HAI923 Machine Learning II
An Overview

Konstantin Todorov
LIRMM - UM - CNRS

2023

Programme HAI923

- Introduction et rappel des notions vues en M1
- Gradient descent
- Réseaux de neurones
- Sous-apprentissage et sur-apprentissage
- Extraction de trajectoire (si on a le temps)
- Deep learning
- Embeddings pour le texte et pour les graphes

Projet: classification d'images à l'aide des réseaux de neurones profonds

Organisation du module HAI923

Résponsables:

- Pascal Poncelet et Konstantin Todorov
  ⇒ prénom.nom@lirmm.fr

MCC:

- Soutenance des projets (en groupe)
- Encadrants: Pasal Poncelet, Salim Hafid

Moodle:

- Notebooks
- Supports
- Rendus
- Infos et actualités

Machine Learning?

X → f(X) → Y

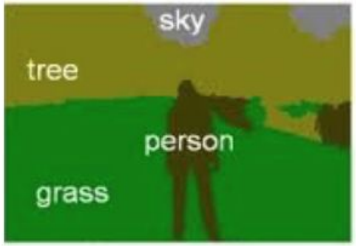
Class labels (Classification)

Real Number (Regression)

Object Recognition

→ {Dog, Cat, Sheep, Bear, Lion, ...}

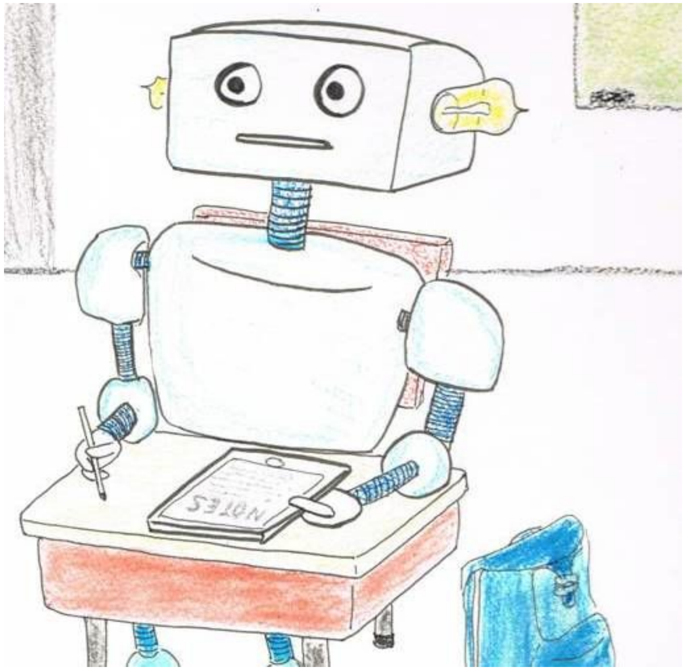
Semantic Segmentation



Sentiment Classification

Machine Learning?

Could computers be made to *learn* and to *improve* automatically with experience?



Can we develop algorithms that can *learn from* and *make predictions* on data?

(Almost) like we humans do...

Types of AI

(A vision of)

Artificial Intelligence

Symbolic
Expert Systems

Logical rules
Computational theory

Sub-symbolic
Machine Learning

DATA
Connectionist theory

Supervised
Classification
Regression...

Unsupervised
Reinforcement Learning
Clustering
Outlier Detection...

A Brief AI History Line

18^{th}– 20^{th}	Advances in probability theory (Bayes, Markov Chains,...)
1950	Turing: a learning machine that can become artificially intelligent
1956	Darthmouth Conference: Minsky, McCarthy
1957	Rosenblatt: the perceptron and its rise and fall
1967	Pattern recognition with nearest neighbours
1970-80s	AI winter, due to unrealised promises of AI research
1980s	Expert systems, rule-based systems for NLP and Computer Vision
1982	Recurrent neural networks
1986	LeCun: back-propagation reinvented
1989	Reinforcement learning
1990s	Vapnik and Cortes: Support Vector Machines shadow NN Statistics/probability-based NLP and CV: Hidden Markov Models, CNNs
2000s	NN regaining popularity due to advanced computational powers
2010s	Rapid acceleration of Deep Learning research
2010-20s	Representation learning, Vaswani's Transformers, Generative AI

Defining the Machine Learning Problem

The Defining Question of ML

How can we build computer systems that automatically improve with experience, and what are the fundamental laws that govern all learning processes?

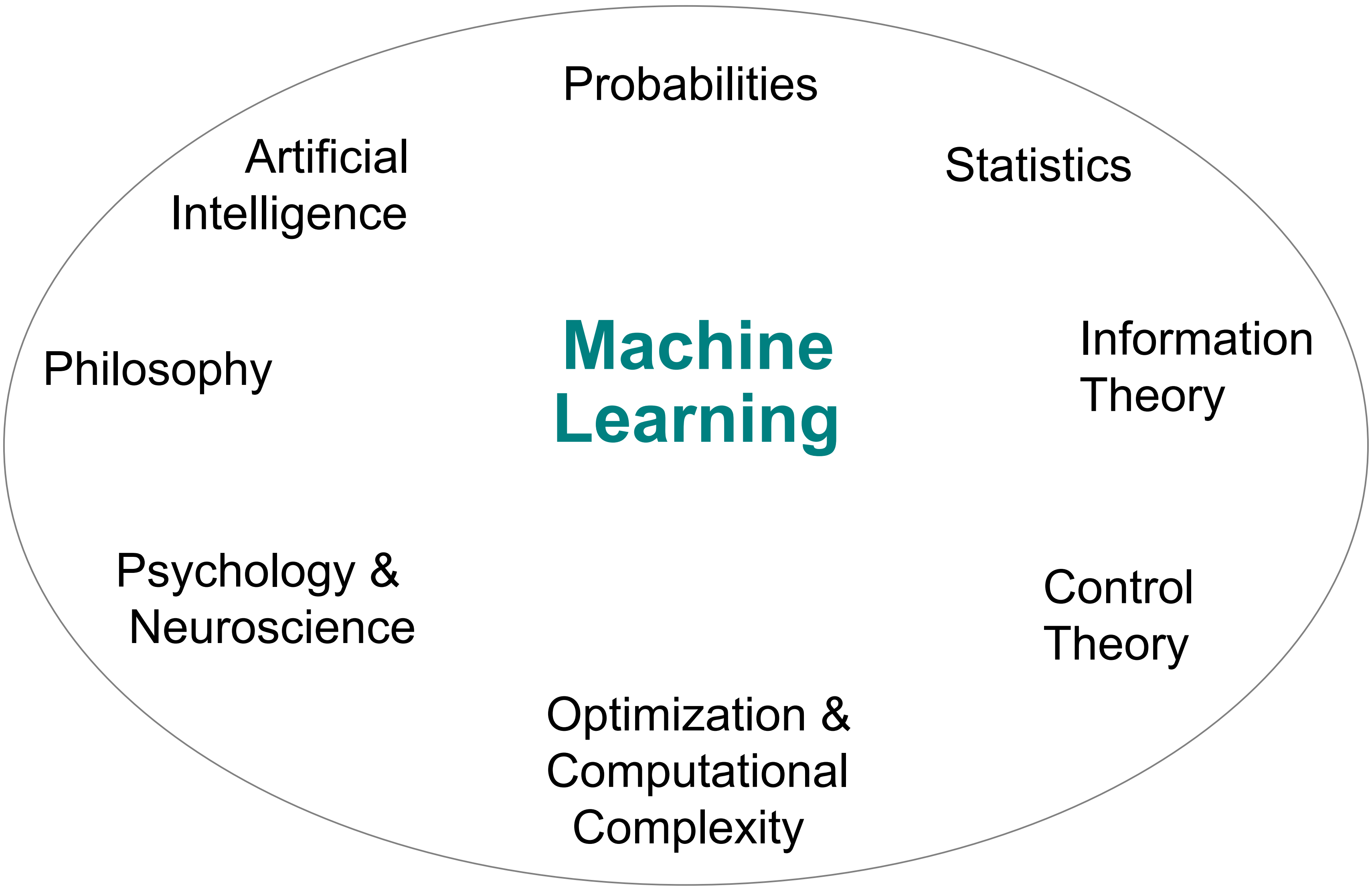
Computer Science

How can we build machines that solve problems, and which problems are inherently tractable?

Statistics

What can be inferred from data and a set of modelling assumptions, with what reliability?

A Multidisciplinary Field



A Definition of Machine Learning

"A computer program is said to learn from experience **E** with respect to some class of tasks **T** and a performance measure **P**, if its performance at tasks in T, as measured by P, improves with experience E."

– Tom M. Mitchell

- An operational, not a cognitive or an etymological definition
- A. Turing: *Can machines **think**?* →
  *Can machines **do** what thinking beings can do?*

Depending on how we define T, P, and E, the learning task might also be called by names such as *data mining*, *classification*, *clustering*, *reinforcement learning*, *etc*...

A Definition of Machine Learning

An example: filtering spam from emails

T task: decide whether an email is spam or not

P performance measure: the percent of correctly filtered emails

E training experience: a dataset of emails with associated classes (spam / email)

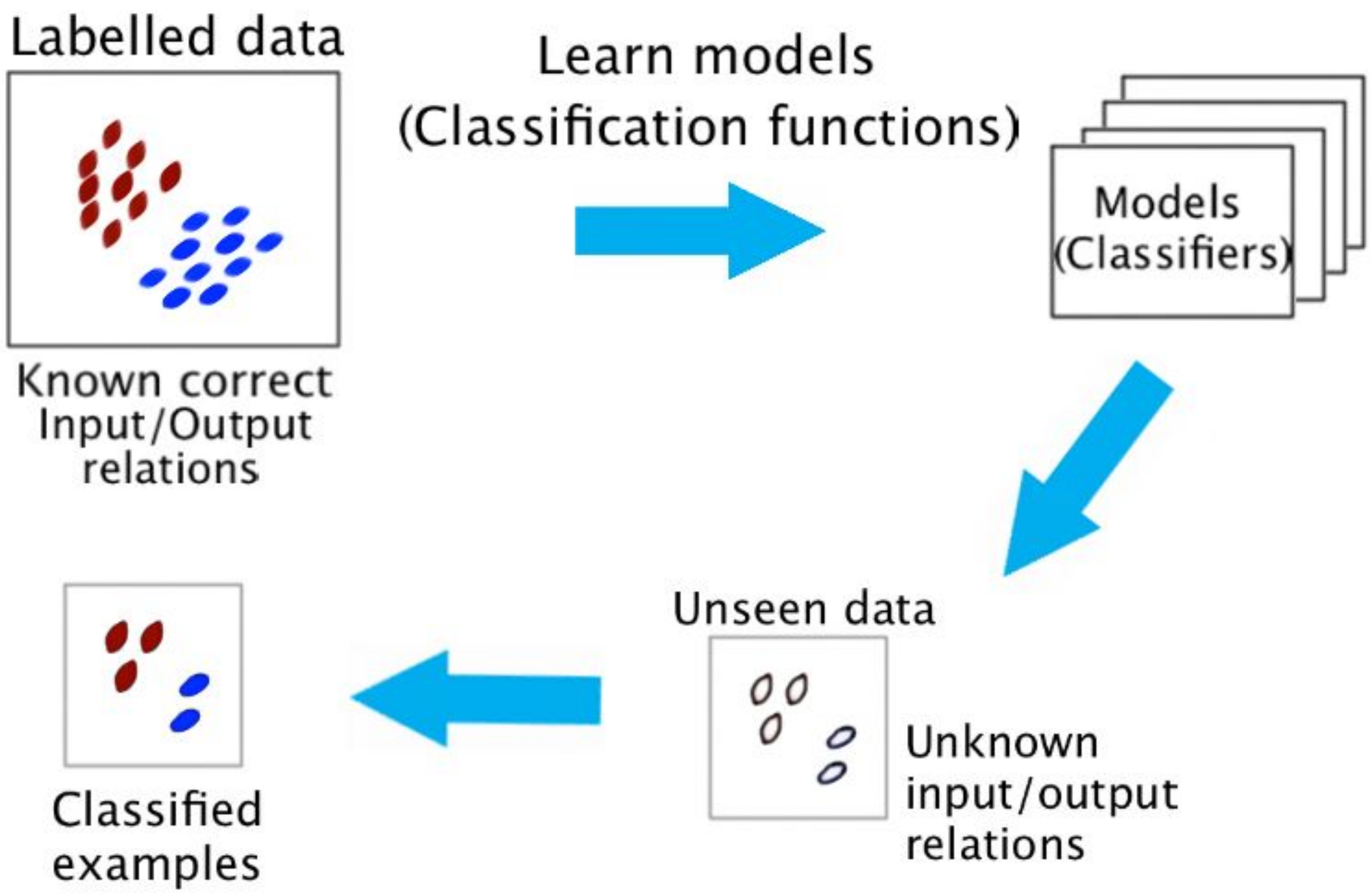
A long list of applications...

web page ranking, recommendation, automatic translation, autonomous cars, diagnostics, face recognition...

