AI, deep learning, computer vision

Ronan Sicre (MCF)
Ecole Centrale Méditerranée / LIS - QARMA

ECM 03/2024

Where is AI ?

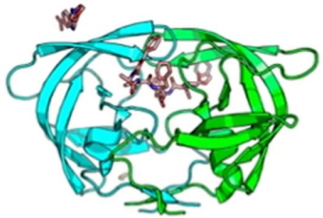
This technology can be found in:

- Games (chess, go, starcraft, dota)
- Web search
- Recommendation
- Audio recognition
- Autonomous driving
- Face/image recognition
- Anti-spam
- Computer-Aided Diagnosis

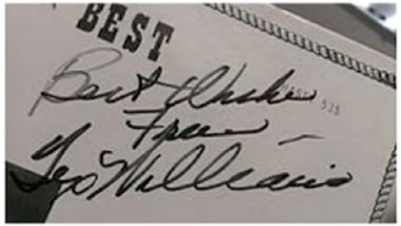
Machine Learning is Everywhere?

AlphaGo

Recommendation systems



Drug discovery



Character recognition

2σ TWO SIGMA

Hedge fund stock predictions

Voice assistants

Assisted driving

Face detection/recognition

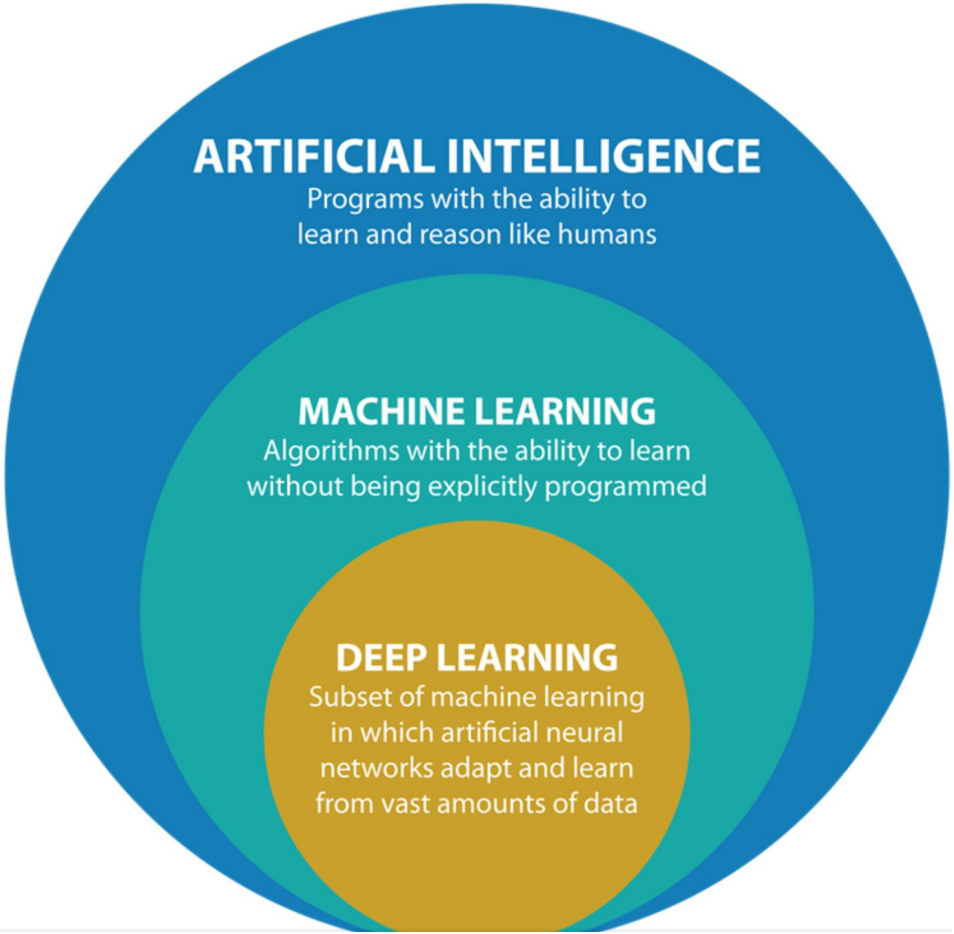
Cancer diagnosis

What is AI ?

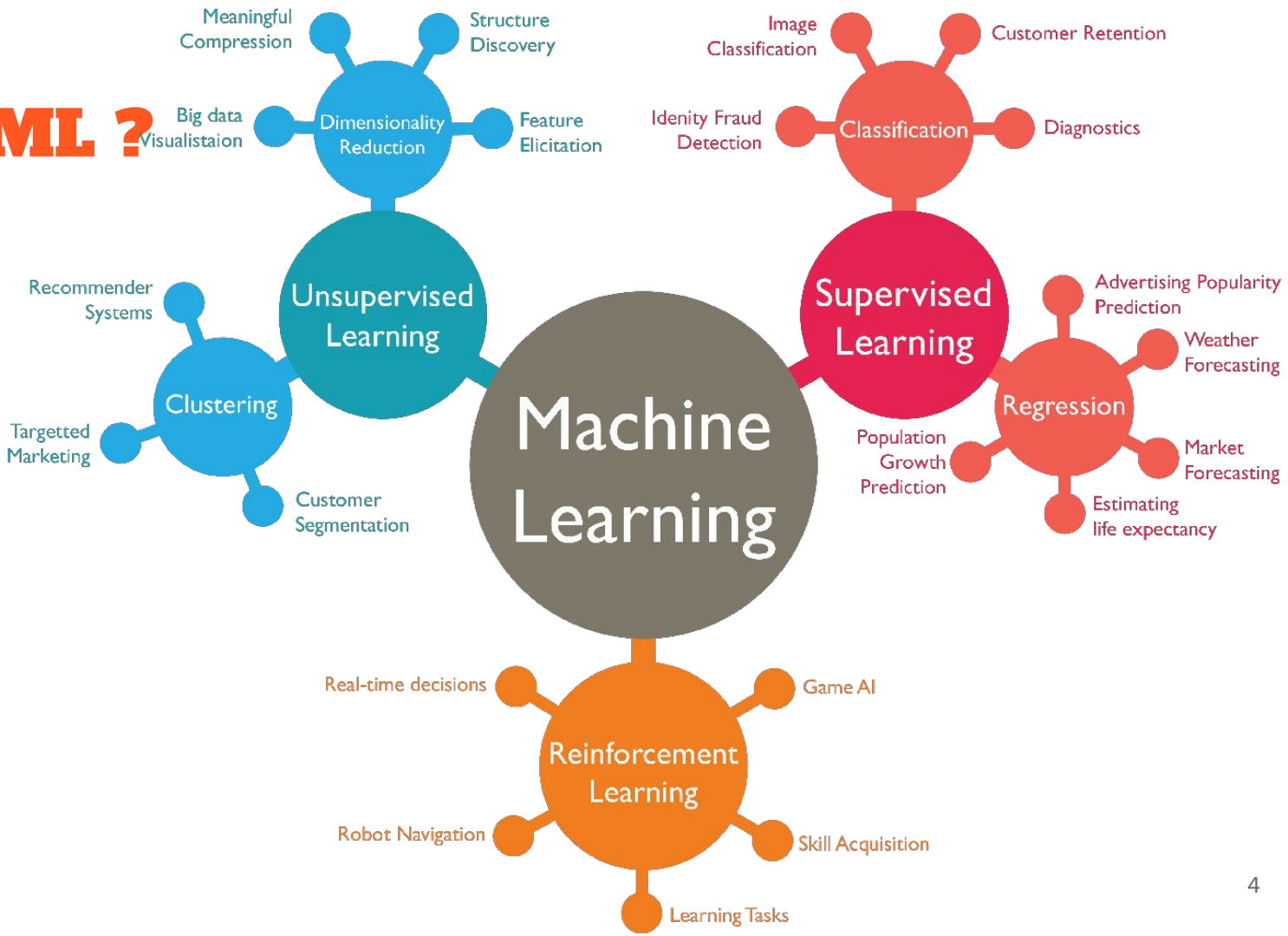
A technology

A science (research field)

- “Old AI”
- Machine learning
- Deep learning



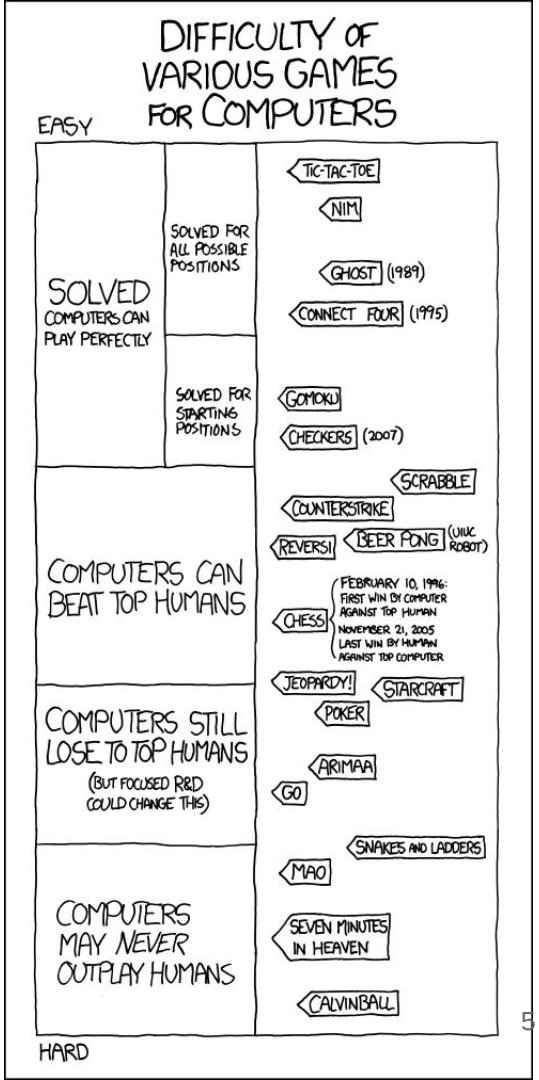
What is ML ?


