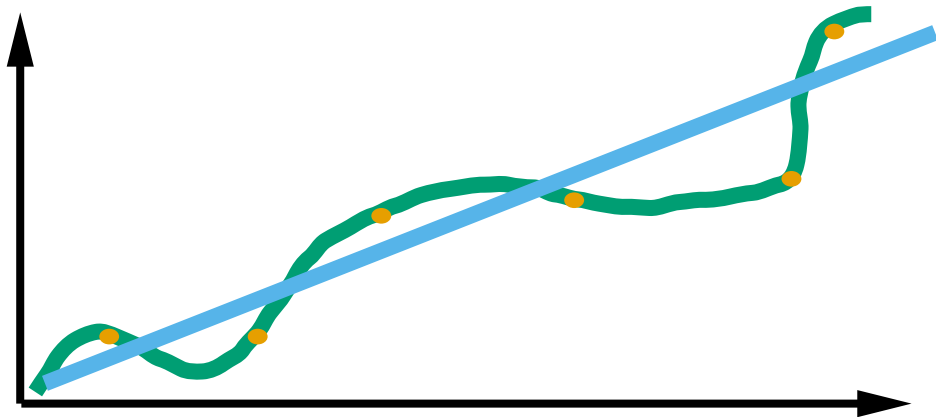
# Longueur de description minimale



# Longueur de description minimale

- Un modèle doit compresser les observations disponibles
- La taille du modèle doit être prise en compte dans la compression

$$K(\text{observations}) \leq K(\text{modele}) + K(\text{observations}|\text{modele})$$

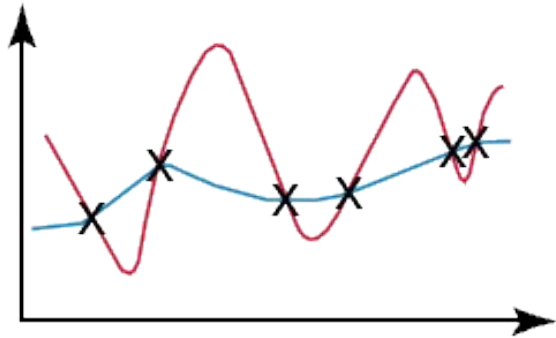


$$y = ax + b \quad \sum \log(\text{erreur})$$

$$y = \sum_k a_k x^k \quad 0$$

# Complexité et apprentissage automatique

- Machine learning is compression
- ML classique : expliquer les données observées grâce à des équations et des paramètres
- → Les équations et les paramètres sont une compression des données



$$y = f(x) = \sum_{i=1}^k a_k x^k$$

1223334444...