Kinds of Machine Learning

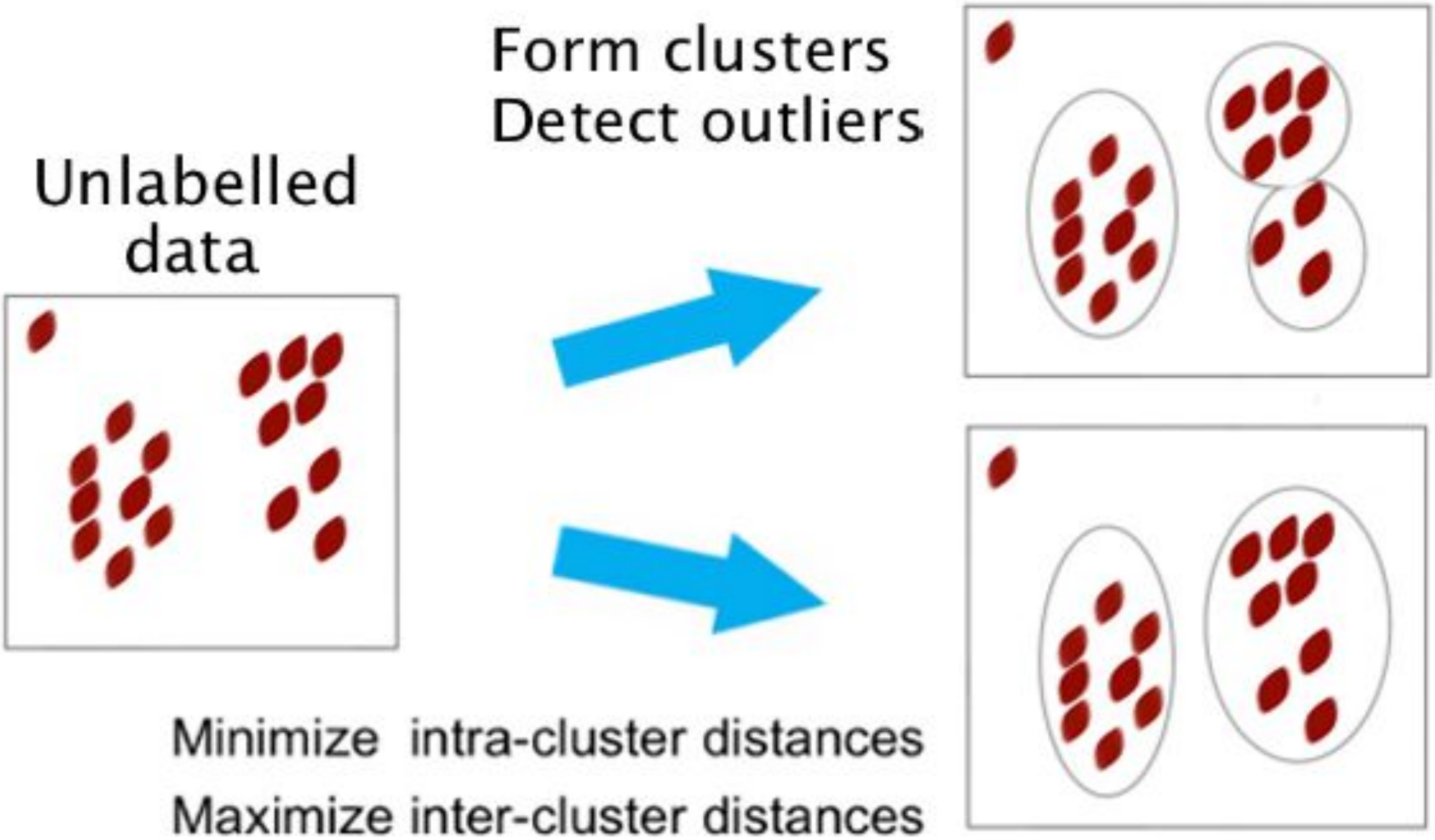
Supervised Machine Learning

Infer input/output functions from ***labelled*** data.



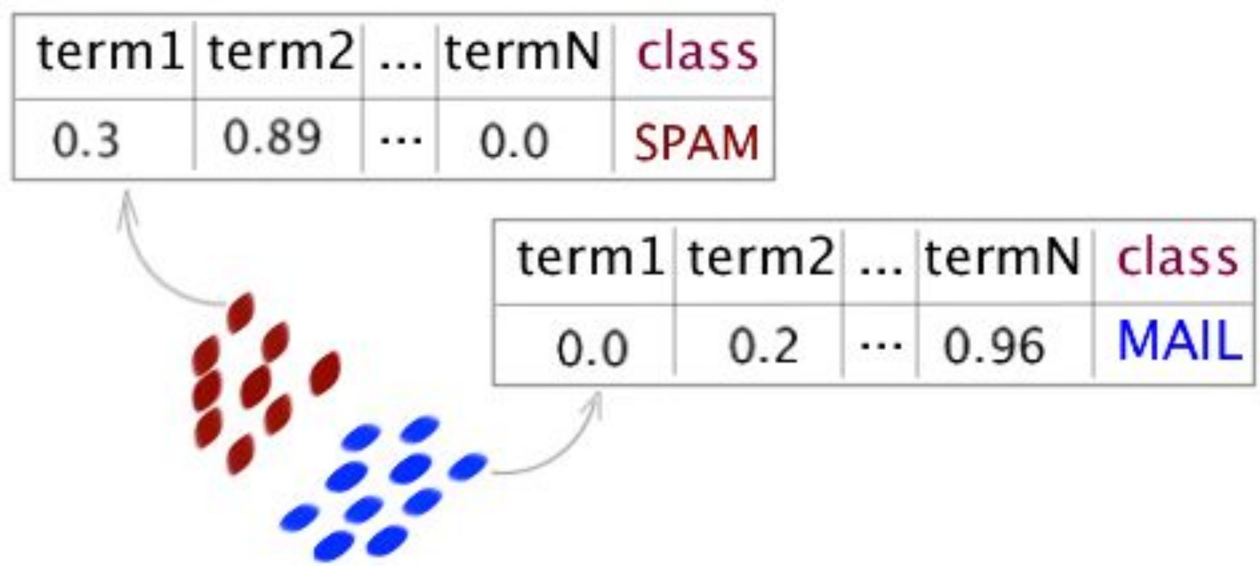
Unsupervised Machine Learning

Infer a latent structure from ***unlabelled*** data.



Data Representation: Features

Remember our spam filtering example: data-points are emails.



Model instances as ***vectors*** described by a number of ***features*** (***variables, attributes***).

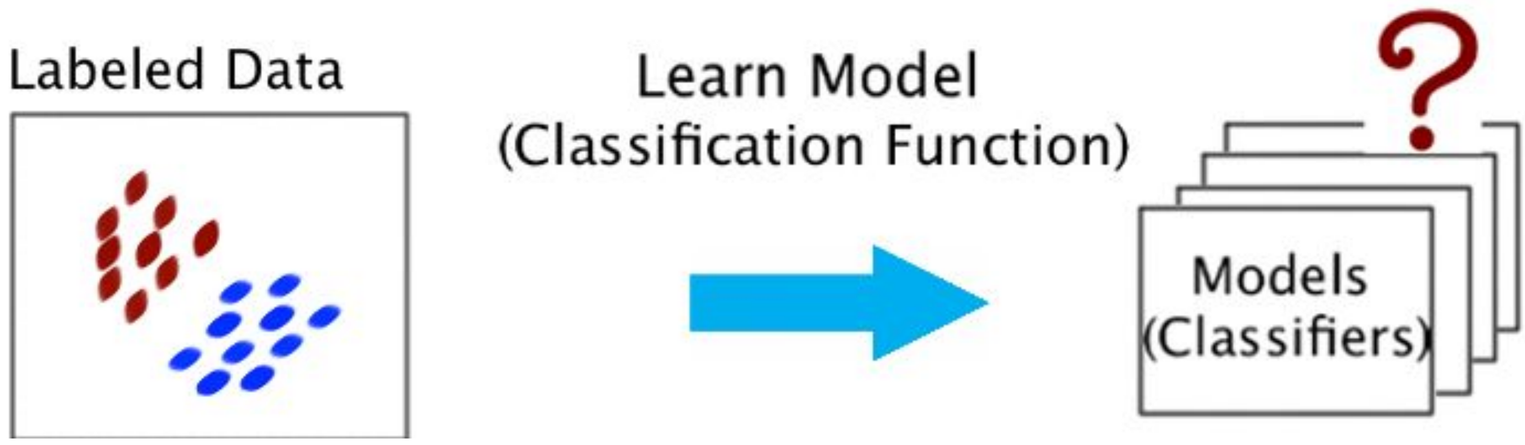
What features best describe instances and allow to separate classes or form clusters? → ***Feature (variable) selection***

- remove noisy features
- analyse their explanatory strength
- reduce dimensionality

Model Selection and Assessment

Overfitting *vs.* **Generalisation**

- how well the learned function performs on *unseen* data?
- select a model (a set of parameters) that *generalises* well
- *evaluate* and avoid overfitting



Model Selection, Model Validation

Model Validation

Confusion matrix and accuracy

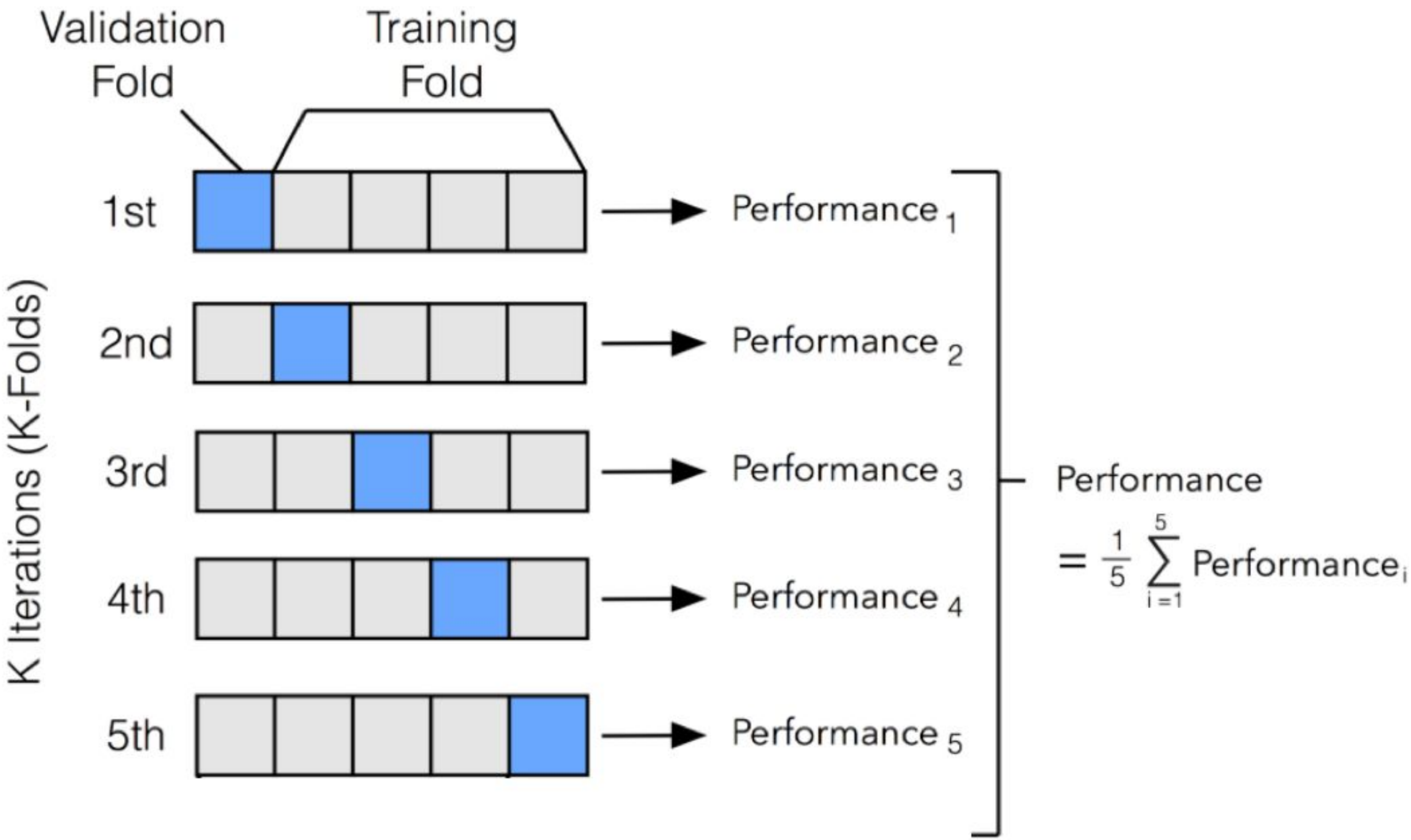
	PREDICTED CLASS		
ACTUAL CLASS		Class=Yes	Class=No
	Class=Yes	a (TP)	b (FN)
	Class=No	c (FP)	d (TN)

- Most widely-used metric:

$$\text{Accuracy} = \frac{a+d}{a+b+c+d} = \frac{TP+TN}{TP+TN+FP+FN}$$

Model Validation

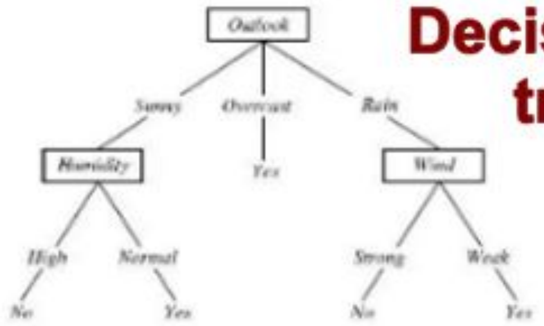
Cross-validation



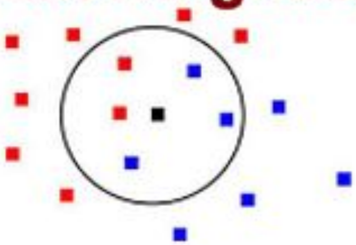
Classical Methods, Tools and Applications (Examples)

Methods and Applications

Decision trees



K Nearest Neighbours

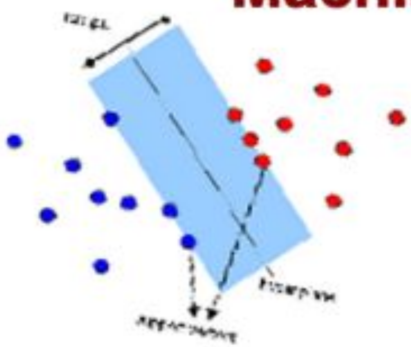


Bayesian Classifiers

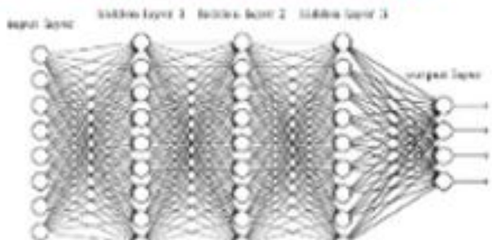
$$P(C \mid A) = \frac{P(A \mid C)P(C)}{P(A)}$$