Machine Learning
Unsupervised Learning
Dimensionality Reduction
Meaningful Compression
Structure Discovery
Big data Visualistaion
Feature Elicitation
Clustering
Recommender Systems
Targetted Marketing
Customer Segmentation
Supervised Learning
Classification
Image Classification
Customer Retention
Idenity Fraud Detection
Diagnostics
Regression
Advertising Popularity Prediction
Weather Forecasting
Market Forecasting
Estimating life expectancy
Population Growth Prediction
Reinforcement Learning
Real-time decisions
Game AI
Robot Navigation
Skill Acquisition
Learning Tasks

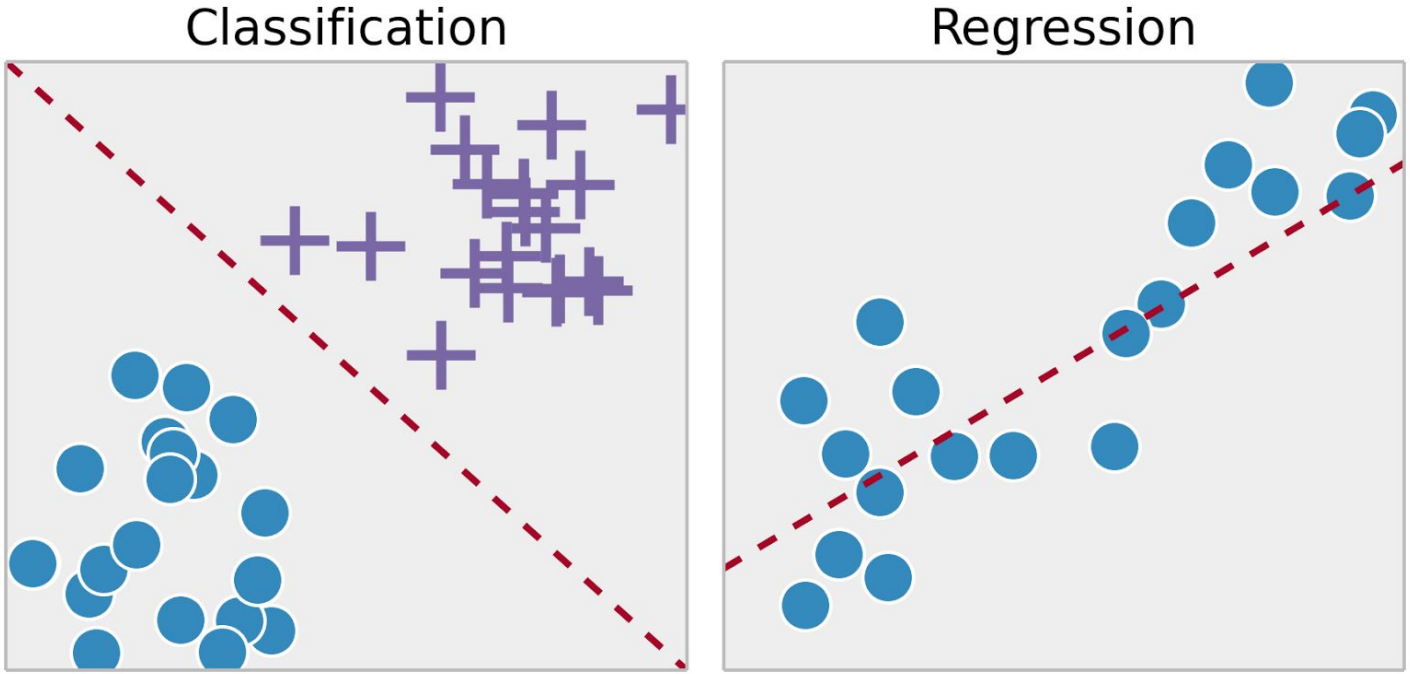
Reinforcement learning

Games, robotics, simulated environment (drive, walk, etc.).

https://www.youtube.com/watch?v=SX08NT55YhA



Supervised learning



Example of classification

Mushrooms:

- Edible
- Poisonous

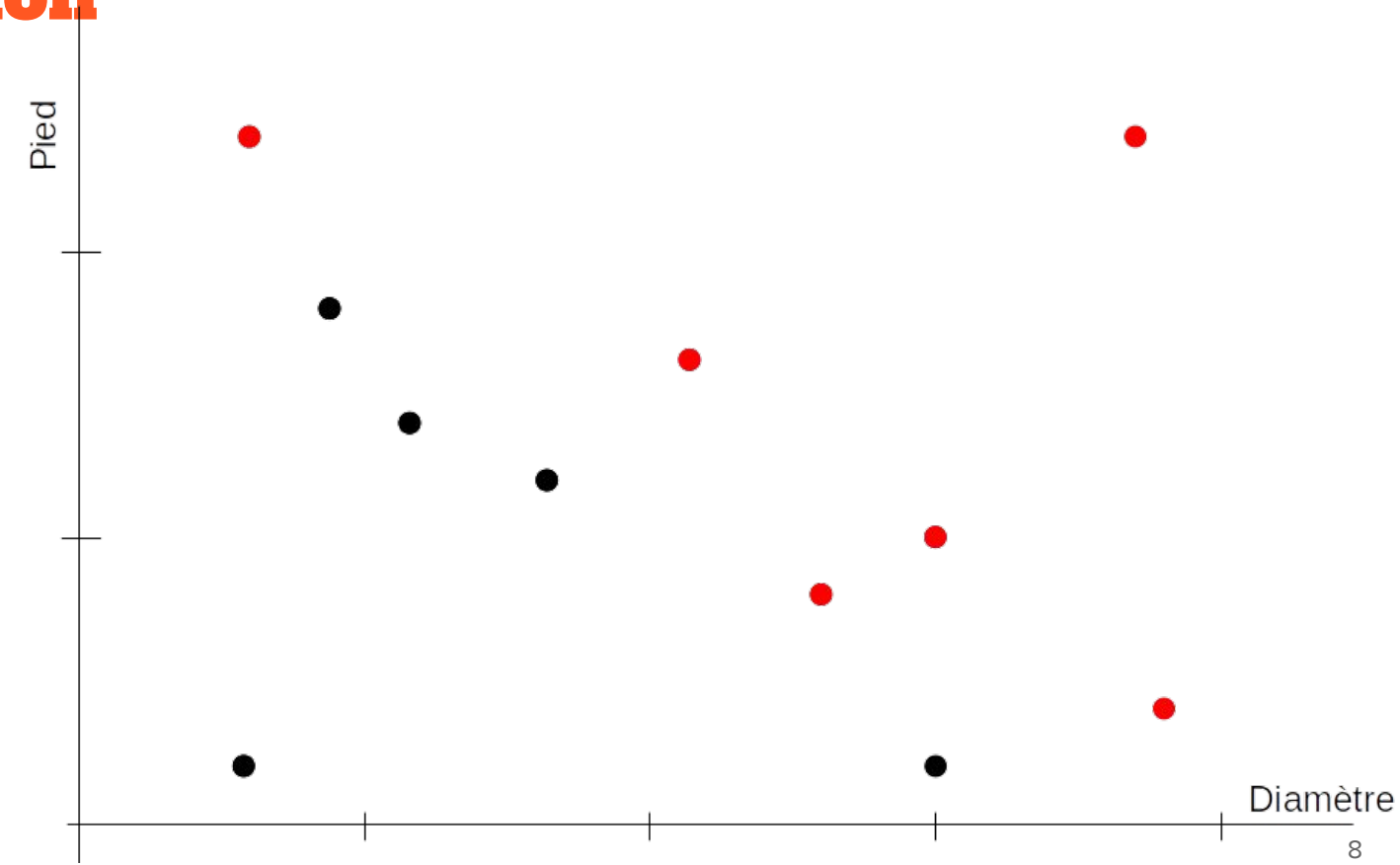
We look for regularities

in the data

Classification

Height of the foot

Diameter

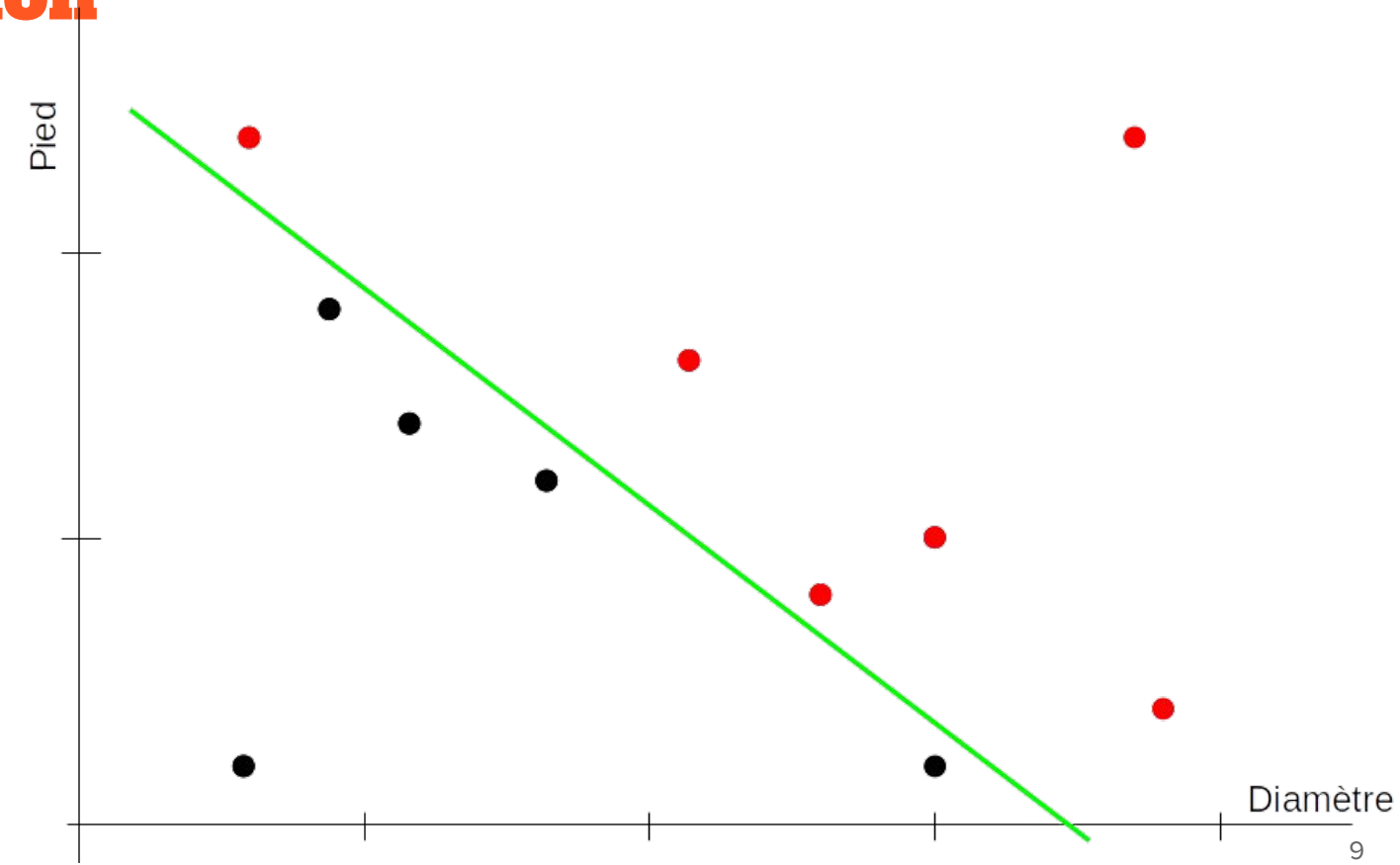


Classification

Training data

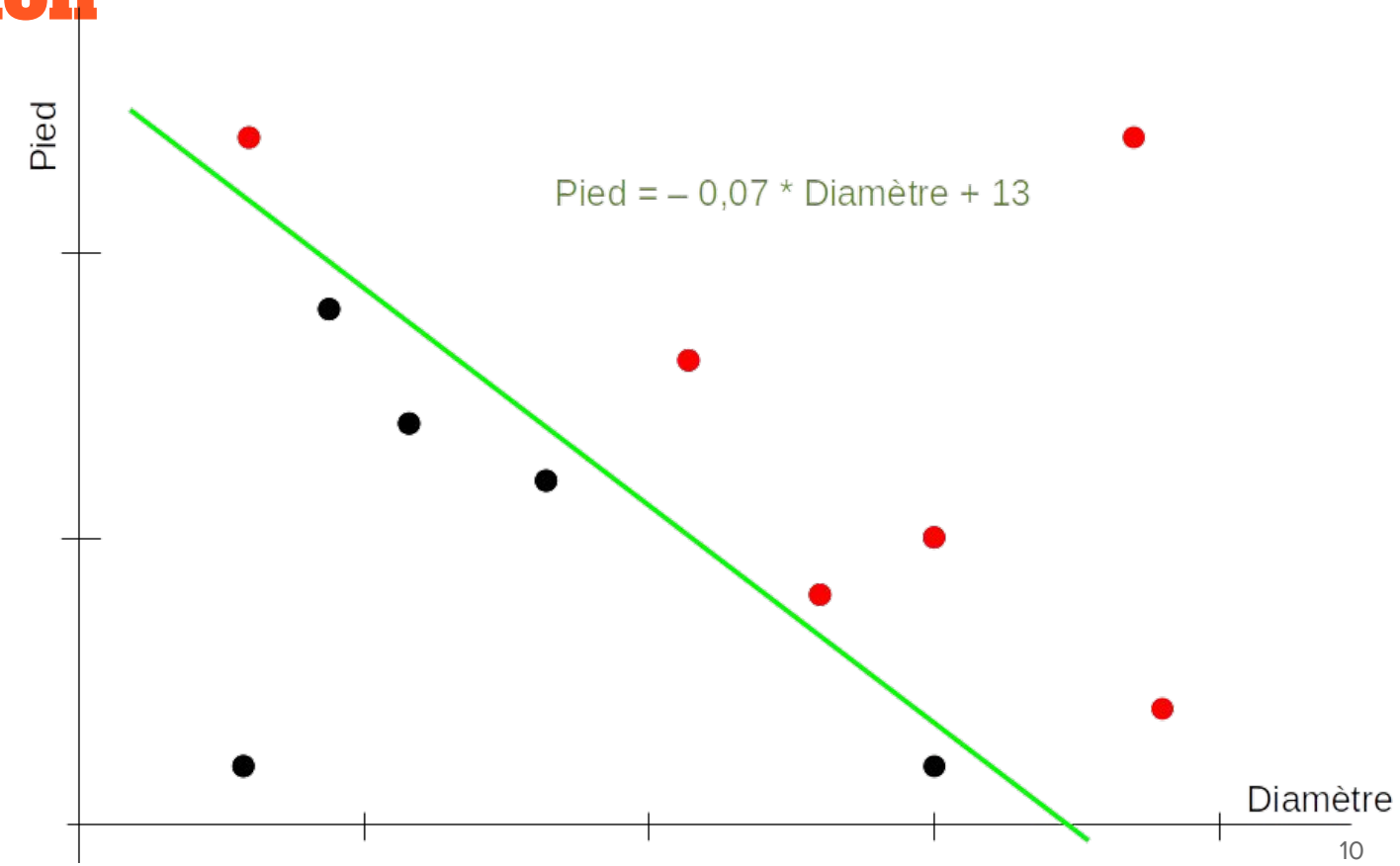
Find a separation

$y = a.x + b$

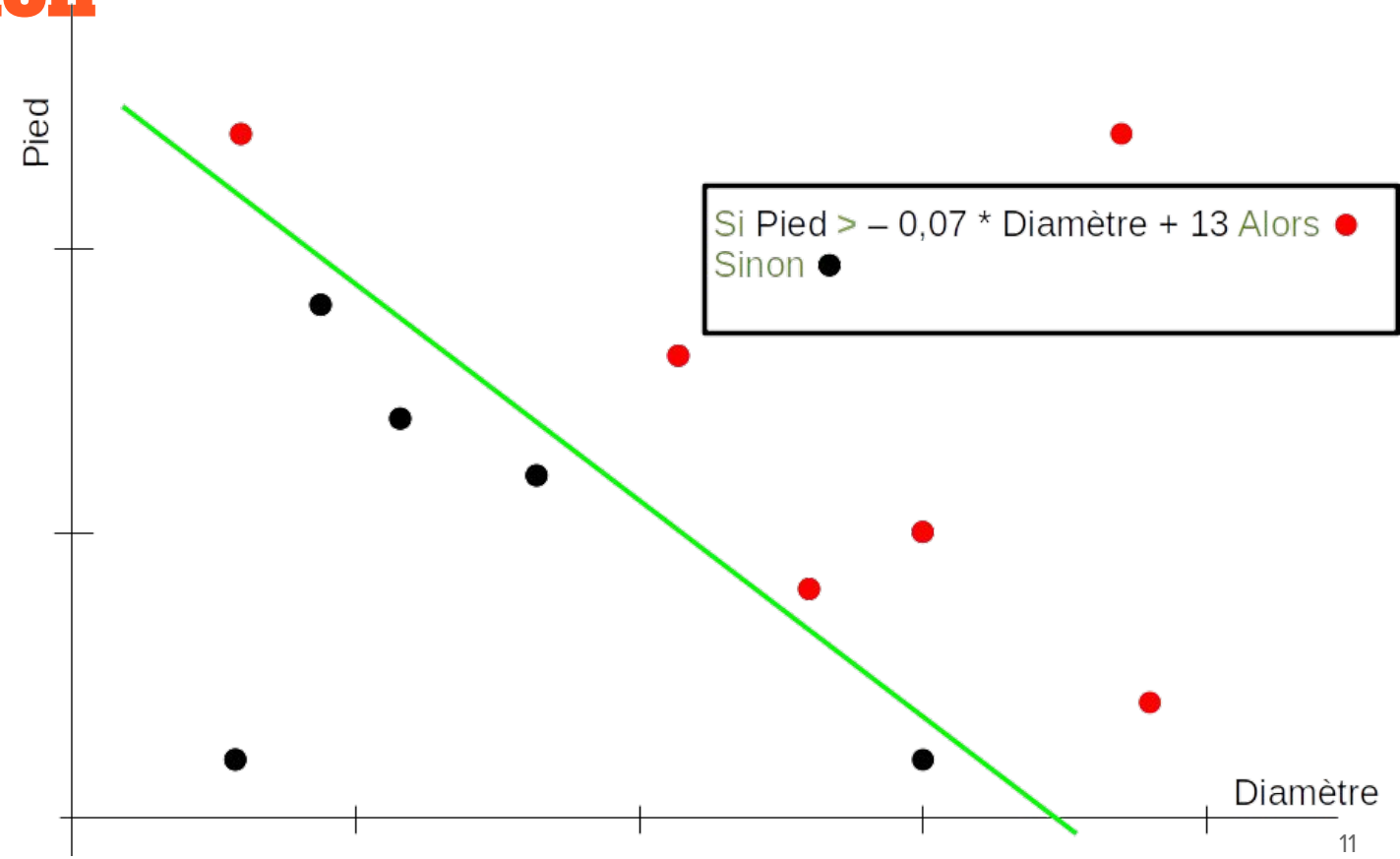


Classification

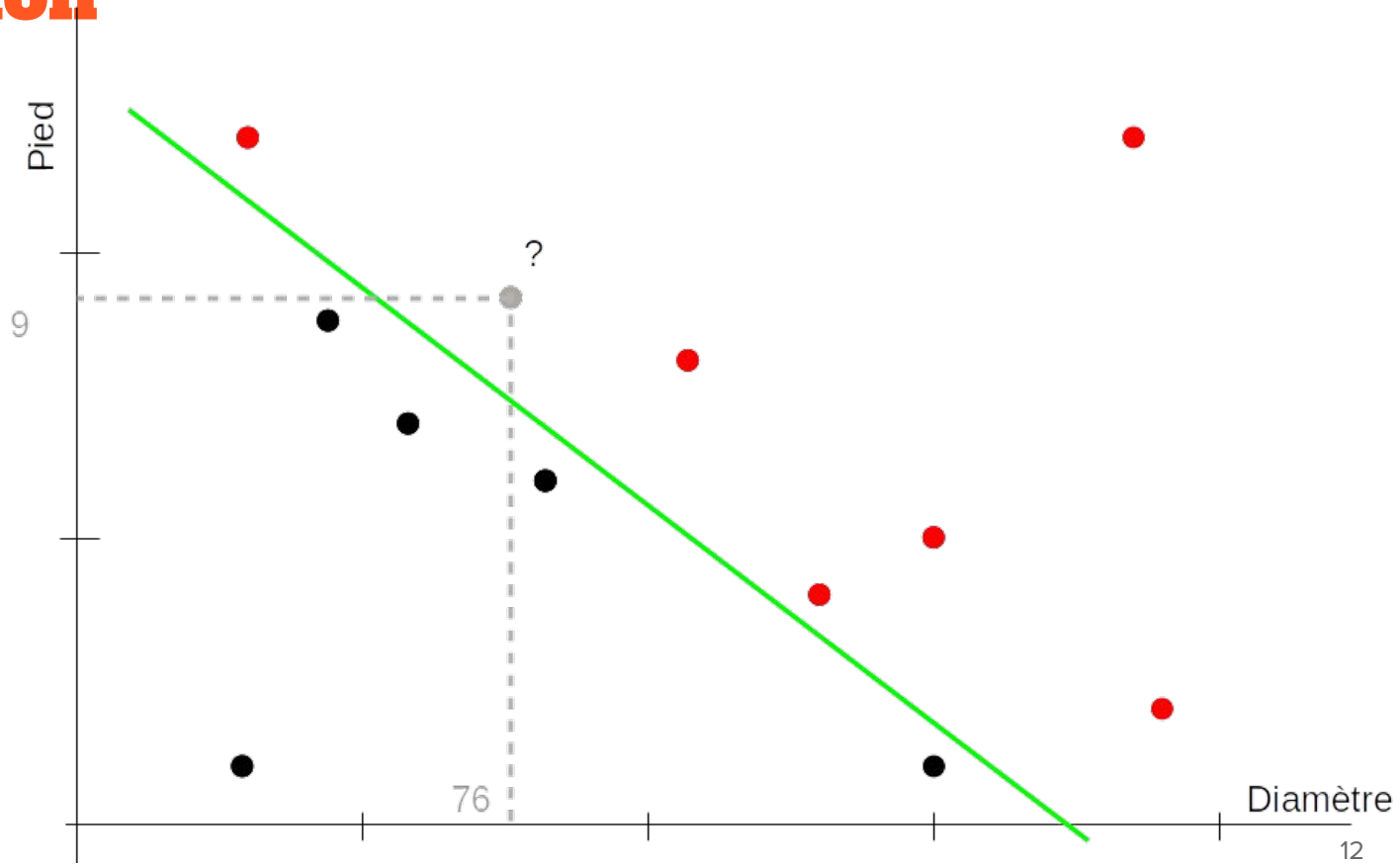
$y = a.x + b$