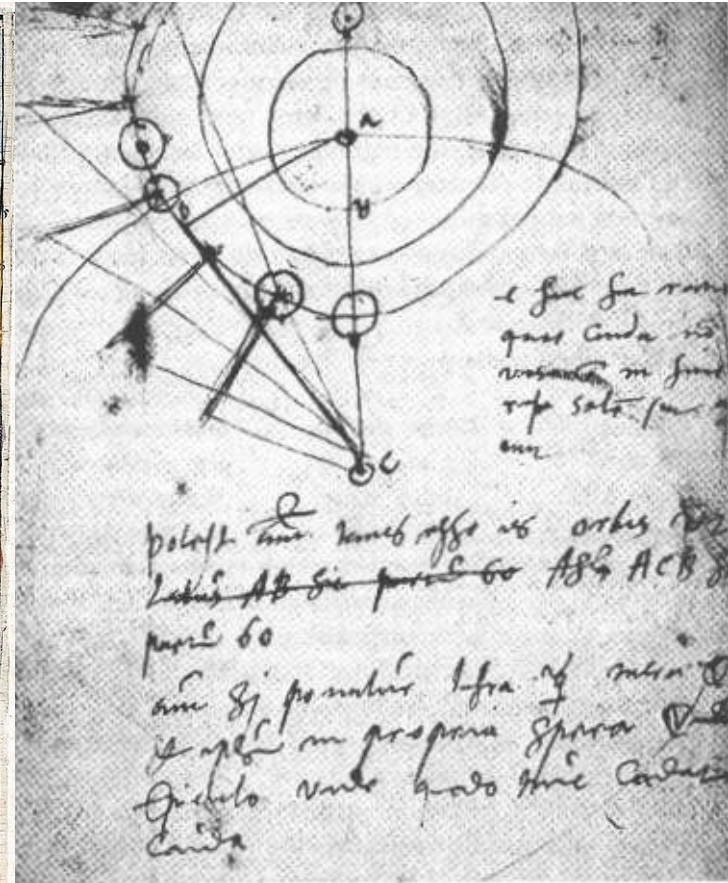
$$\bigoplus_{m=1}^n m^m$$



# Longueur de description minimale



Tycho Brahe



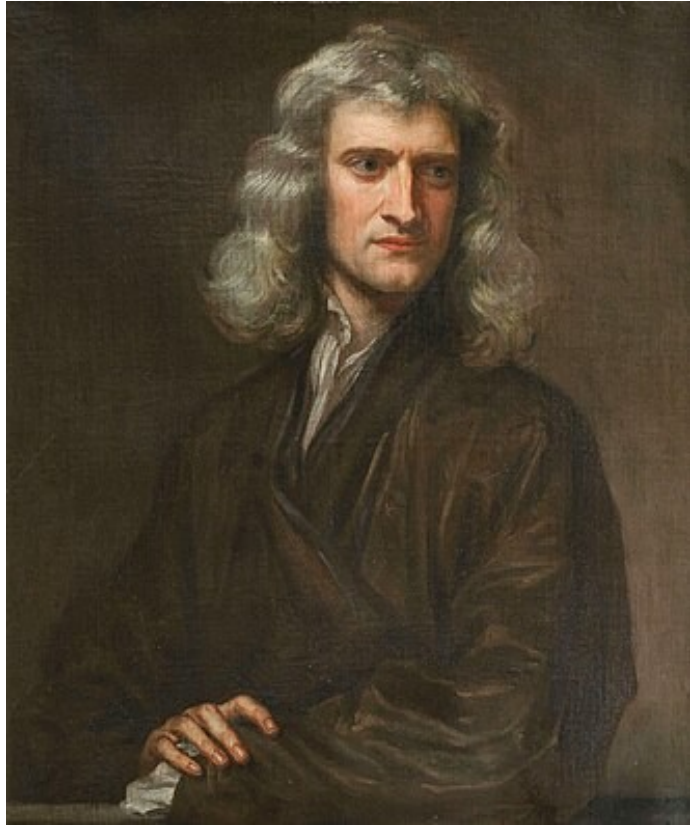
# Longueur de description minimale



- 1) Les planètes du système solaire décrivent des trajectoires elliptiques, dont le Soleil occupe l'un des foyers.
- 2) Des aires égales sont balayées dans des temps égaux.
- 3)  $\frac{T^2}{a^3} = \text{constante}$

Johannes Kepler

# Longueur de description minimale



Isaac Newton

1) Tout corps persévère dans l'état de repos ou de mouvement uniforme en ligne droite dans lequel il se trouve, à moins que quelque force n'agisse sur lui, et ne le contraigne à changer d'état.