Reinforcement Learning

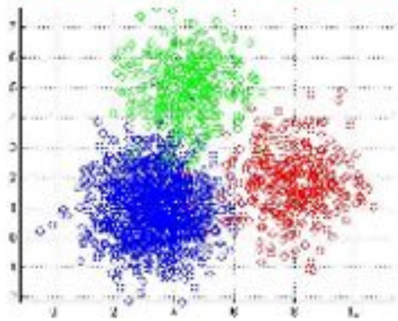
Support Vector Machines



Neural Networks / Deep Learning

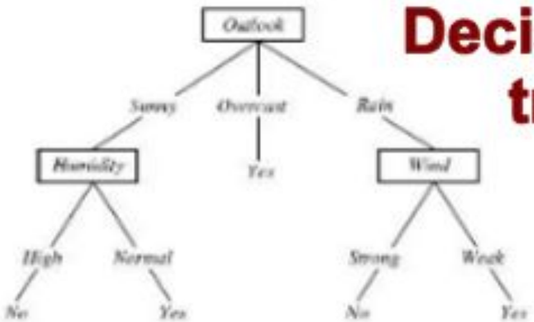


Cluster Analysis

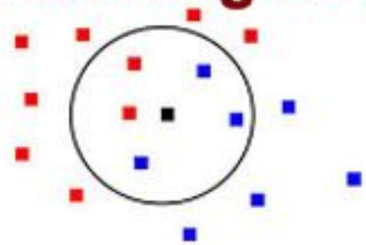


Methods and Applications

Decision trees



K Nearest Neighbours



Financial distress prediction

Process control

Fraud detection

Medical diagnosis

Robotics

Driving autonomous cars

Playing games

Trading stradegies

Bayesian Classifiers

$$P(C \mid A) = \frac{P(A \mid C)P(C)}{P(A)}$$

Text Categorization

Automatic translation

Web search

Computer vision

Image retrieval

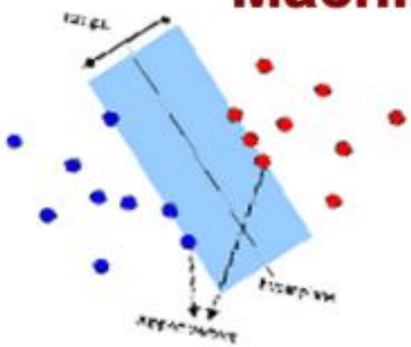
Gene clustering

Topic discovery

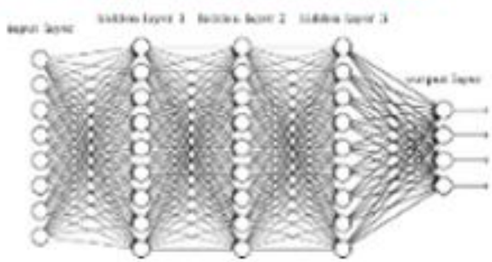
Market segmentation

Reinforcement Learning

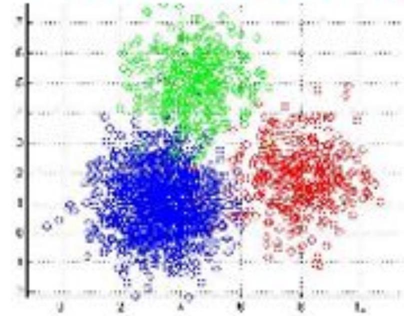
Support Vector Machines



Neural Networks / Deep Learning



Cluster Analysis

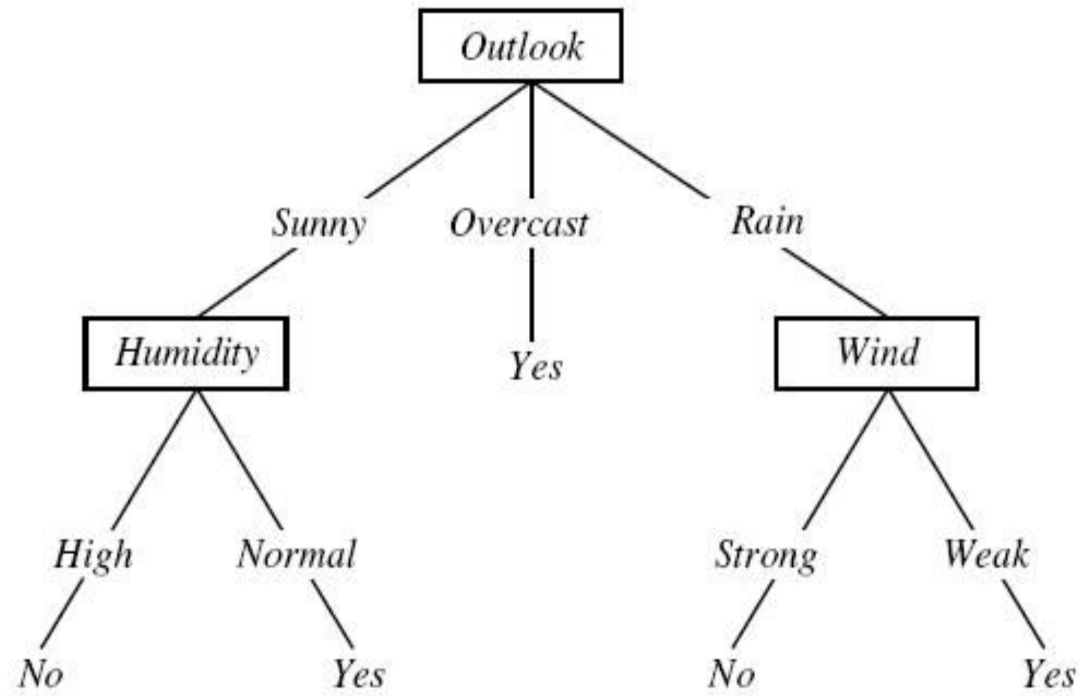


ML Methods and Applications

Decision trees

Supervised / Classification

Fits data into a tree
Attributes → nodes
Values → branches
Easy to interpret
Overfitting occurs often



From T. Mitchell's "Machine Learning"

Applications

Biomedical engineering: selecting features for implantable devices
Manufacturing, production: process control
Molecular biology: analyzing amino acid sequences
Fraud detection

ML Methods and Applications

Bayesian Classifiers

Supervised / Classification

$$p(\text{class} \mid \text{data}) = \frac{p(\text{data} \mid \text{class}) \times p(\text{class})}{p(\text{data})}$$

Creates a model per class, using probability theory.
Attributes are assumed independent.
Probabilities are estimated from data.

Applications

Text categorisation
Speech recognition
Automatic medical diagnosis

ChatGPT

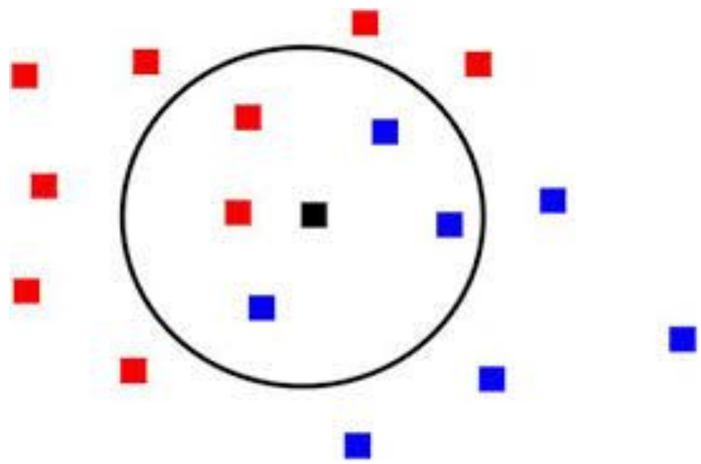
Les classificateurs bayésiens sont des algorithmes de machine learning qui utilisent des probabilités pour classer des données dans des catégories. Ils se basent sur le théorème de Bayes pour calculer la probabilité qu'une donnée appartienne à une catégorie spécifique en fonction des caractéristiques de cette donnée. Le classificateur choisit ensuite la catégorie avec la probabilité la plus élevée comme prédiction.

ML Methods and Applications

K Nearest Neighbours

Supervised / Classification

Lazy instance-based learners.
Uses distance calculation over all instance pairs.



Applications

Cancer diagnosis
Financial distress prediction
Computer vision

ChatGPT

K Nearest Neighbors (K-NN) est un algorithme de machine learning utilisé pour la classification et la régression. Il se base sur le principe suivant : pour classer ou prédire une nouvelle donnée, regardez les "voisins" les plus proches de cette donnée parmi les données d'entraînement.

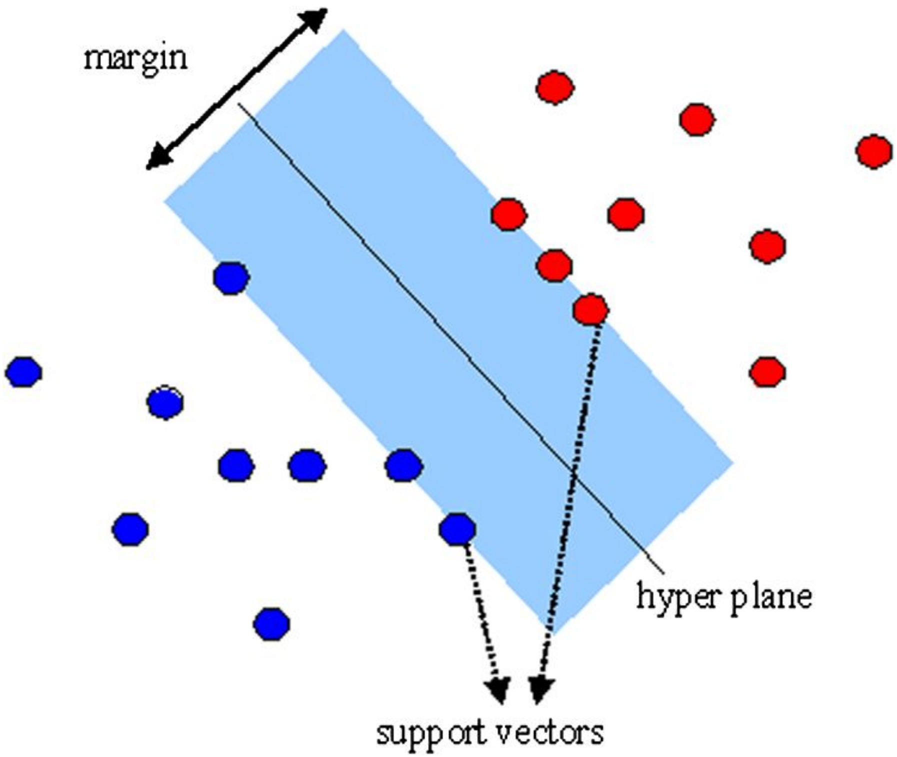
Pour classer une nouvelle donnée, K-NN identifie les K exemples les plus proches dans l'ensemble d'entraînement. Ensuite, il compte combien de ces exemples appartiennent à chaque classe. La classe majoritaire parmi ces voisins devient la prédiction pour la nouvelle donnée.

ML Methods and Applications

Support Vector Machines

Supervised / Classification

Learns a maximum margin separating hyperplane.
Deals with non-linearly separable data
Uses kernels



Applications

Text categorisation
Automatic translation
Computer vision
Handwriting / face / facial expression recognition
Content-based image retrieval

L'objectif principal d'un SVM est de trouver la meilleure séparation (ou hyperplan) entre différentes classes de données. Cette séparation est choisie de manière à maximiser la marge, c'est-à-dire la distance entre l'hyperplan et les points de données les plus proches de chaque classe. Ces points de données les plus proches sont appelés "vecteurs de support", d'où le nom "Support Vector Machines".

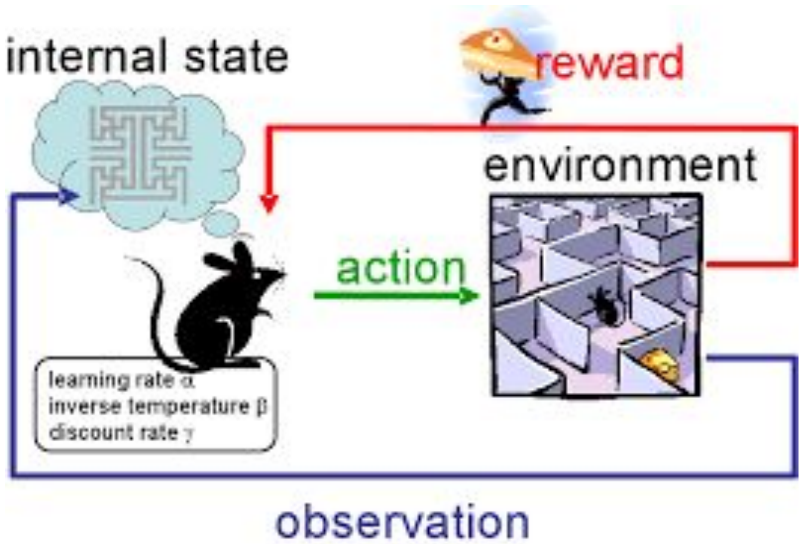
Une fois que l'hyperplan est trouvé, il peut être utilisé pour classer de nouvelles données en les plaçant de part et d'autre de l'hyperplan. Si une nouvelle donnée se trouve du côté de la classe correspondante à l'hyperplan, elle est classée dans cette classe.

ML Methods and Applications

Reinforcement Learning

Unsupervised or Semi-supervised

Take actions according to rewards.
Behaviour optimisation with respect to the environment.