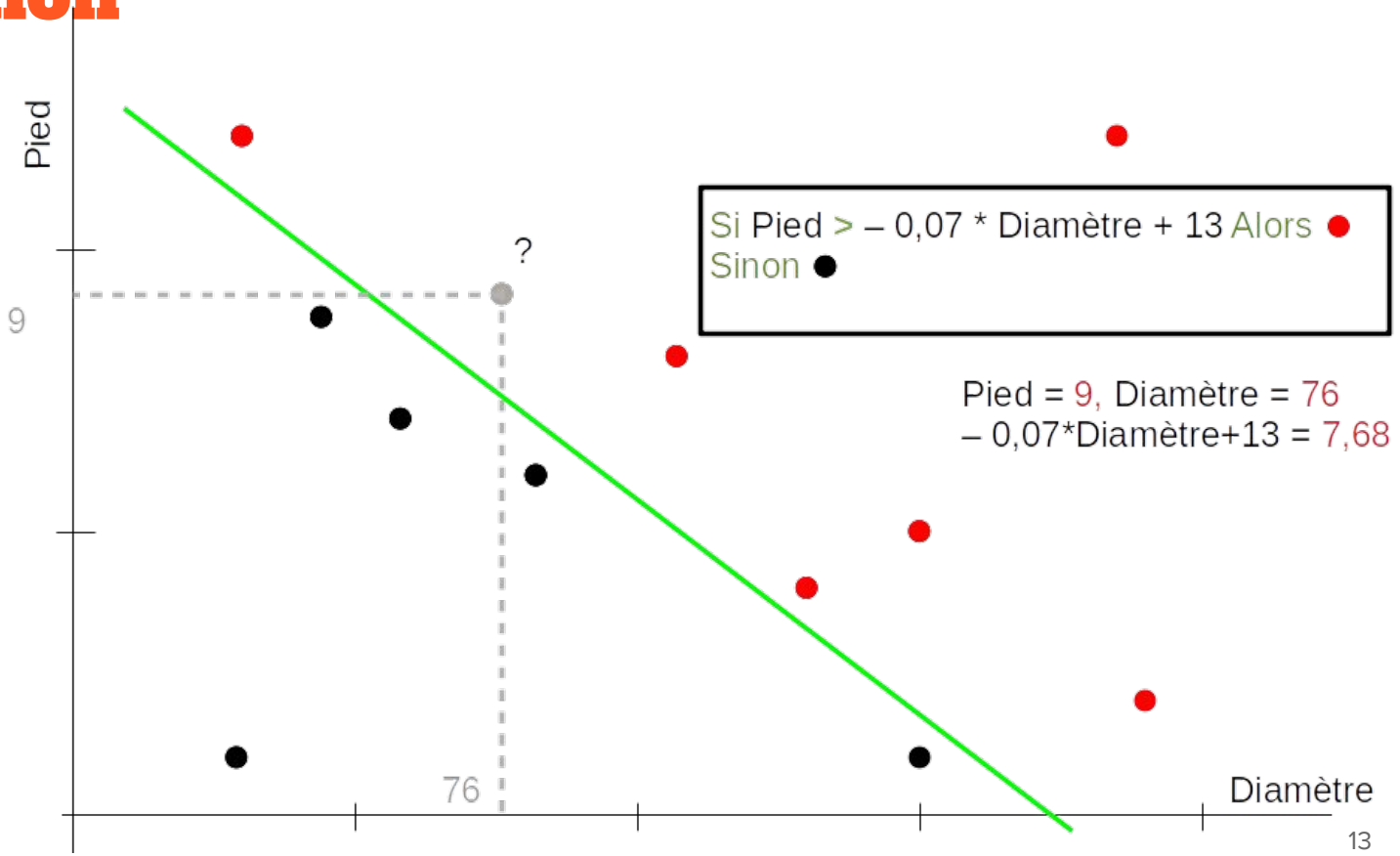
Classification



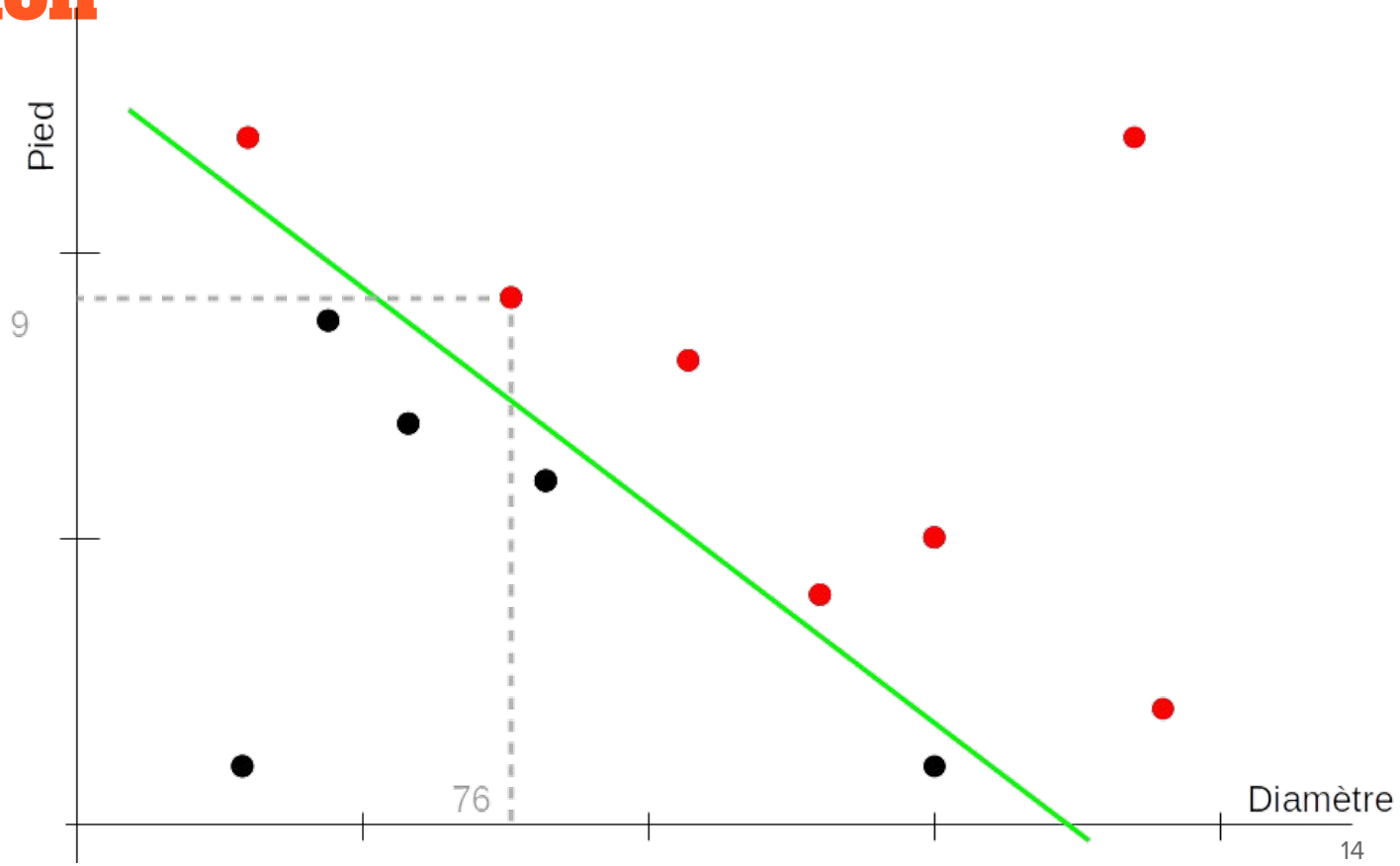
Classification



Classification

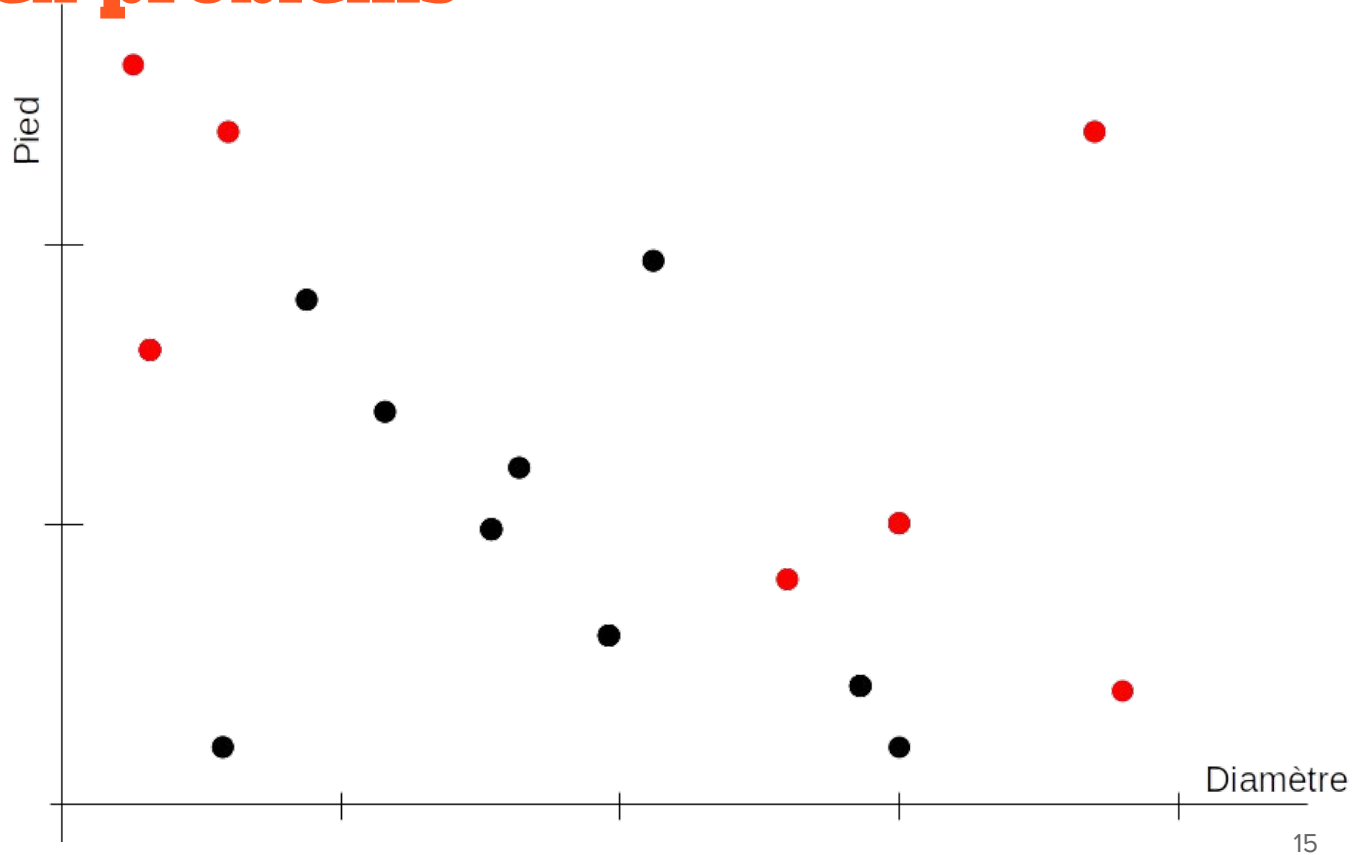


Classification

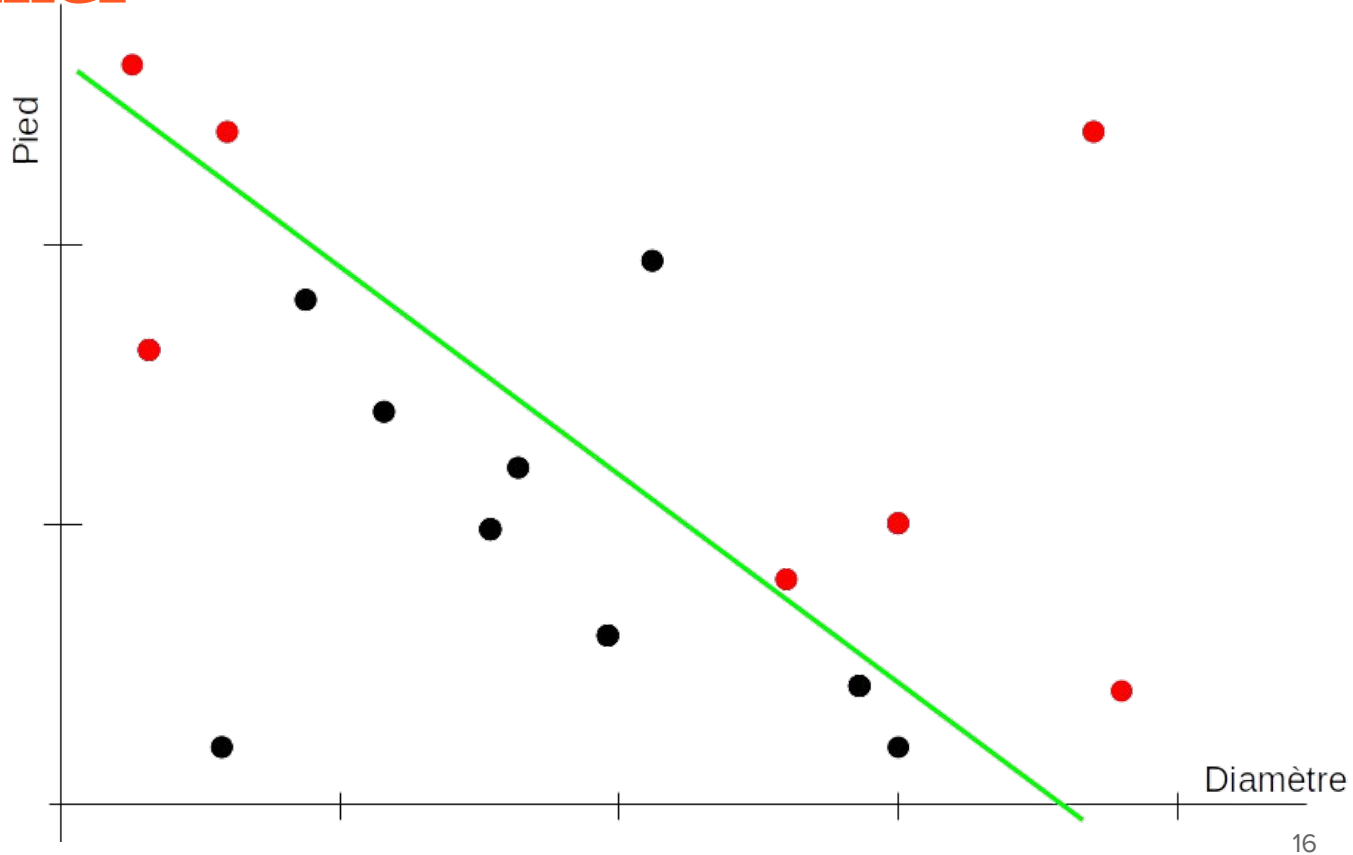


More complex problems

Not linearly separable



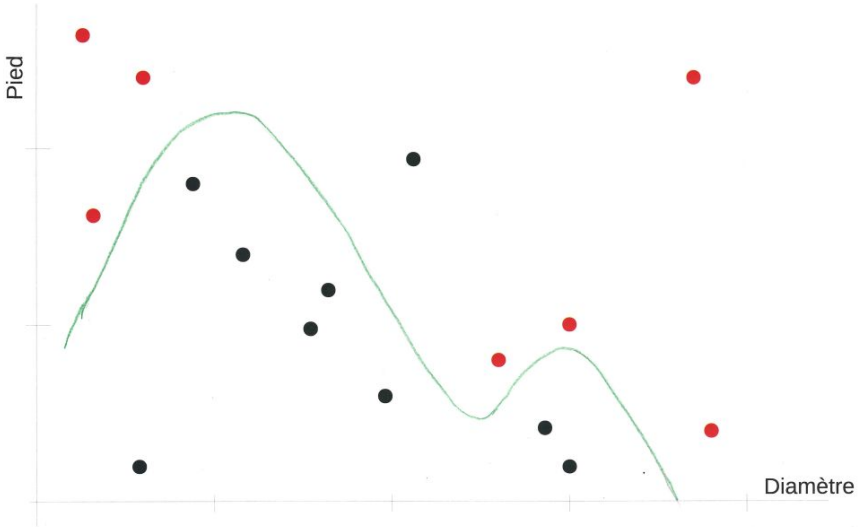
Linear classifier



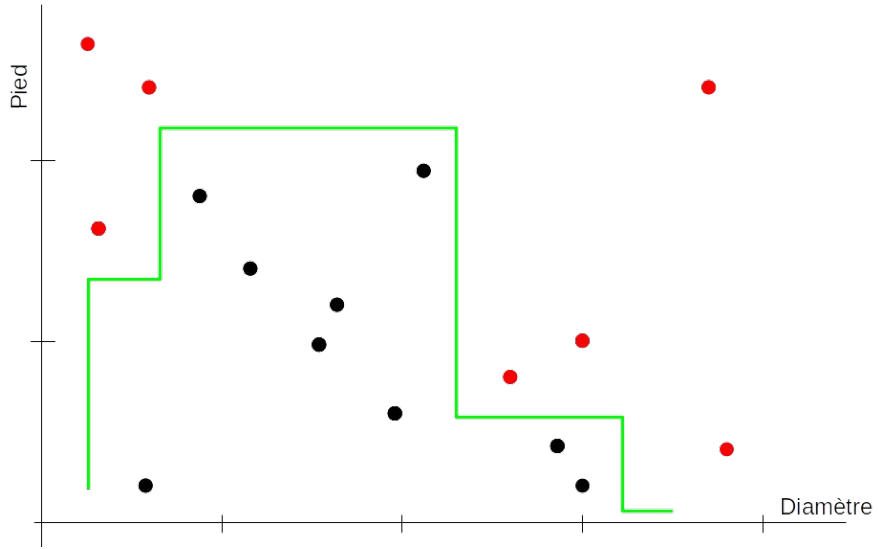
More classifiers