$$2) \frac{d\vec{p}}{dt} = \sum_i \vec{F}_i$$

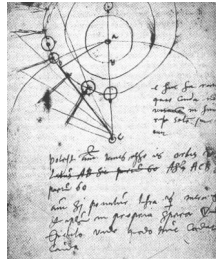
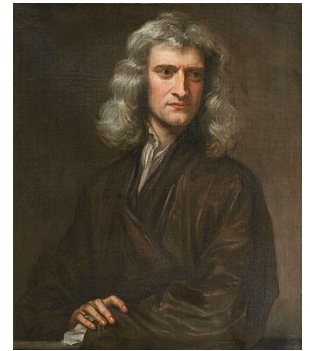
$$3) \vec{F}_{A \rightarrow B} = -\vec{F}_{B \rightarrow A}$$

$$\vec{F}_{A \rightarrow B} \cdot \vec{AB} = 0$$

# Longueur de description minimale



Réalité → Observations → Lois empiriques → Principes



$$\frac{T^2}{a^3} = \text{constante}$$

$$\frac{d\vec{p}}{dt} = \sum_i \vec{F}_i$$

# Longueur de description minimale

GPT-3  $\approx$  explication des données trouvées sur internet

ANNALS OF TECHNOLOGY

## CHATGPT IS A BLURRY JPEG OF THE WEB

*OpenAI's chatbot offers paraphrases, whereas Google offers quotes. Which do we prefer?*

By Ted Chiang

February 9, 2023

# Longueur de description minimale

Quel est le meilleur modèle ?

- Une équation avec 8 paramètres qui prédit 92% des observations
- Un réseau de neurones avec 12M de paramètres entraîné sur 1M d'exemples qui prédit 96% des observations

# Longueur de description minimale

Soit le modèle suivant:

*Il y a deux sortes d'individus :*

- 1. Ceux qui se lèvent avant 7h*
- 2. Ceux qui se lèvent après 7h*

Ce modèle :

- A. Est un premier pas vers la compréhension du rôle du sommeil
- B. Ne sert à rien

# Analogies

g h i       $\longrightarrow$       g h j

uu vv ww       $\longrightarrow$       uu vv xx  
uu vv wx  
uu vv jj  
uu vv wj



# Analogies

$a : b :: c : d$

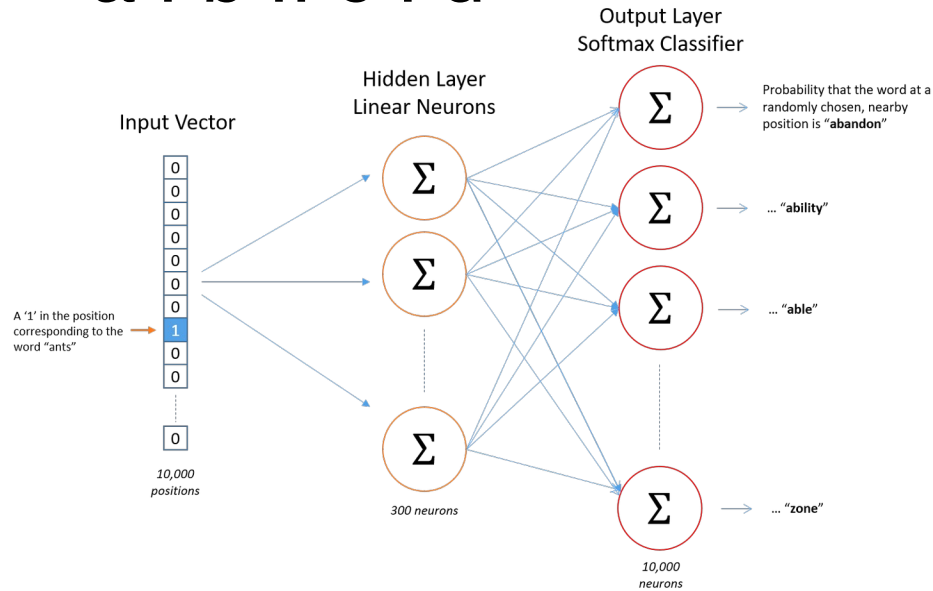
$a : b :: c : x$

$x = ?$

- king : queen :: man : woman
- rosa : rosam :: vita : vitam [latin]
- setzen : setzte :: lachen : lachte [allemand]
- solve : solves :: get : gets [anglais]
- guru : guru-guru :: pelajar : pelajar-pelajar [indonésien]
- puhua : puhuu :: juhlia : juhlii [finlandais]
- apte : inapte :: élu : \*inélu