Applications

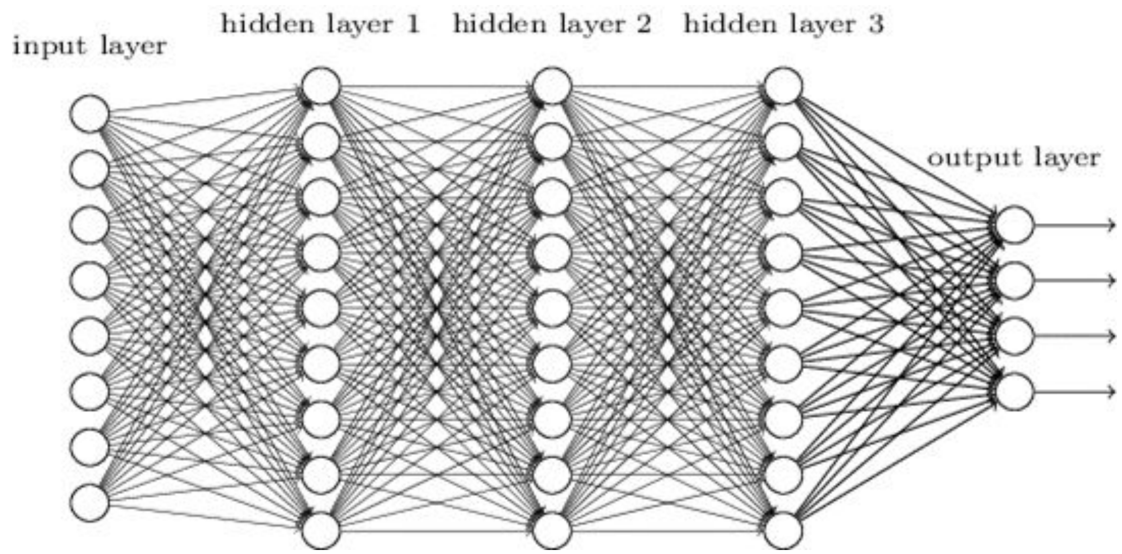
Driving autonomous vehicles
Robot vision
Playing games

Agent : L'agent est l'entité qui prend des actions dans un environnement pour atteindre un objectif spécifique.
Environnement : L'environnement représente le monde dans lequel l'agent opère. Il peut être réel, virtuel ou simulé.
Actions : L'agent prend des actions pour interagir avec l'environnement. Chaque action a un impact sur l'état actuel de l'environnement.
États : Les états décrivent la situation actuelle de l'environnement, ce qui permet à l'agent de prendre des décisions en fonction de ce qu'il perçoit.
Récompenses : L'environnement fournit des récompenses à l'agent en fonction de ses actions. L'agent cherche à maximiser la récompense cumulative au fil du temps.
Politique : La politique est une stratégie qui guide l'agent dans le choix de ses actions en fonction des états. L'objectif est de trouver une politique optimale qui maximise les récompenses.
Apprentissage : L'agent apprend à améliorer sa politique au fil du temps en explorant différentes actions et en évaluant leurs conséquences sur les récompenses. Cela se fait souvent par le biais de techniques d'exploration et d'exploitation.

ML Methods and Applications

Neural Networks / Deep Learning

Supervised and Unsupervised



Bio-inspired: a complex net of interconnected neurones

Applications

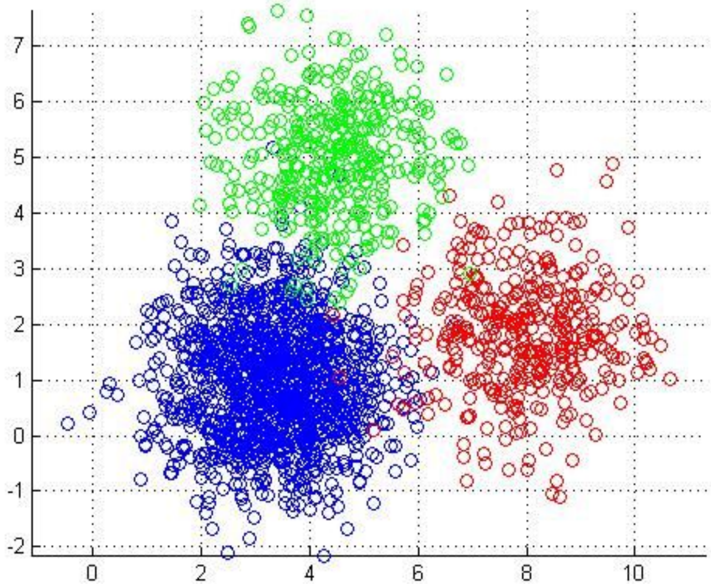
Driving autonomous vehicles
Computer vision
Speech / face / handwriting recognition
Sensor data interpretation
Image retrieval
Text and language models

ML Methods and Applications

Cluster Analysis

Unsupervised / Clustering

Group together instances into subsets
Maximise intra-cluster instance
similarities and inter-cluster distances.
    K-means, DBSCAN,
    Descriptive Statistics, ...



Applications

Market segmentation
Gene clustering
News summarisation
Topic discovery

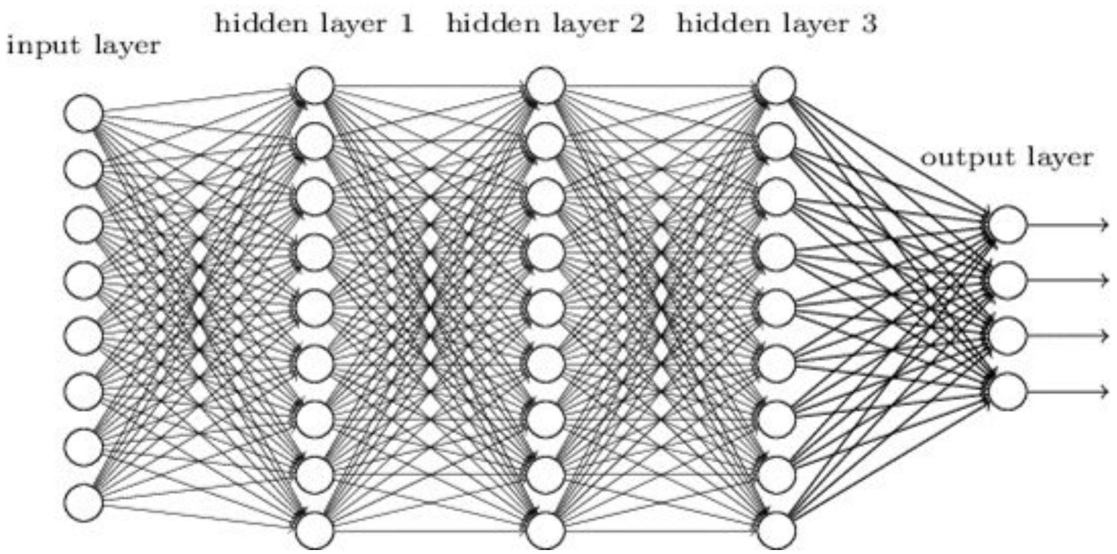
Deep neural networks and representation learning

Deep Learning?

Used in many domains: language models, computer vision, robotics, natural language processing (NLP), music, arts

For example, in NLP and speech:

- Sentiment Analysis
- Machine Translation (Google Translate)
- Question Answering Systems (Bot)
- Language models to generate Text (GPT-3)
- Speech Recognition - automatic subtitling (Youtube)



Deep Learning

Traditional machine learning: heavy feature engineering to represent data

- Text Analysis: Bag of Words
- Image Analysis: Hog (Histogram of Oriented gradient), SIFT (Scale Invariant Feature Transform)

Hand-Crafted Features → **Simple Trainable Classifier (SVM, RF, NB, …)**

Deep learning: no need for hand-crafted features

Trainable Feature Extractor → **Trainable Classifier**

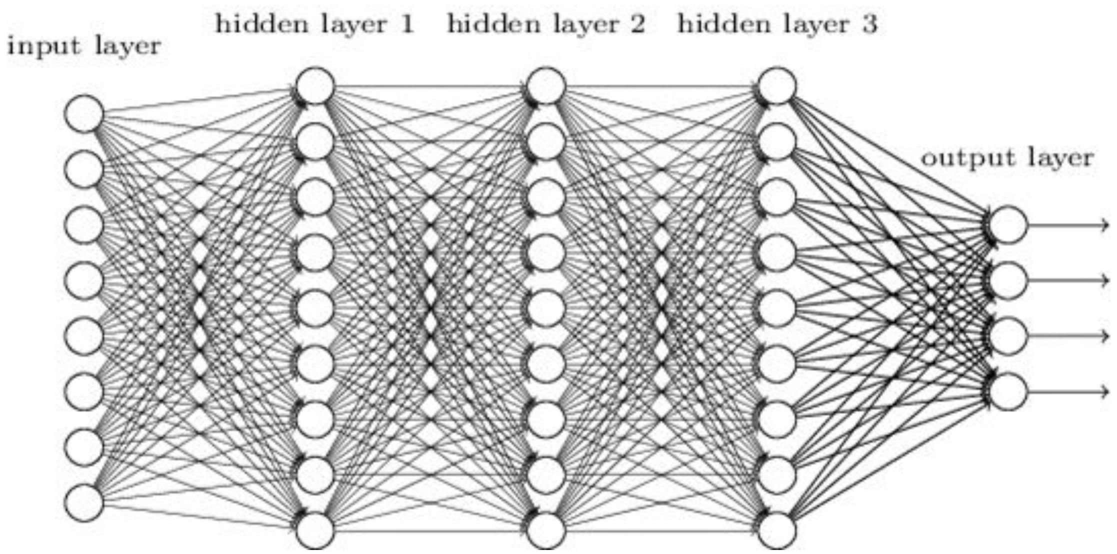
Deep Learning Model

Deep Learning / Representation Learning

⇒ Learn automatically features / vector representations
⇒ Perform various prediction / classification tasks
⇒ Do both! Start doing machine ***learning***, instead of 80% manual feature design and selection

A class of machine learning techniques

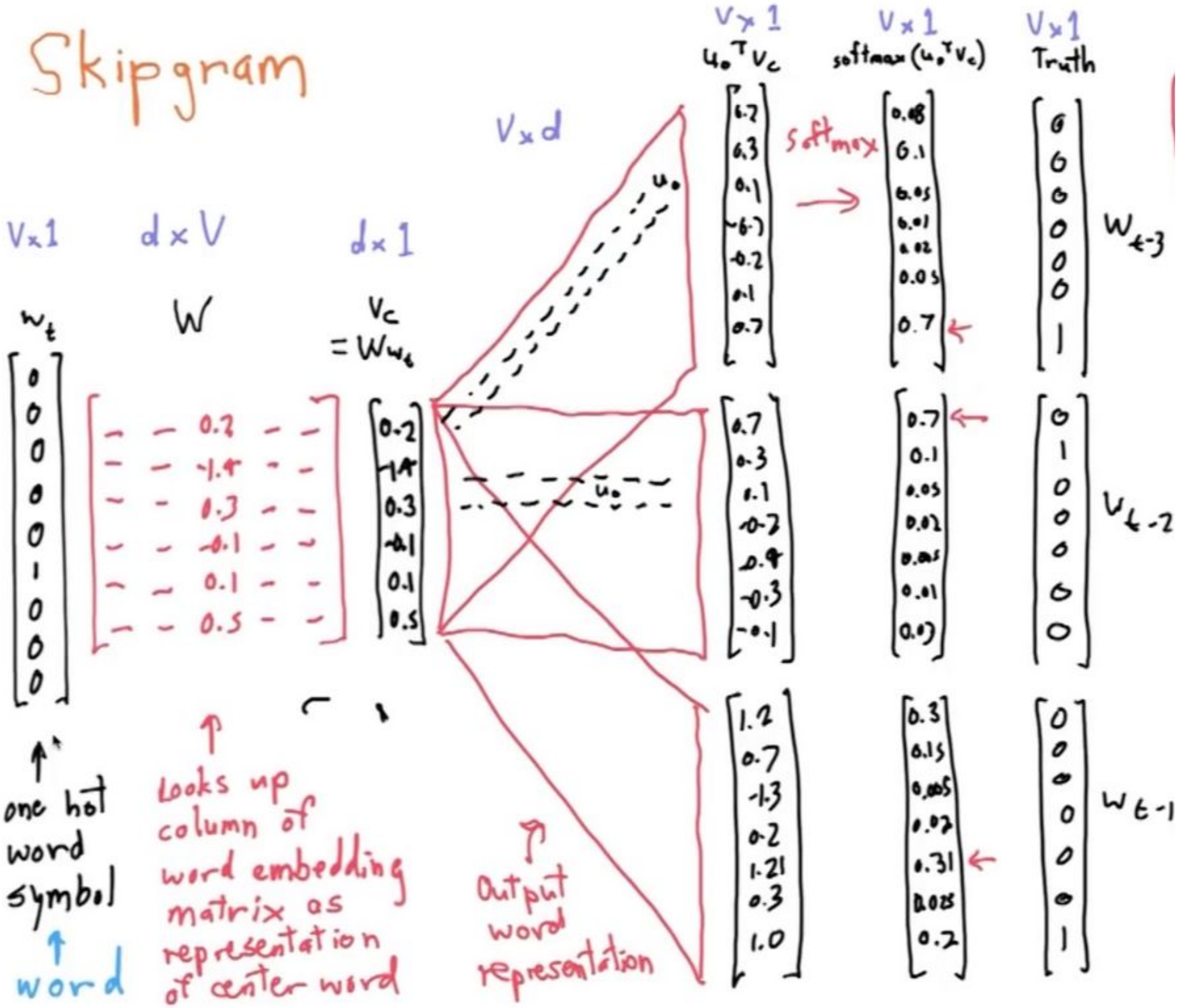
Exploit many layers of (non-)linear information processing for supervised or unsupervised feature extraction and transformation and pattern analysis and classification.



Most commonly, based on neural networks with several hidden layers.