More complex models, more parameters:

Polynomial $y = \sum_{i=0}^{n} a_i x^i$

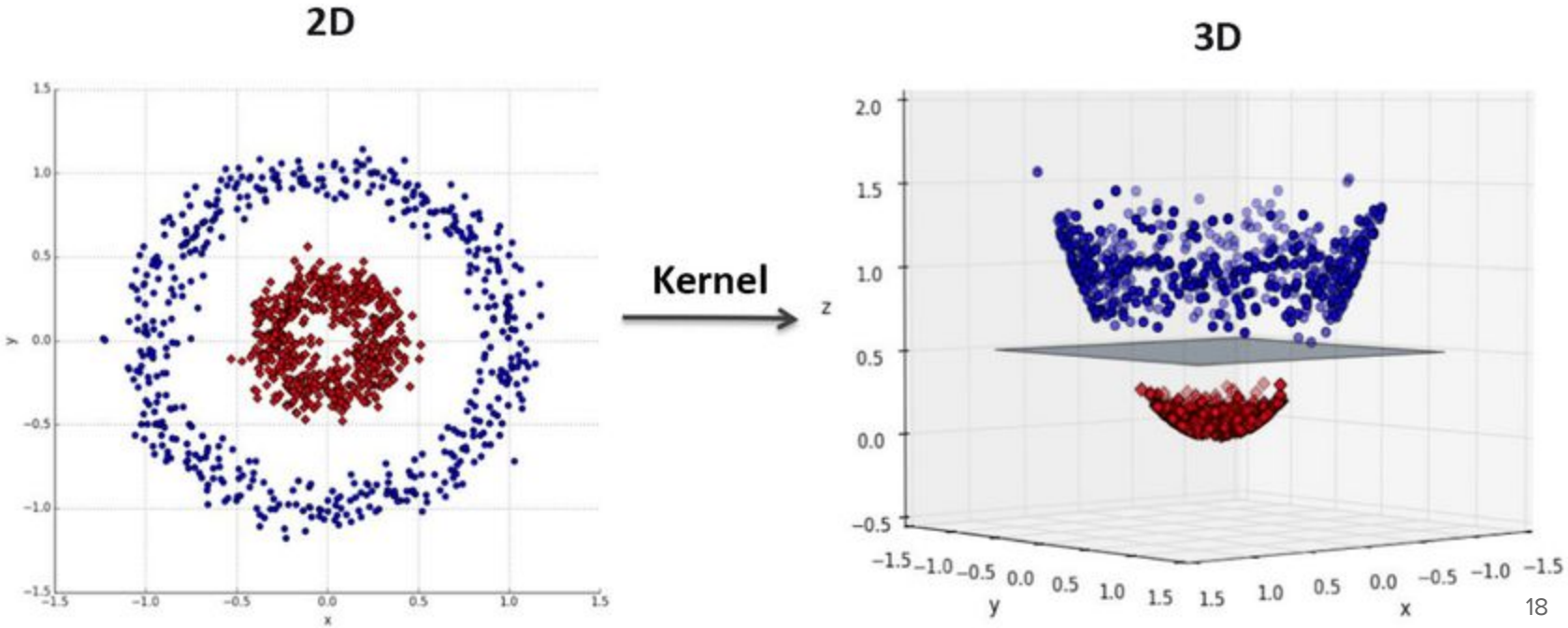


Piecewise linear



Kernel trick

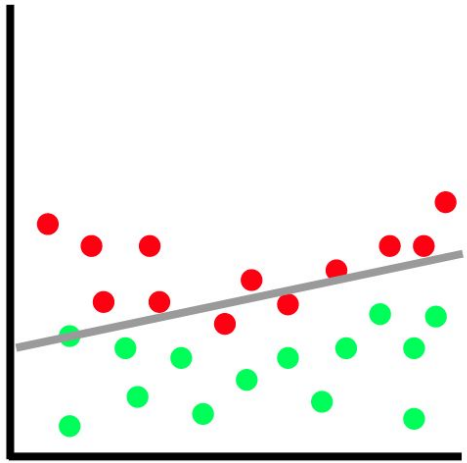
Project your data into a new space of higher dimension

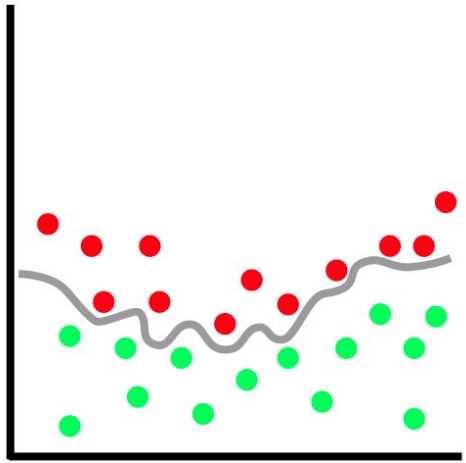


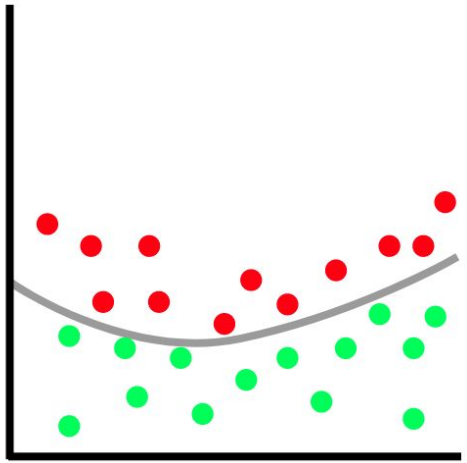
Choice of the model

Simple model that works well on training data.