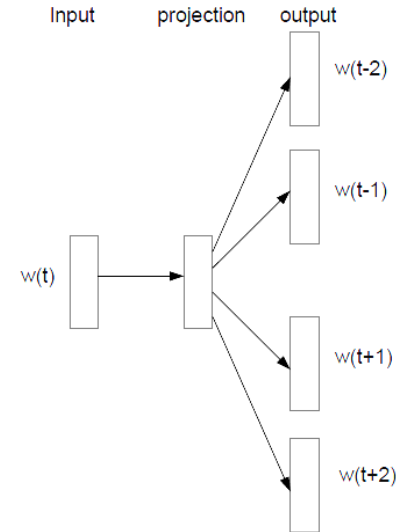
# Analogies

$a : b :: c : d$



$a : b :: c : x$

$x = ?$



Skip-gram model architecture

“Madrid” - “Spain” + “France” = “Paris”

# Analogies

- Des millions de paramètres
- 6 milliards de mots

1. CLASS-INCLUSION  
Taxonomic



1. CLASS-INCLUSION  
Class:Individual



2. PART-WHOLE  
Object:Component



2. PART-WHOLE  
Collection:Member



3. SIMILAR  
Synonymy



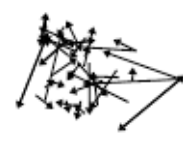
3. SIMILAR  
Dimensional Similarity



4. CONTRAST  
Contrary



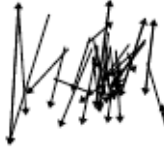
4. CONTRAST  
Reverse



5. ATTRIBUTE  
Item:Attribute



5. ATTRIBUTE  
Object:State



6. NON-ATTRIBUTE  
Item:Nonattribute



6. NON-ATTRIBUTE  
Object:Nonstate



7. CASE RELATIONS  
Agent:Instrument



7. CASE RELATIONS  
Action:Object



8. CAUSE-PURPOSE  
Cause:Effect



8. CAUSE-PURPOSE  
Cause:Compensatory Action



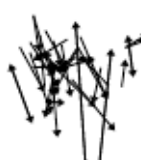
9. SPACE-TIME  
Item:Location



9. SPACE-TIME  
Time Associated Item



10. REFERENCE  
Sign:Significant



10. REFERENCE  
Representation



Chen et al, [Evaluating vector-space models of analogy](#),  
Cogsci-2017.

# Analogies

$$a : b :: c : x \quad x = ?$$

$$x = \underset{y}{\operatorname{argmin}} K(a, b, c, y)$$

$a = f(b); x = f(c)$  tel que  $K(f)$  est minimale

$$a \xrightarrow{f} b$$

$$c \xrightarrow{f} x$$

# Analogies

let, 'abc', let stores the string abc in memory.

mem, 0 retrieves the last item stored in memory  
(here: abc).

let, ?0, ?1, let stores an operation (here: concat).

```
$ python
>>> import analogy as A
>>> A.generate_string("abc, ijk")
>>> 'abcijk'
>>> A.generate_string("let, ?0, -, ?0, let, mem, 0, orang")
>>> 'orang-orang'
>>> A.instruction_complexity("let, ?0, -, ?0, let, mem, 0, orang")
35
```

# Analogies

let, 'abc', let stores the string abc in memory.

mem, 0 retrieves the last item stored in memory  
(here: abc).

let, ?0, ?1, let stores an operation (here: concat).