Representation Learning

- A set of methods that allow a machine to be fed with raw data and to automatically discover the needed <u>vector representation</u> for detection, classification or prediction tasks.



Distributional semantics

- Look at the neighbourhood (context) of a word in ***many*** different documents (large corpora).

“You shall know a word by the company it keeps”

(J. R. Firth 1957: 11)

One of the most successful ideas of modern statistical NLP

government debt problems turning into	banking	crises as has happened in
saying that Europe needs unified	banking	regulation to replace the hodgepodge

banking

Predict the textual context == understand the meaning of a word.

Word meaning defined by vectors

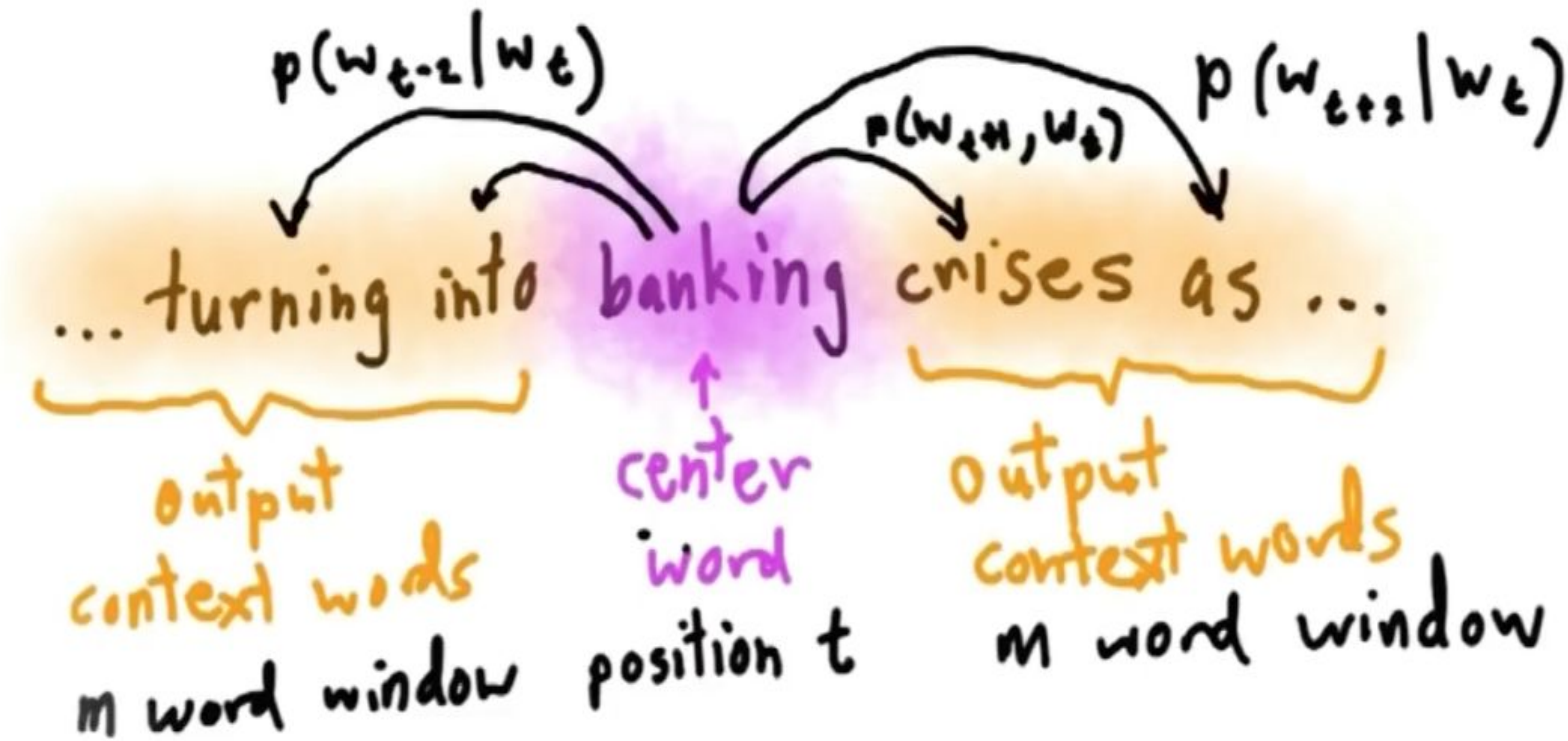
We will build a dense vector for each word type, chosen so that it is good at predicting other words appearing in its context

... those other words also being represented by vectors ... it all gets a bit recursive

$$
\textit{linguistics} = \begin{pmatrix} 0.286 \\ 0.792 \\ -0.177 \\ -0.107 \\ 0.109 \\ -0.542 \\ 0.349 \\ 0.271 \end{pmatrix}
$$

Skip-gram prediction model

Goal: → Choose vector representations of words that maximize the probability distributions of context words.

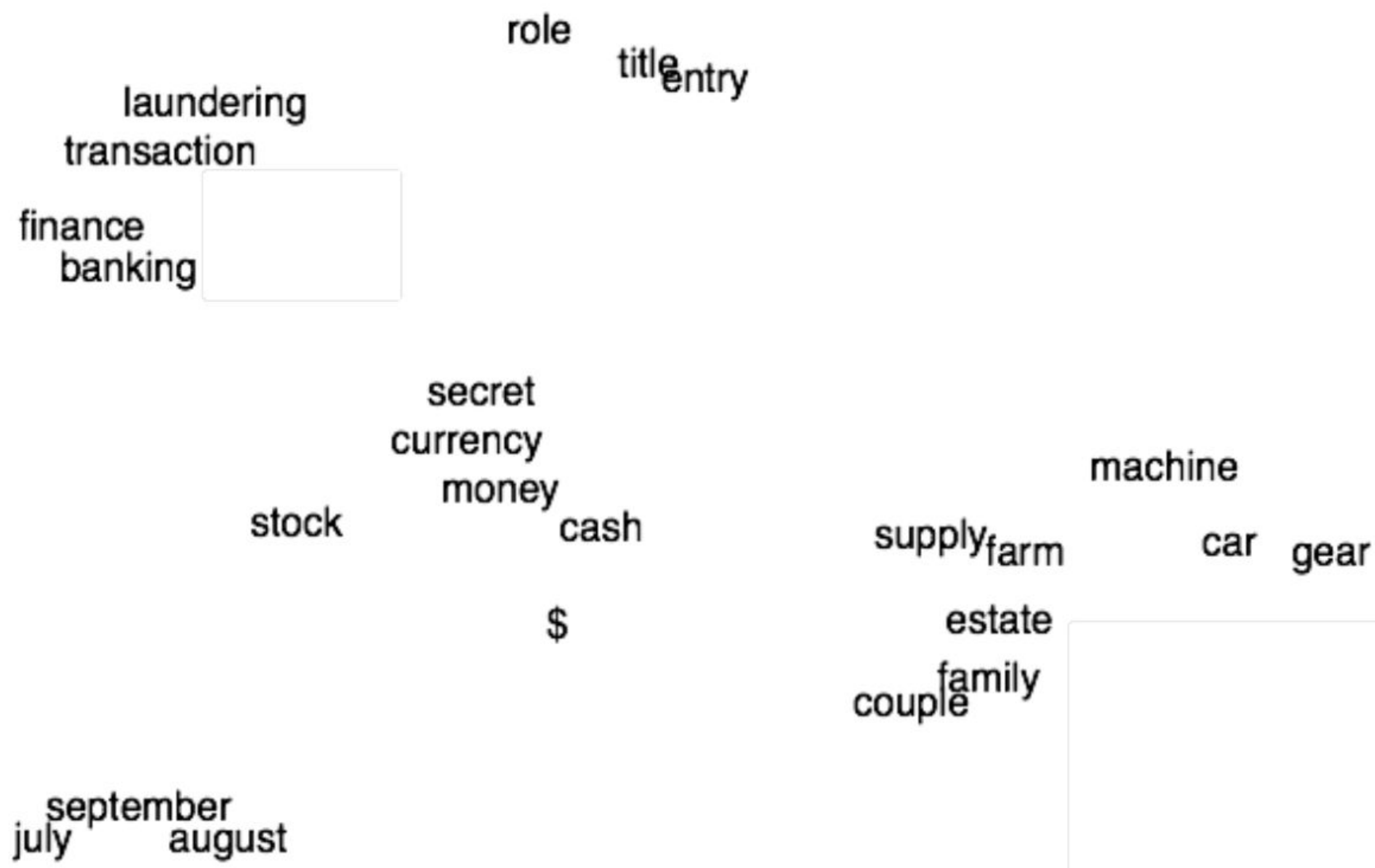


Assumption:
→ there is 1 probability distribution (all words follow the same law)

Word meaning in a vector

The result: → close in meaning words are represented by close in a vector space points.

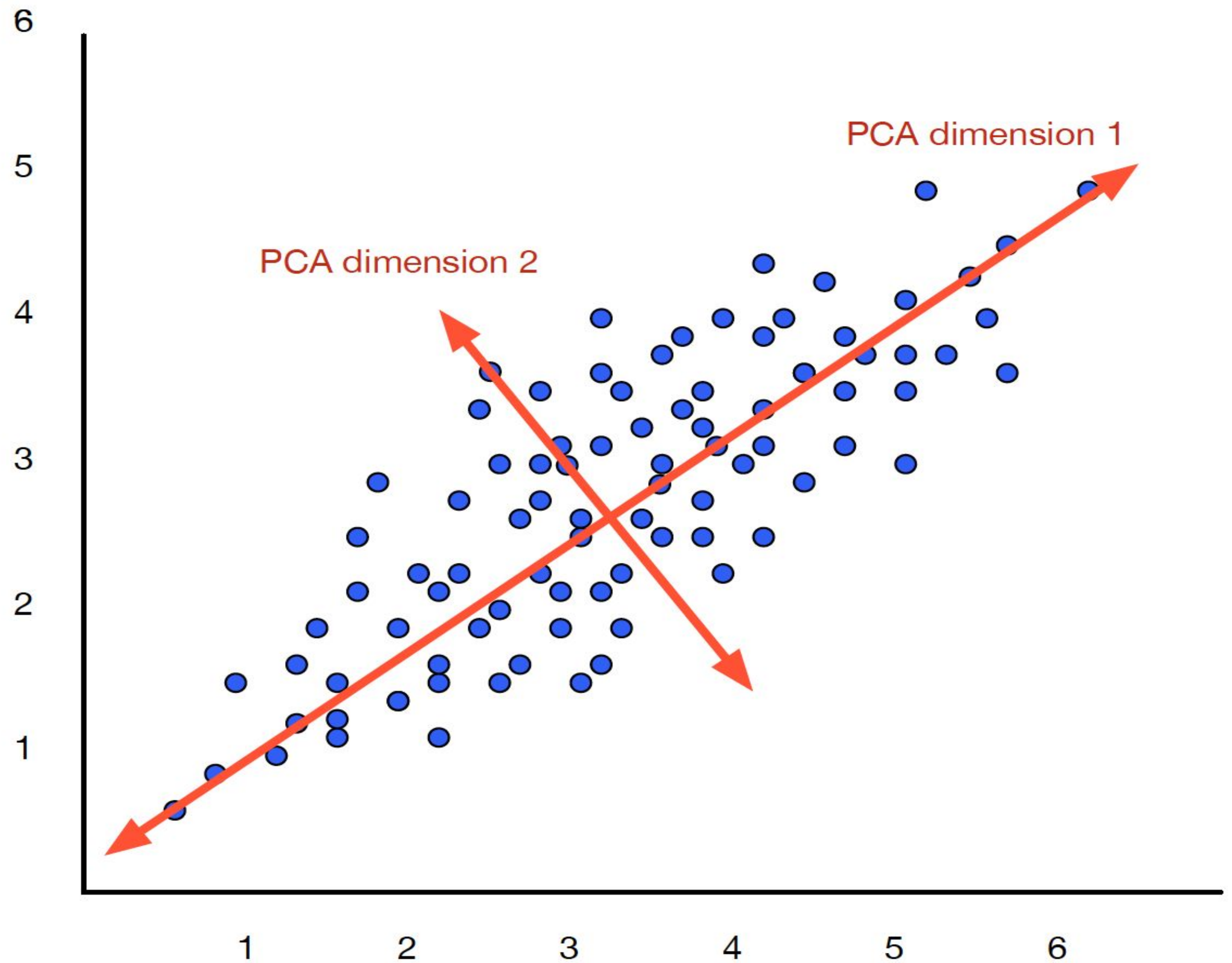
$$\textit{currency} = \begin{pmatrix} 0.286 \\ 0.792 \\ -0.177 \\ -0.107 \\ 0.109 \\ -0.542 \\ 0.349 \\ 0.271 \end{pmatrix}$$



Vectors of high dimensions

Difficult to interpret…

→ Dimensionality reduction techniques (e.g. PCA)