More importantly, model that can **generalize** well to new data.

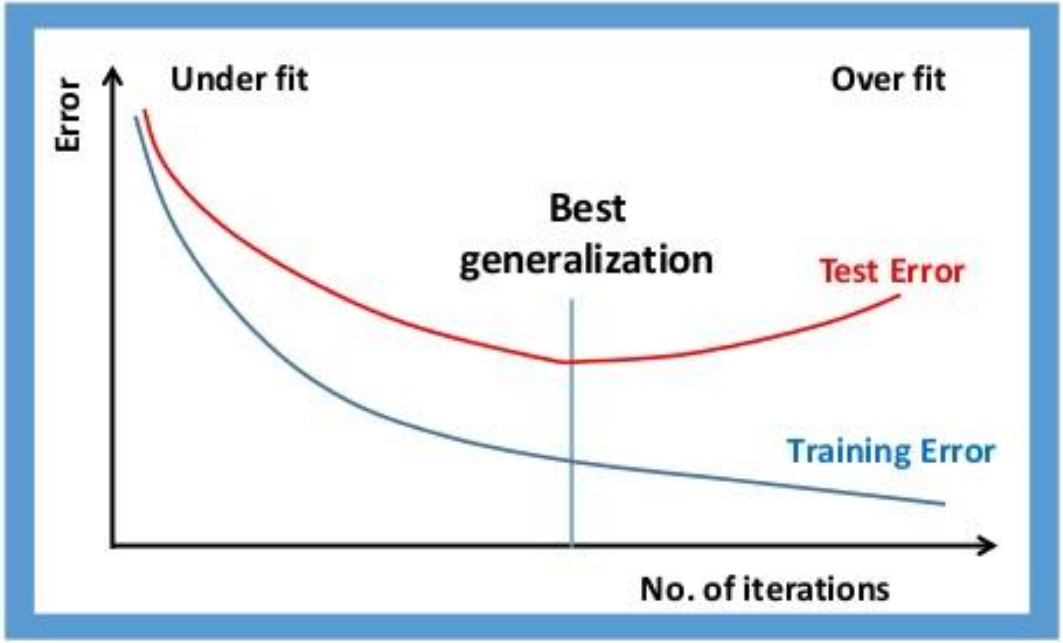






Training a model

Generalization



Train, validation, test

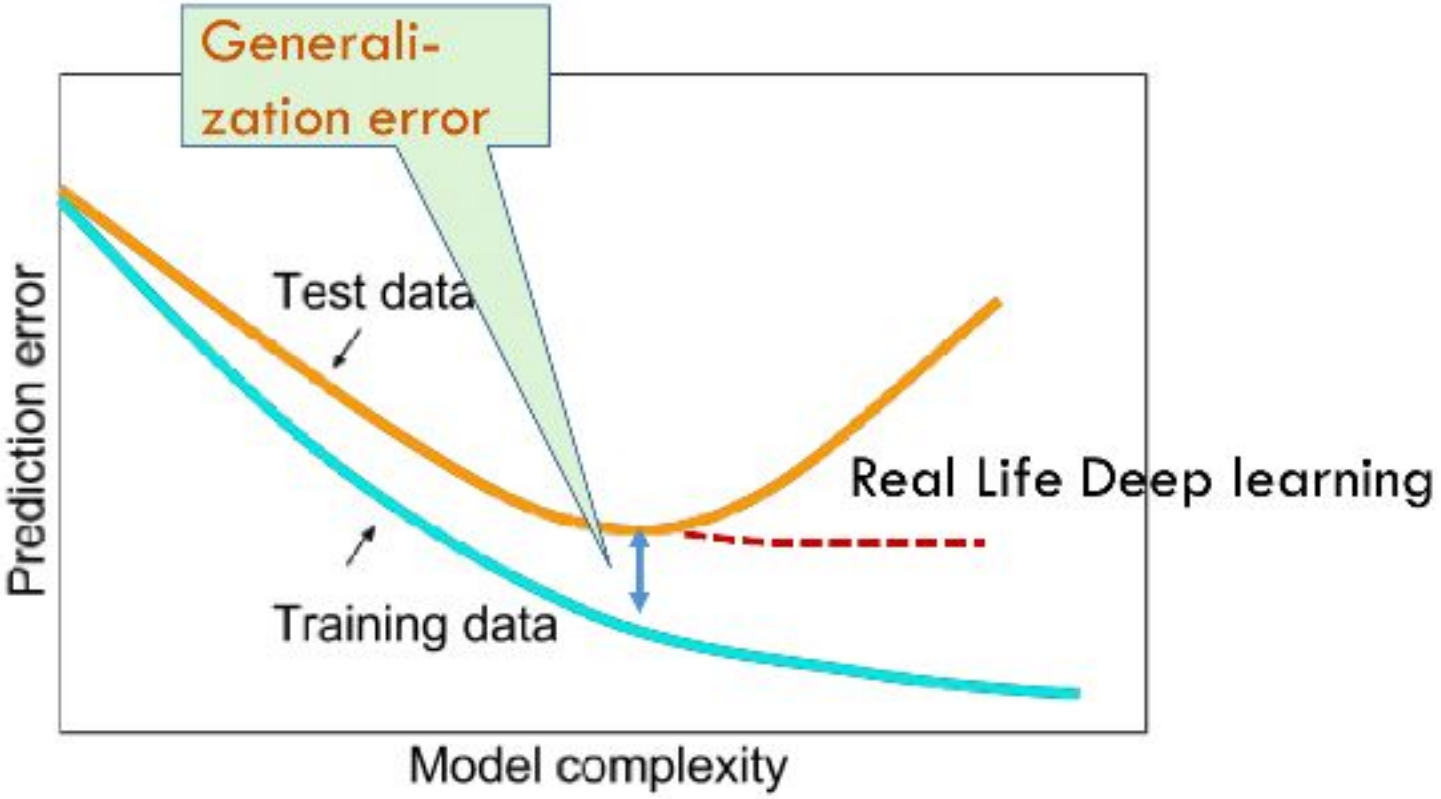
Train and validation: try out numerous combination of hyperparameters.

Adjust based on the validation performance.

Then use test data for final results.

Lots of experiments - comparison on open datasets

Training a model



The deep learning field

Supervised learning: classification, regression

Unsupervised learning

Reinforcement learning

Self-supervised learning

Generative models (image generation, deep fakes, chat GPT)

Deep learning (recent) history

Neural Networks were studied in the 90s then disappear in the 2000s

In 2011, AlexNet wins the ImageNet Challenge: image classification (1M images, 1k categories).

Why Convolutional Neural Networks work !

- Lots of data
- Lots of computing power (parallelization on 2 GPUs)

Types of data

Image // image, video recognition (Computer vision)

Text // translation, information extraction, classification (Natural Language Processing)

Audio // speech, sound, music

Time series // weather, stock market

Graphs // social medias, brains

Multi-view, multi-modal

Architectures

Convolutional Neural Networks (CNN),

Transformers,

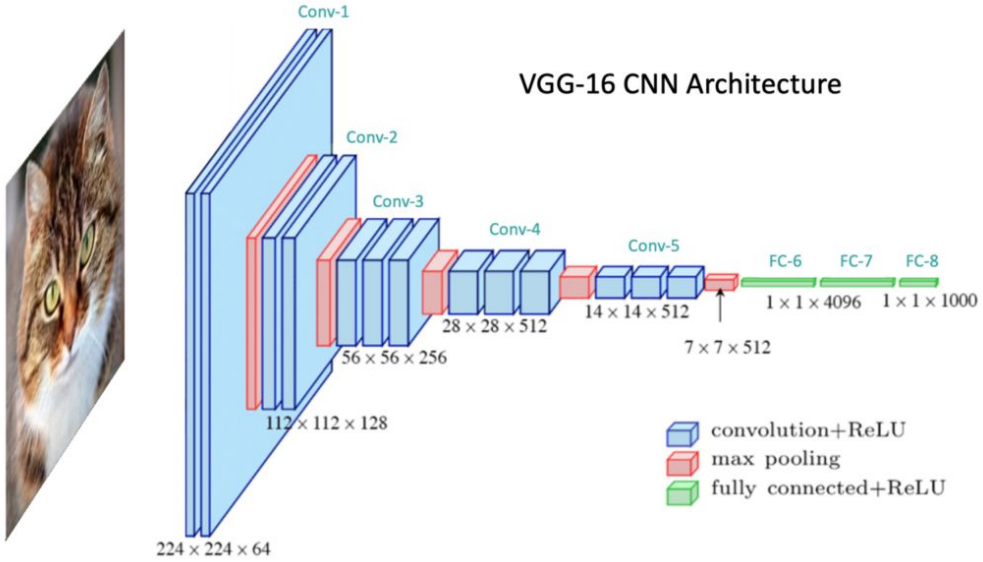
Graph Neural Networks (GNN, GCN),

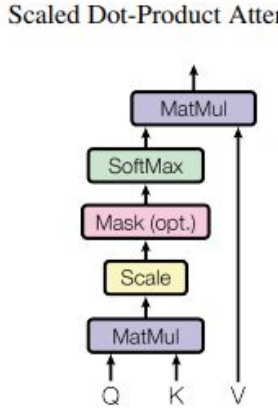
Spiking Neural Networks

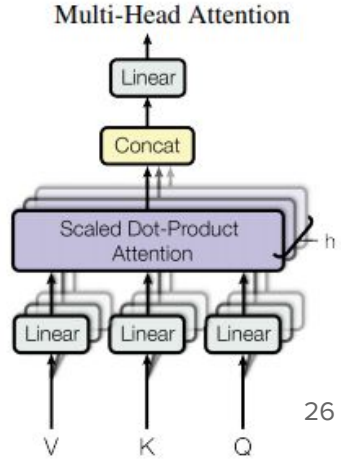
Recurrent Neural Networks (RNN),

Gated Recurrent Unit (GRU),

Long Short-Term Memory (LSTM)