 What is the output of the following program?

```
let, ab, ?0, ?1, let, a, mem, 0, a, b
```

- A. abba
- B. aabba
- C. abab
- D. aabab

# Analogies

Language	#analogies	NLG_COMP	NLG_PROP	NLG_ALEA
Arabic	165,113	87.18%	<b>93.33%</b>	81.91%
Finnish	313,011	<b>93.69%</b>	92.76%	78.75%
Georgian	3,066,273	<b>99.35%</b>	97.54%	88.42%
German	730,427	<b>98.84%</b>	96.21%	95.42%
Hungarian	2,912,310	<b>95.71%</b>	92.61%	86.02%
Maltese	28,365	<b>96.38%</b>	84.72%	91.84%
Navajo	321,473	81.21%	<b>86.87%</b>	78.95%
Russian	552,423	96.41%	<b>97.26%</b>	95.46%
Spanish	845,996	<b>96.73%</b>	96.13%	94.42%
Turkish	245,721	<b>89.45%</b>	69.97%	70.06%
<b>Total</b>	<b>9,181,112</b>	<b>96.41%</b>	94.34%	87.93%

Table 2: Proportion of correct answers when solving analogies from the dataset SIGMORPHON'16 using our method NLG\_COMP and two state-of-the-art methods NLG\_PROP [Fam and Lepage, 2018] and NLG\_ALEA [Langlais *et al.*, 2009].

Murena, P.-A., Al-Ghossein, M., Dessalles, J.-L. & Cornuéjols, A. (2020). [Solving analogies on words based on minimal complexity transformation](#). *IJCAI*, 1848-1854.

rosa : rosam :: vita : x

orang : orang-orang :: burung : x

# Détection d'anomalies

1, 2, 3, 4, 5, 5, 6, 7, 8, 9, 10, 11, 12

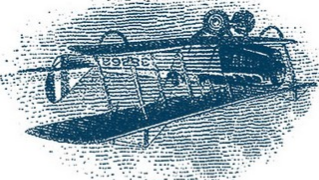


# Détection d'anomalies

17.2	21.6	19.3
20.7	18.9	20.6
19.2	18.1	19.1
18.8	18.6	20.1
18.5	15.4	14.8
16.0	18.1	18.4
20.2	20.6	18.4
15.4	17.5	18.4
19.3	15.6	18.4
18.8	17.5	18.4

# Détection d'anomalies


**INVERTED JENNY**



**The World's Most Famous Stamp**  
HISTORY OF FIRST U.S. GOVERNMENT AIRMAIL  
PRODUCTION OF FIRST U.S. AIRMAIL STAMP  
ROBEY'S DISCOVERY AT THE POST OFFICE  
SALE RECORDS OF THE 100 INVERTED JENNYS  
OWNERS' BIOGRAPHIES AND STORIES

---

**Society of Inverted Jenny Owners**



Membership requires documentation of past or present ownership of a genuine Inverted Jenny.

**MEMBER SIGN IN**    **Not a member? REGISTER HERE**



2 000 €



2 000 000 €

*Simple pour une raison simple*

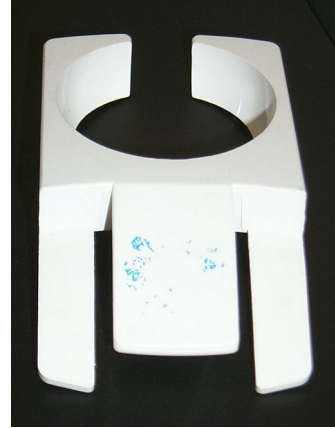
$$K(s) \leq K(f) + K(s | f)$$

# Apprentissage avec connaissances

## One-shot learning

A child learns about four+  
new words a day

Goulden, R., Nation, P. & Read, J. (1990).  
How large can a receptive vocabulary be?  
*Applied linguistics*, 11 (4), 341-363.



# Conclusion

- (Machine) learning is compression
  - Lossy compression
  - Lossless compression
- La complexité donne des critères de compression
  - Exploitation des redondances/de la structure
  - Minimum description length