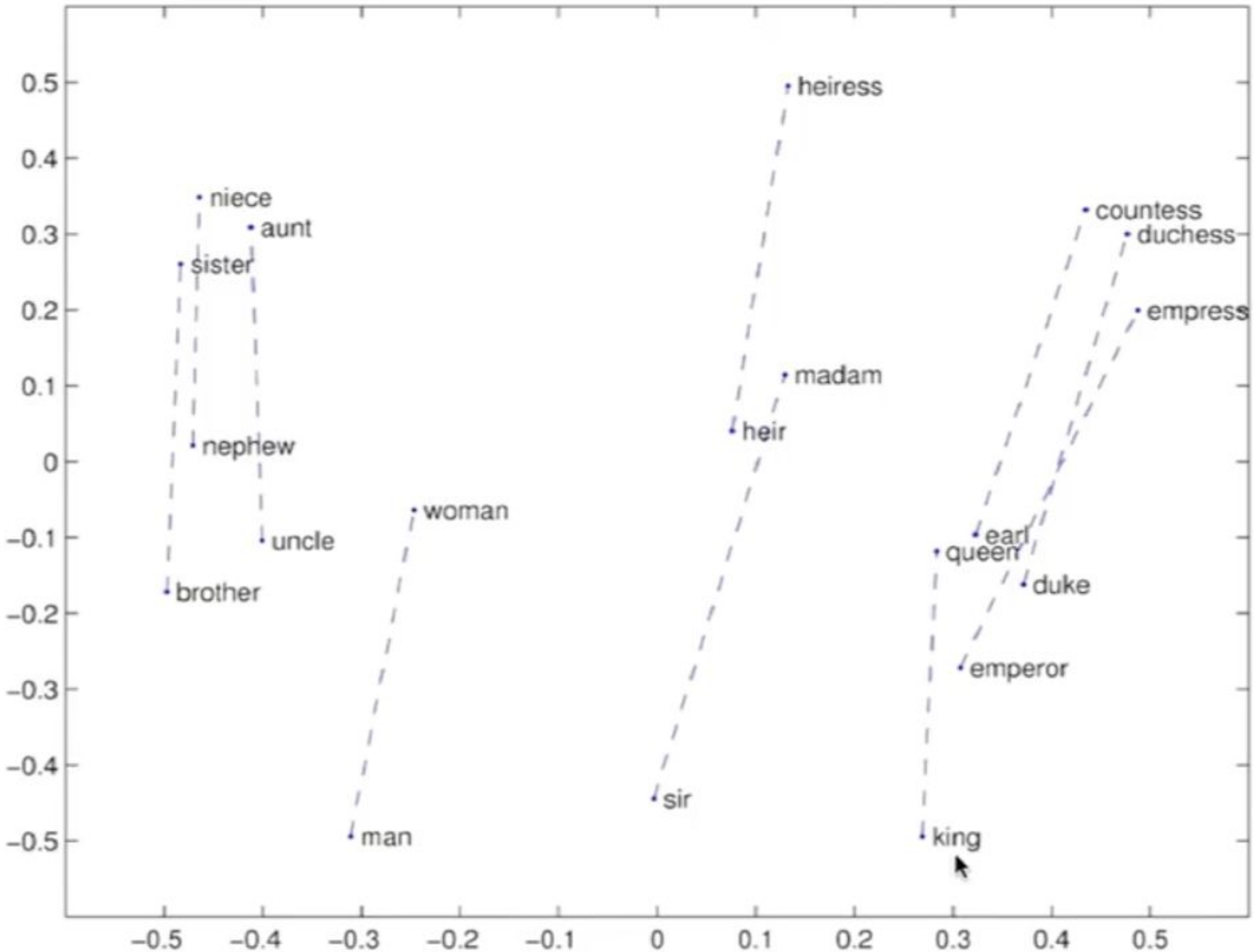


Embeddings are astonishing

Some interesting outcomes... for example, analogies

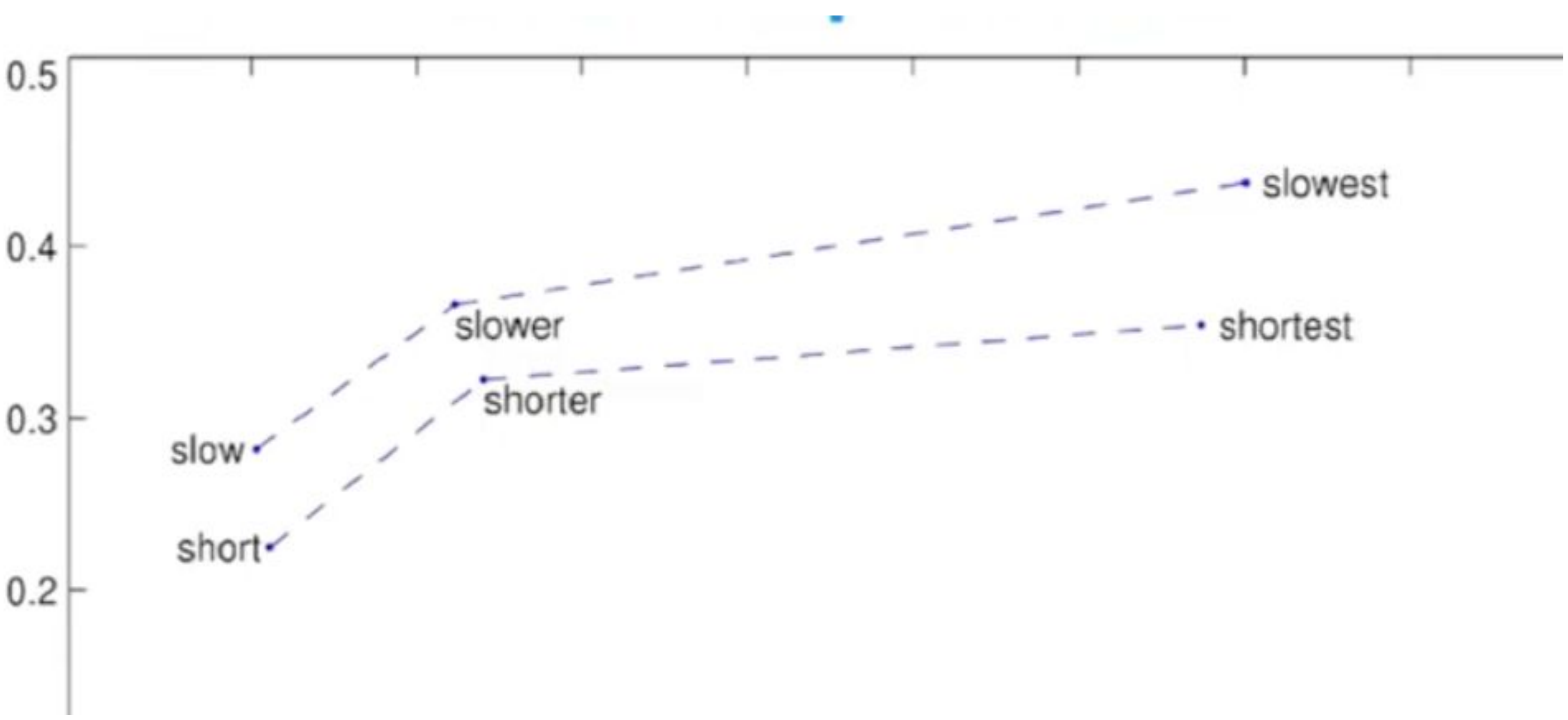
man:woman :: king : x?

vec(man) - vec(woman) + vec(king) = ???



More analogies

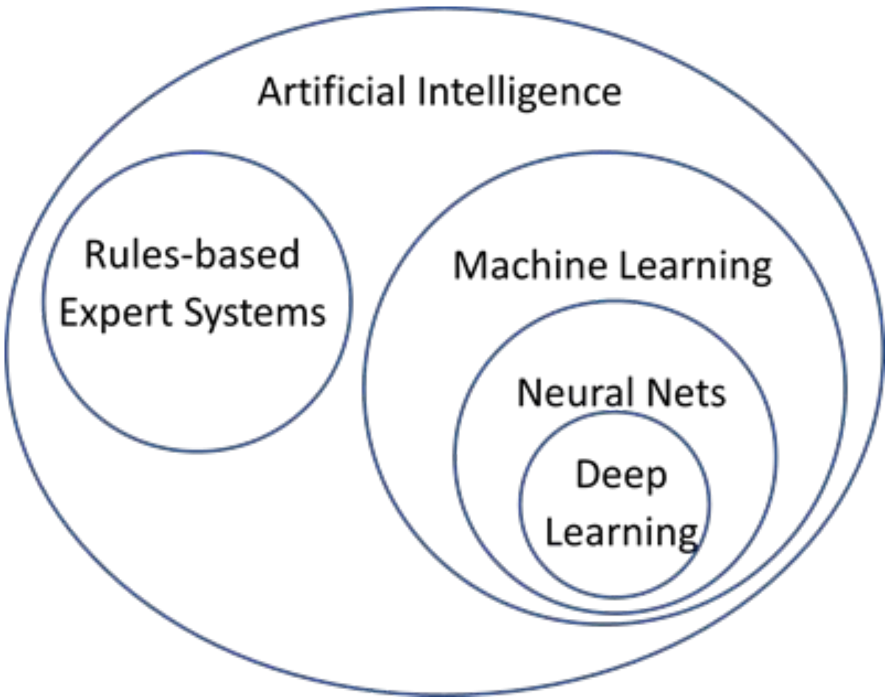
Paris – France + Italy -----> Rome
bigger – big + cold ------> colder
sushi – Japan + Germany ----> bratwurst



THIS IS YOUR MACHINE LEARNING SYSTEM?
YUP! YOU POUR THE DATA INTO THIS BIG PILE OF LINEAR ALGEBRA, THEN COLLECT THE ANSWERS ON THE OTHER SIDE.
WHAT IF THE ANSWERS ARE WRONG?
JUST STIR THE PILE UNTIL THEY START LOOKING RIGHT.
DATA
ANSWERS

Hybrid AI: Combining Knowledge Representation with ML

Hybrid AI?



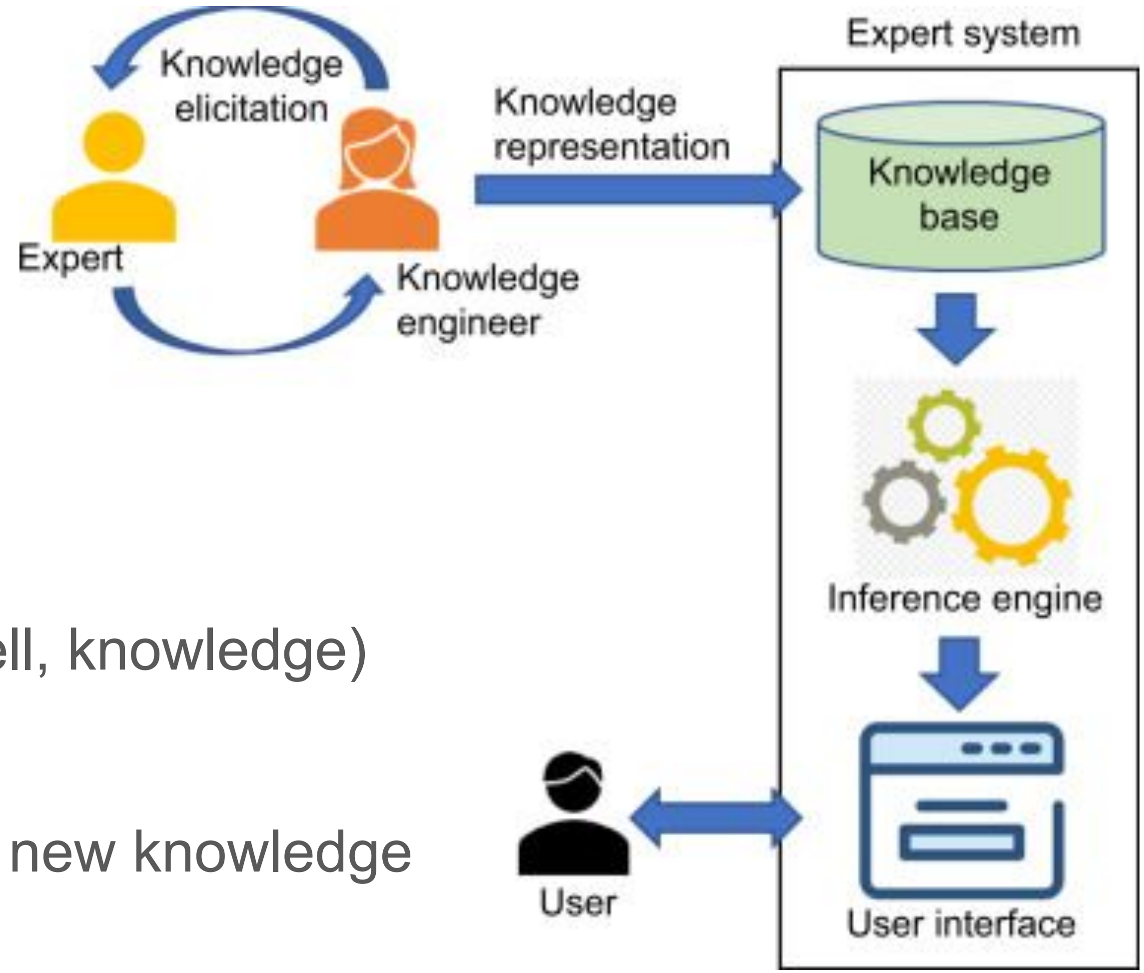
- In the 70s-80s: neural networks are failing
  ⇒ the future belongs to symbolic approaches!

- In the early 2000s: symbolic approaches don't seem to work
  ⇒ long live the deep neural networks!

How about taking the best of combining machine learning with symbolic approaches?

Expert systems

Mimic the decision making abilities of humans in a specific field (medicine, finance, etc.)