Training

Loss function:

- Adapted to the task: classification, regression, reconstruction
- Regularizations

Optimization: Stochastic gradient descent (SGD)

- Optimizer choice: learning rate evolution

Computer Vision

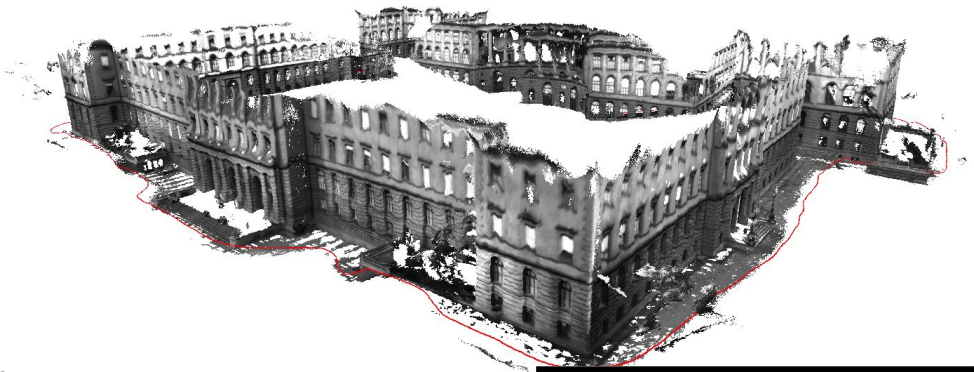
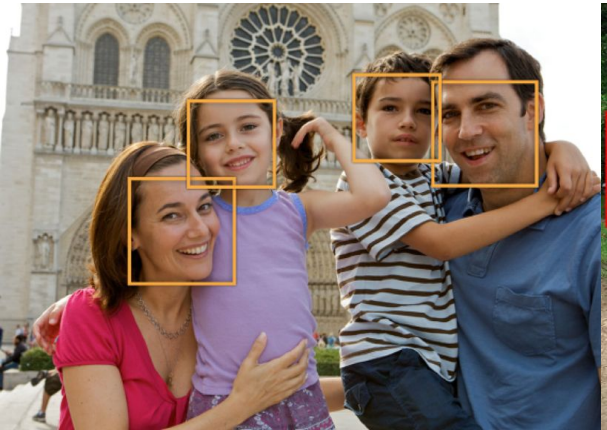
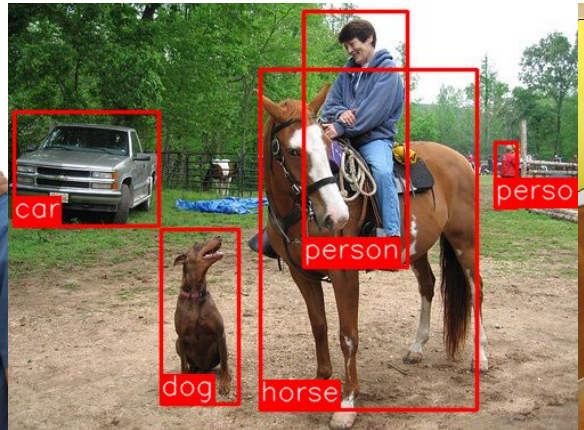


Image classification, object detection,

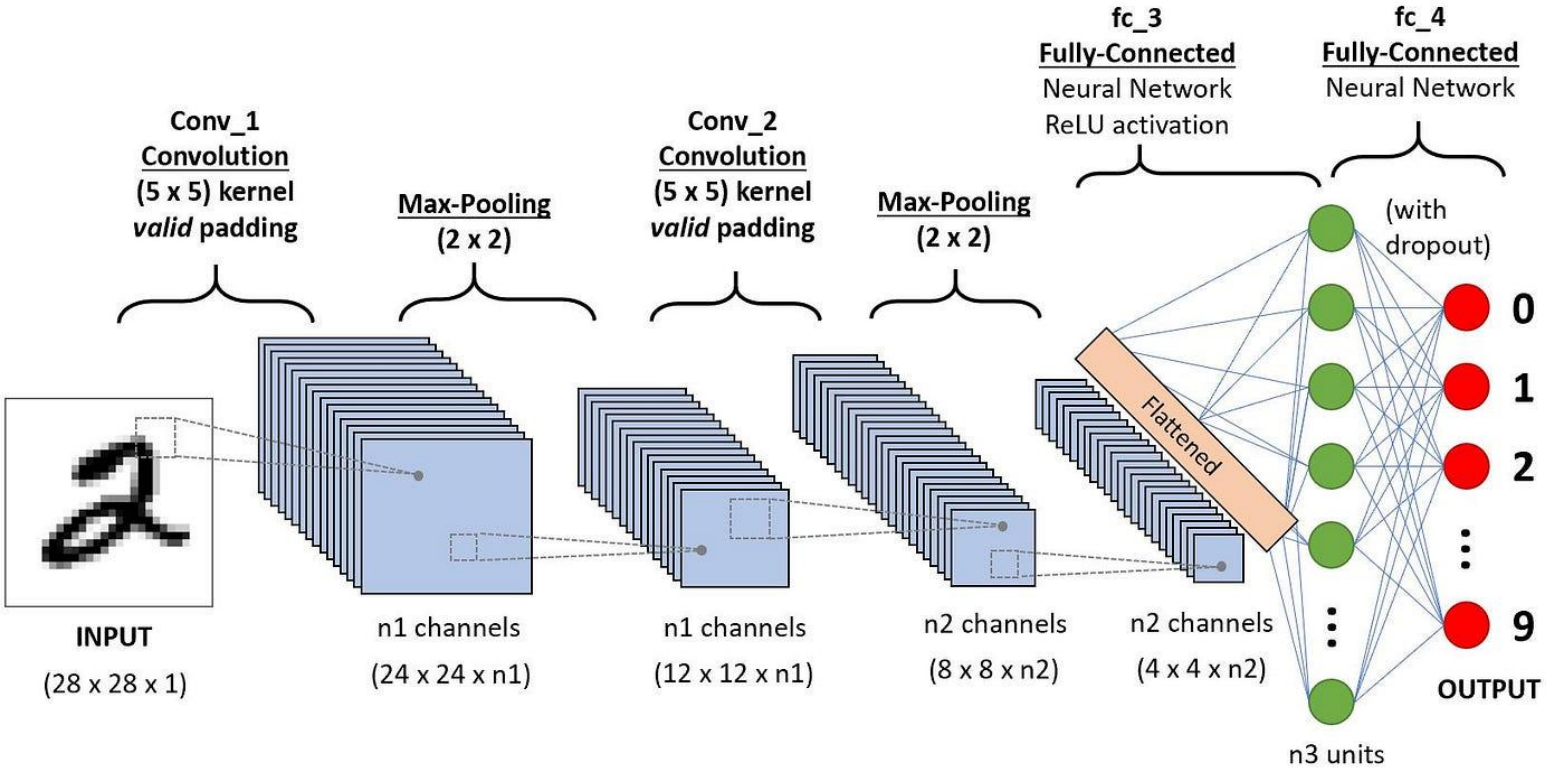
Instance segmentation, generative models,

Videos: object tracking, action recognition.





Example of image classification



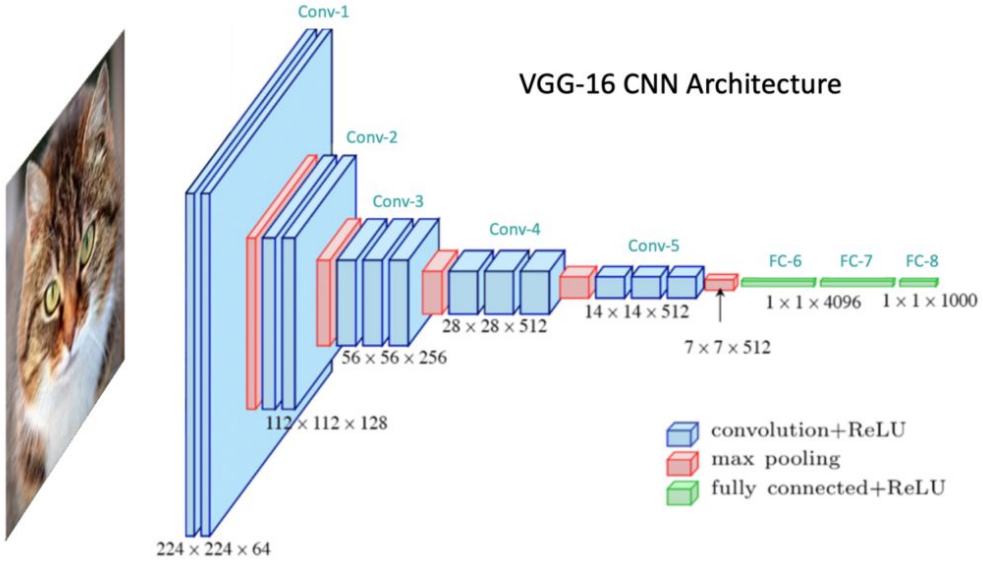
Architectures: CNN, Transformers

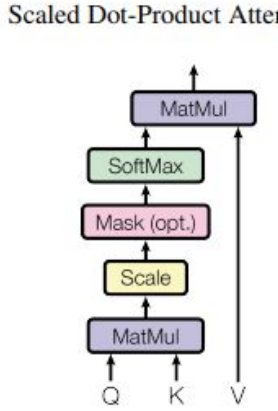
CNNs: LeNet, AlexNet, VGG,

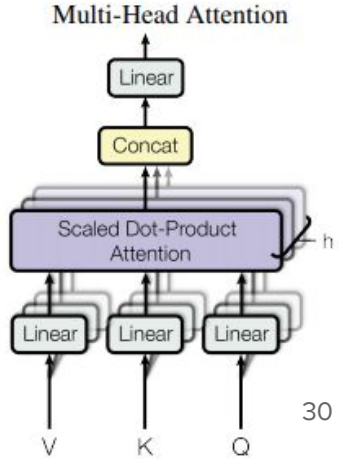
GoogleNet, ResNets, DenseNet,

NAS-Net, Efficient Net…

Transformers: ViT



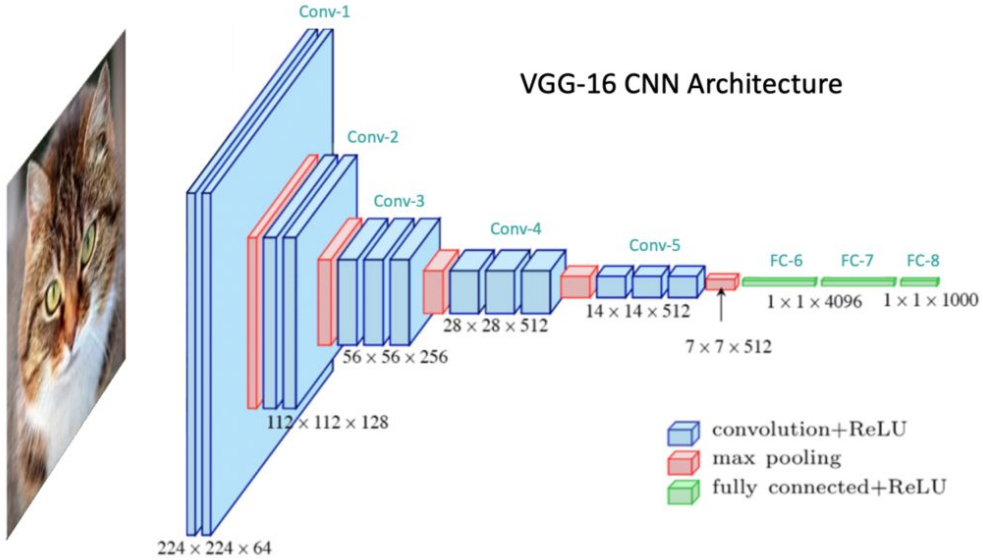




Train an image classification model

1 Million image: train from scratch

Transfer learning: use a pretrained Network (ImageNet) and fine-tune on your data.