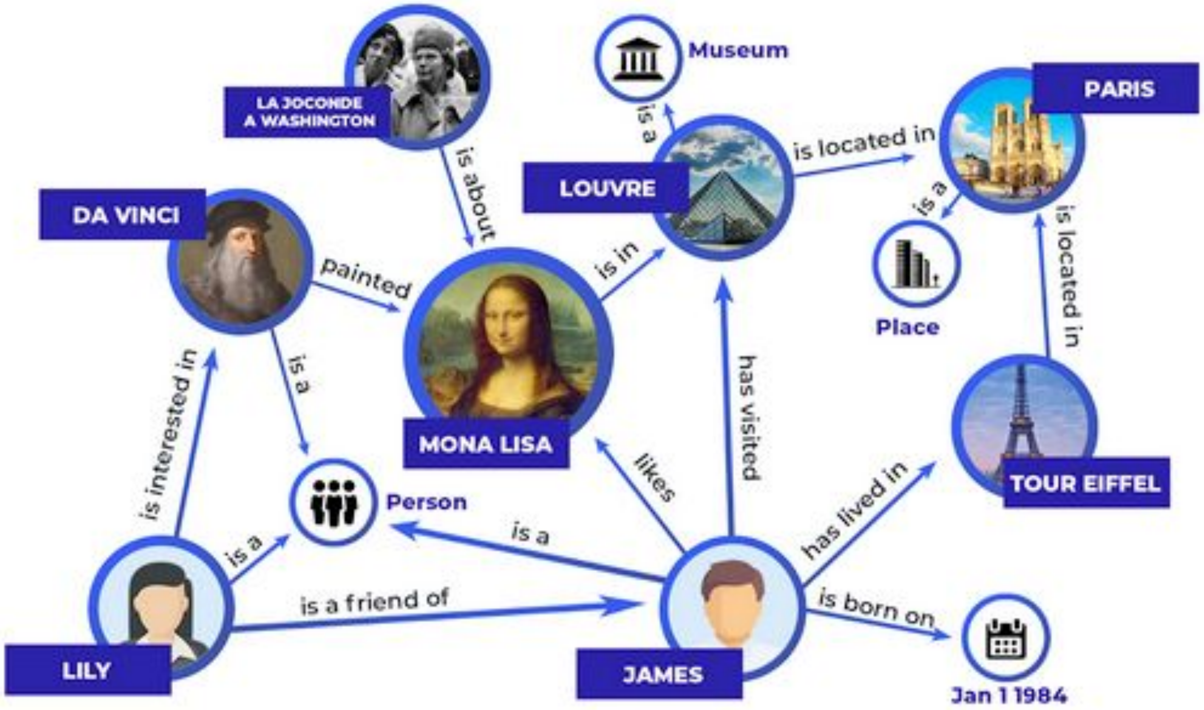
A knowledge base of facts (well, knowledge) about the specific field

A reasoning engine that infers new knowledge or shapes a decision

Knowledge Bases / Knowledge Graphs

Still in the focus today

- growing importance
- a compelling abstraction for organizing world's knowledge over the internet
- a way to integrate information extracted from multiple data sources

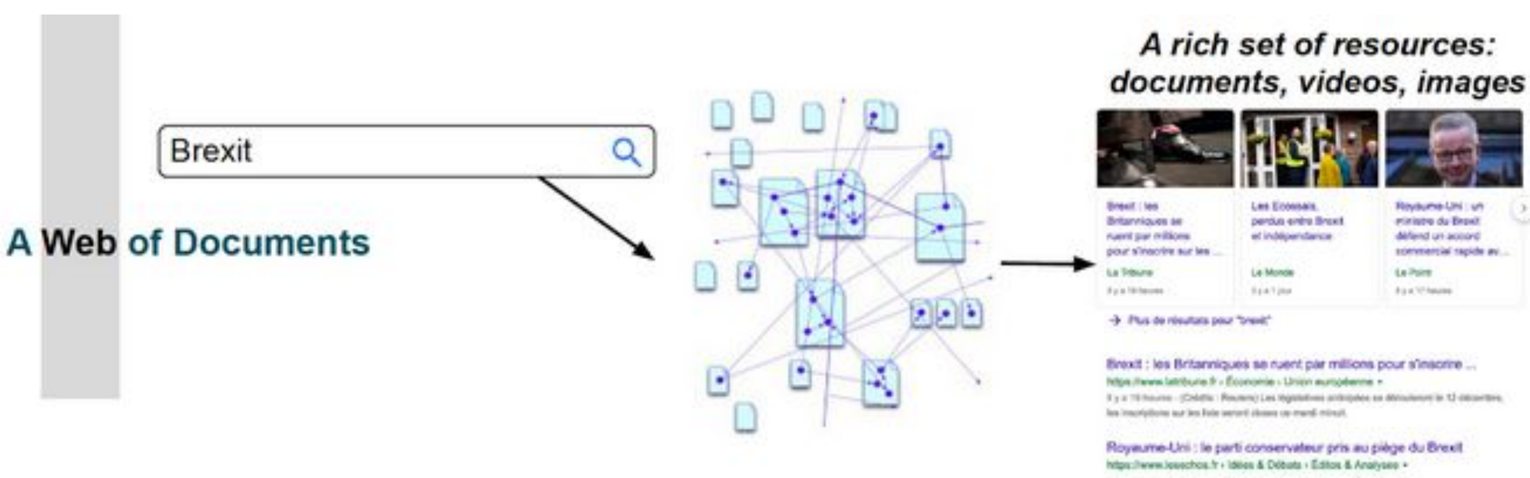


- relations between entities
- play a central role in **machine learning** as a method to incorporate world knowledge and for explaining what is learned

High level human structured and curated semantics
Precious sources of knowledge for machines and algorithms ***and*** humans

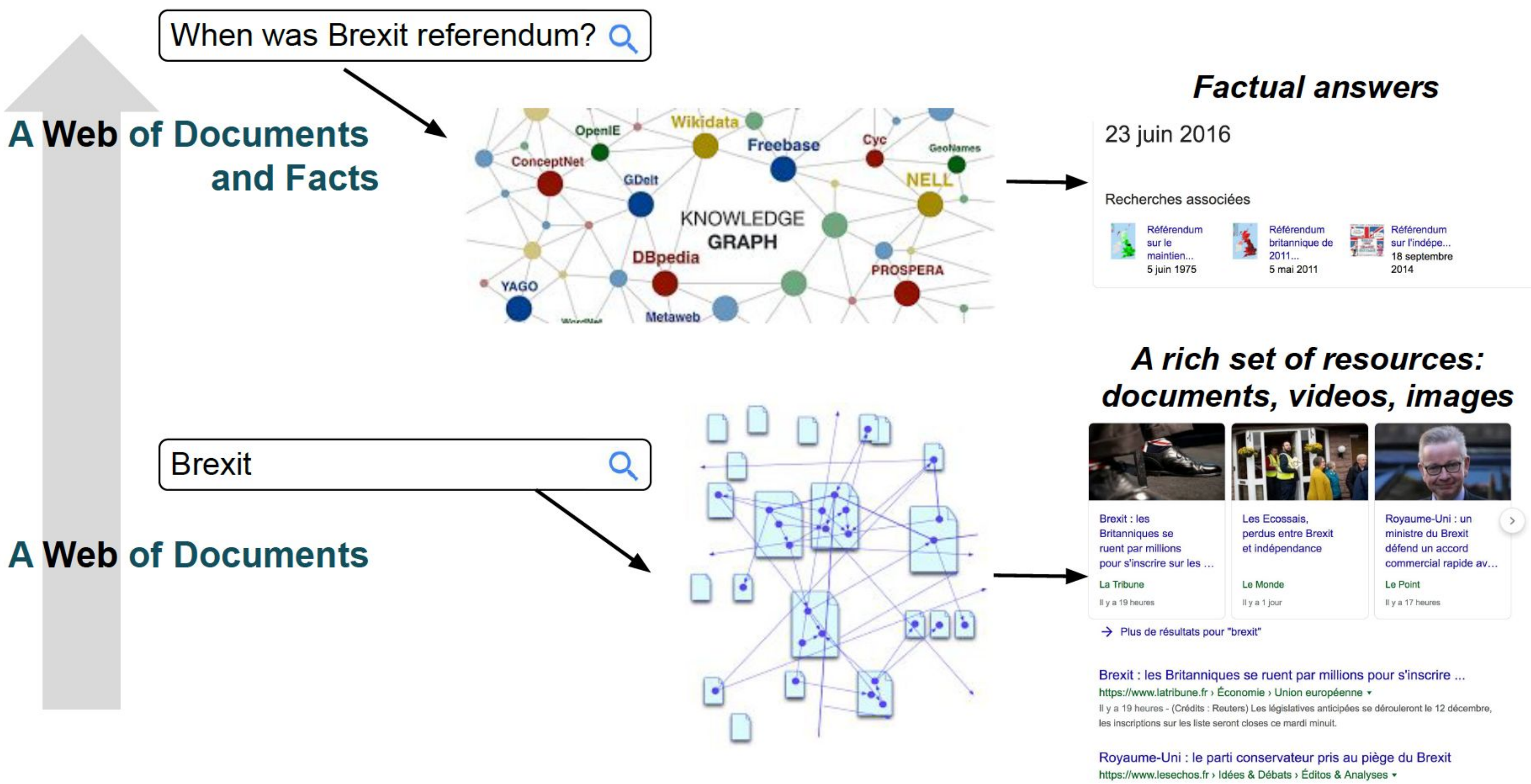
The web and hybrid AI?

From a web of documents towards a web of structured knowledge and facts



A web of structured knowledge ***and*** facts

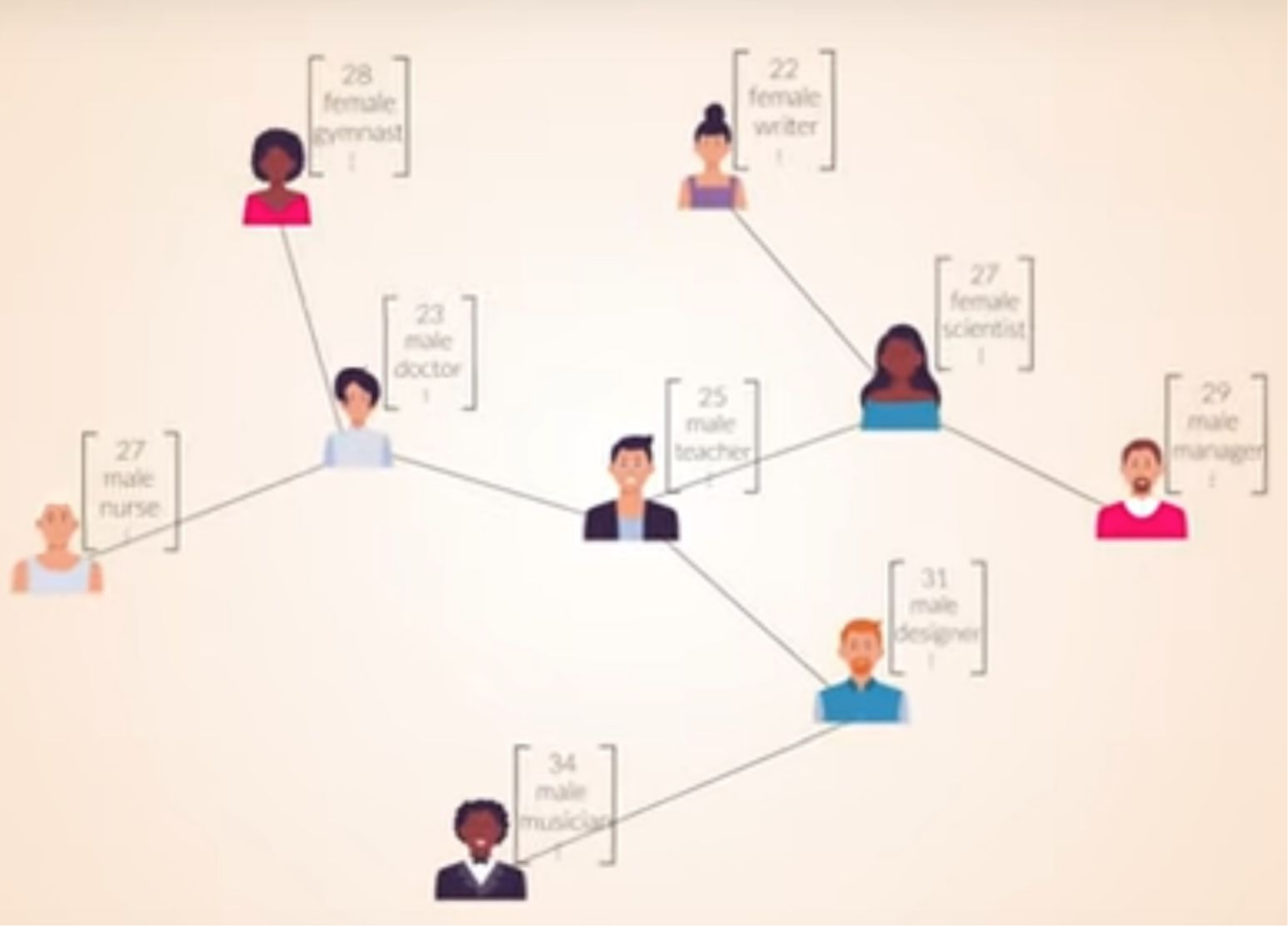
A paradigm shift: from keywords-based to ***entity-centric search***



A growing effort in ***building and publishing structured data*** on the Web

Graph Neural Networks

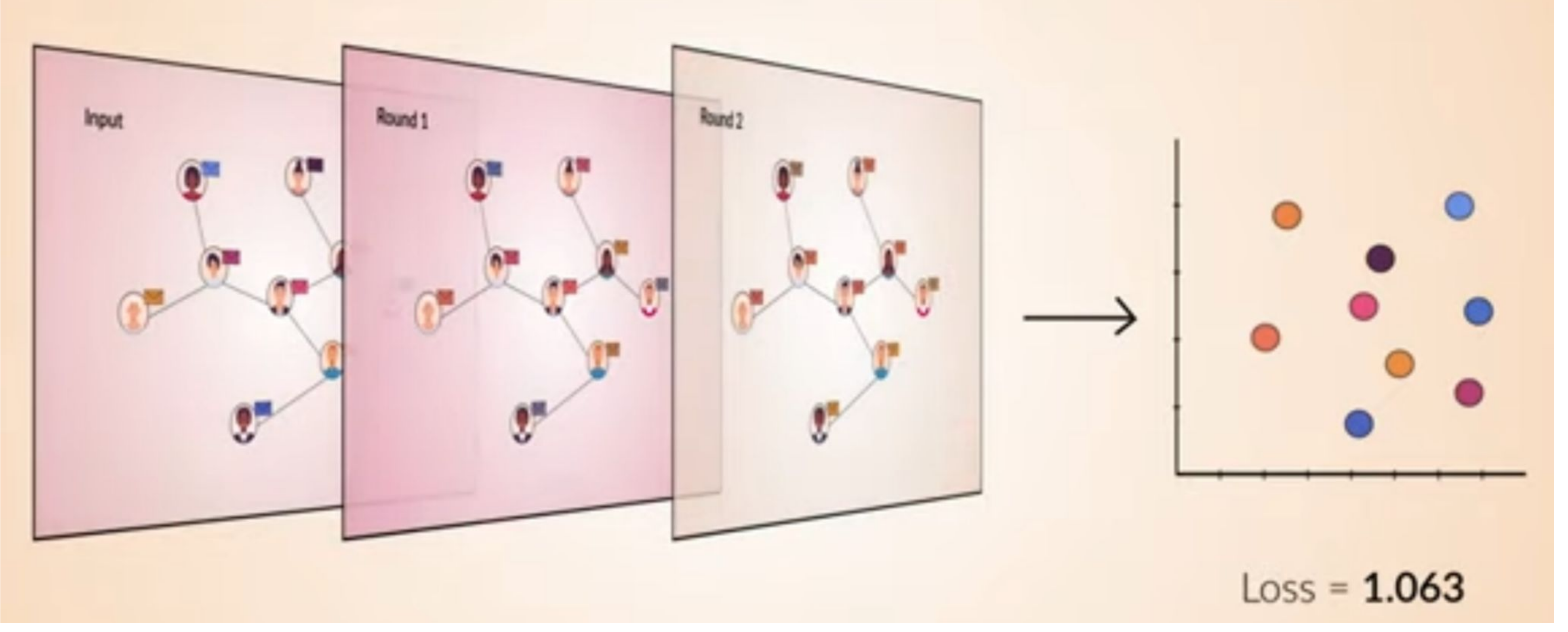
Discover automatically new drugs, predict gene adaptation to new diseases, improve transportation, model and predict social network behaviour, study online discourse…



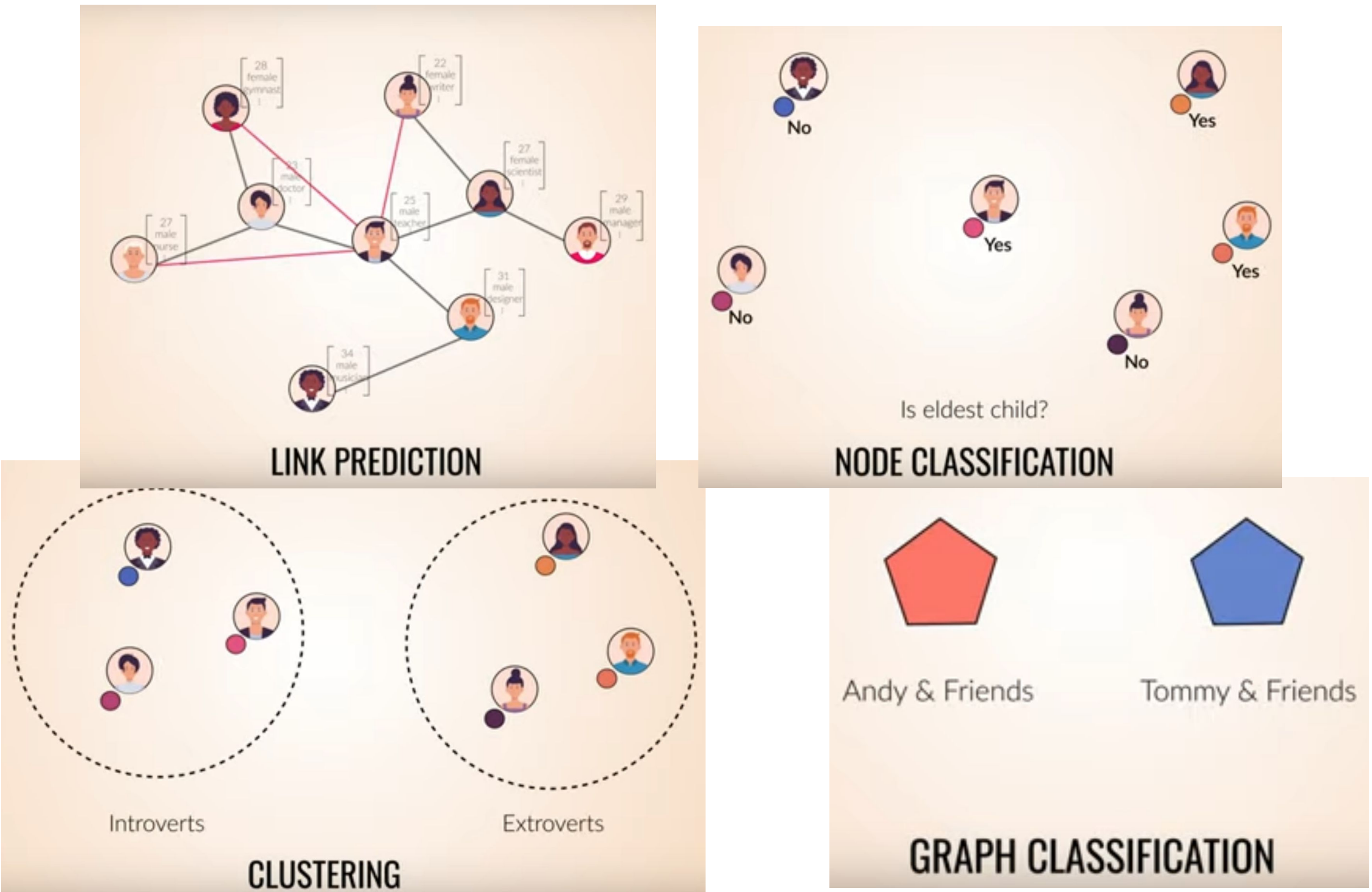
Graph Neural Networks

Represent nodes and relations as **vectors** - ***graph embeddings***.
Learning these vectors by requiring them to have certain properties

- *e.g. place more "similar" nodes closer in the embedding space*



Applying Graph Embeddings



Tools for Doing ML

Tools and Their Usefulness

A long list of open source tools and software...

Scikit-learn, Torch, Keras, Weka, R, RapidMiner*,...

Often balck boxes for users.

→ How to implement a given ML solution (which API)?

Algorithms don't change from one API to another...

→ What and how much data is needed? How to select a model? Which method for what problem?

An empirical science... with some heuristics.

→ How deep an understanding of the algorithms is required?

Investing in statistical inference:

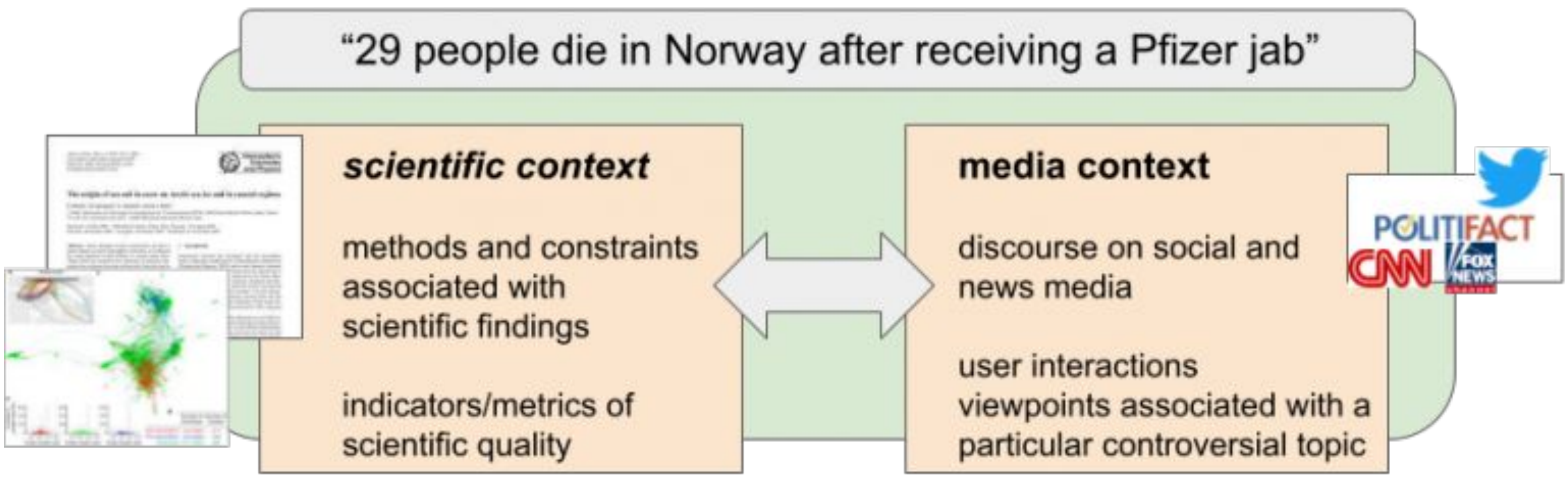
– hiring a statistician / data scientist, training engineers

ML and Hybrid AI projects at LIRMM

AI4Sci: A Hybrid AI Approach for Interpretation of Scientific Online Discourse

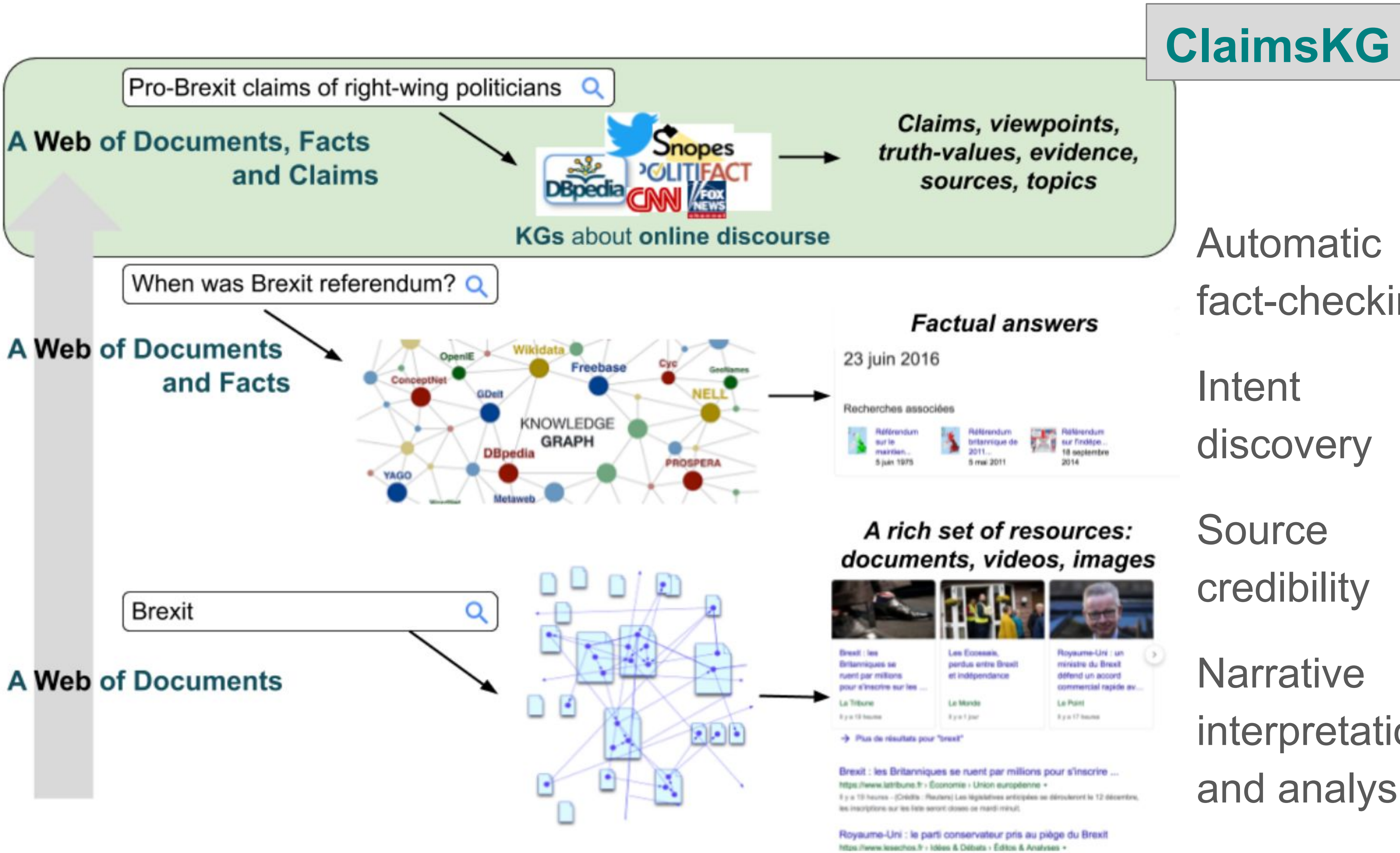
AI4SCI MESRI co-funded **ANR French-German** project

- computational methods at the intersection of **machine learning, distributional semantics** and **structured knowledge**
- **trace, detect, interpet, link** and **classify** scientific claims in online news & social media



Fighting back **science-related mis- and disinformation** online
Create tools for **social scientists and journalists**

Beyond Facts: a Web of Claims and Discourse



Automatic fact-checking

Intent discovery

Source credibility

Narrative interpretation and analysis

ANR DACE-DL: Data Centric AI Driven Data Linking

Doing linked knowledge graphs by using graph neural networks and ML

Establishing typed links between resources across two knowledge graphs.

⇒ Difficult when data are highly heterogeneous or domain specific!

Example:

`http://yago-knowledge.org/resource/Ludwig_van_Beethoven,`
owl:sameAs, `http://dbpedia.org/resource/Ludwig_van_Beethoven`

Human genetics, agro-ecology…

Collaboration with the IGH - the Human Genetics Institute Montpellier

- Predict the adaptive defensive strategy and immune response of human organisms when exposed to pathogens

Collaboration with Elzeard, a Bardeaux-based start-up in agro-ecology

- Assist farmers in the culture rotation processes in order to decrease the use of pesticides and optimize crops

Thank you for listening.