Self-supervised learning

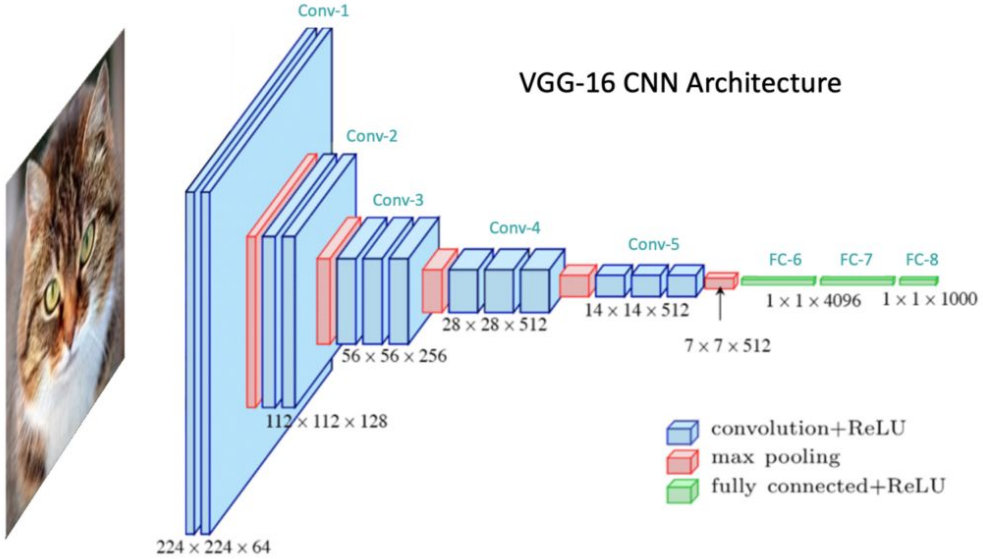
Learn representations

- Without labels
- With good transfer capabilities

Clustering, pretext tasks,

augmentations and contrastive loss,

distillation, masking...



Object detection

2 stages: Faster R-CNN

1 stage: **Yolo** v1...9

https://www.youtube.com/watch?v=ZyKK4o4HaAM

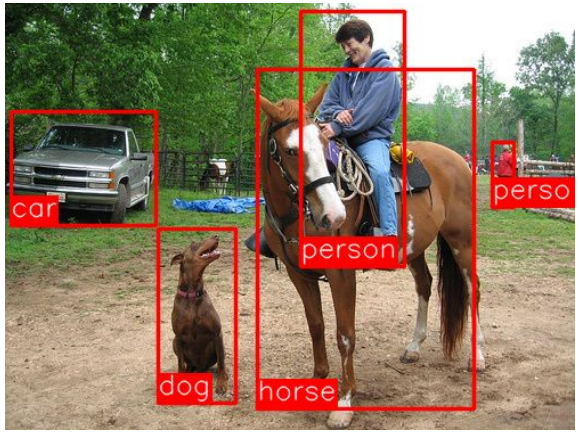
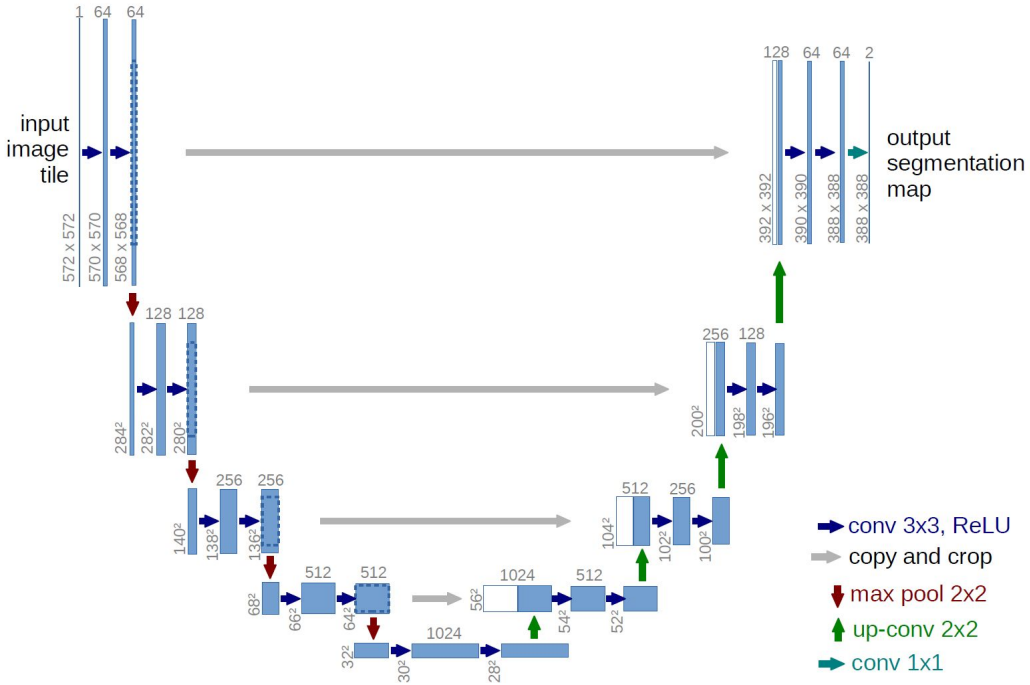
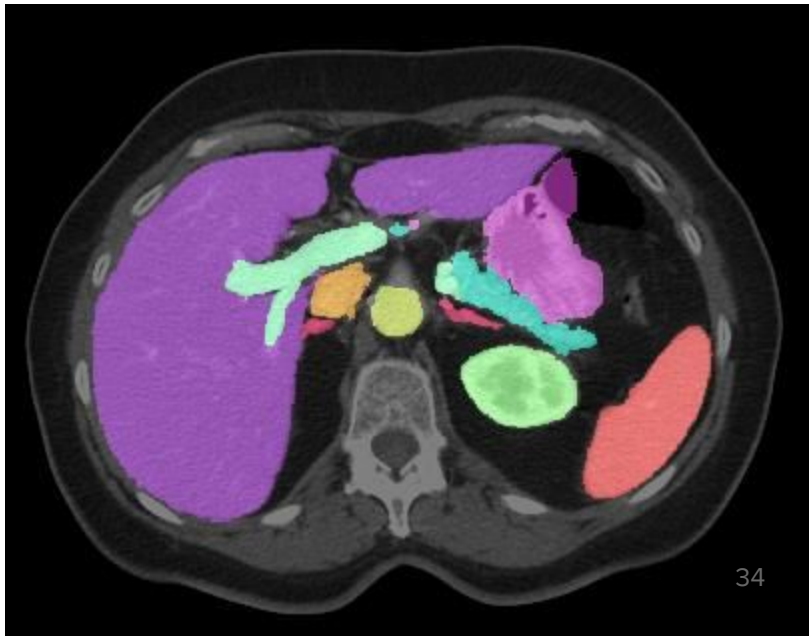


Image / instance segmentation

U-Net





Generative models

Generative Adversarial Networks (GANs), Variational auto-encoder (VAE), Masking auto-encoder (MAE), Denoising auto-encoder (DAE), Diffusion models.

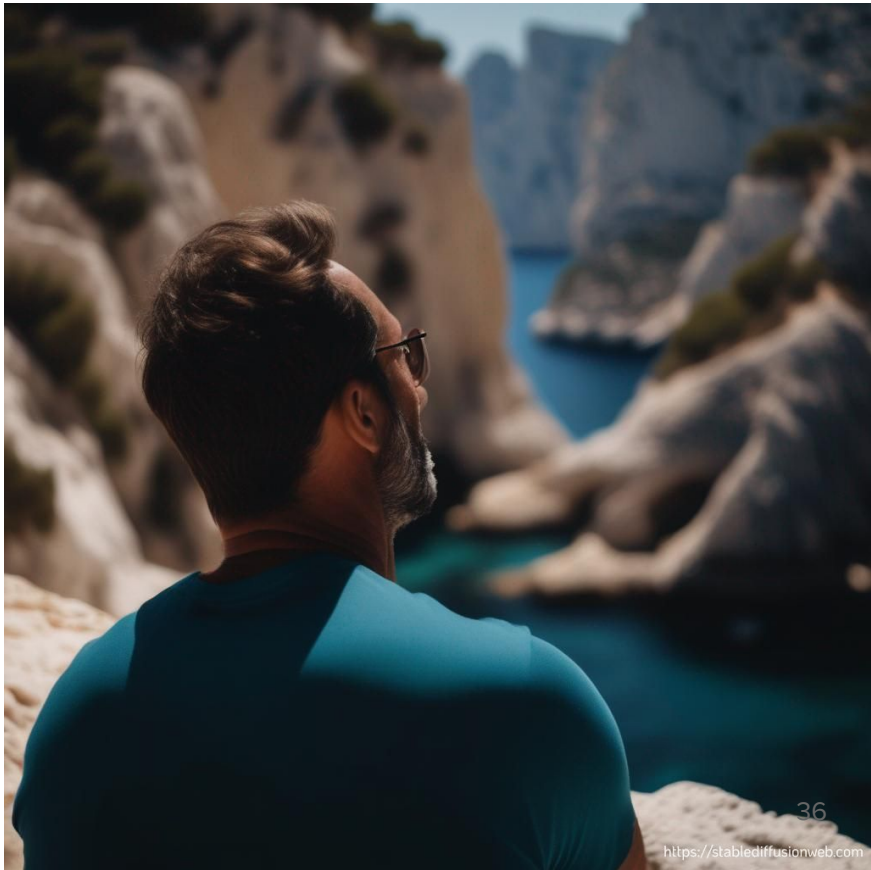
https://thispersondoesnotexist.com/

Generative models

Stable diffusion:

image generation from prompt

“**a guy giving a seminar in the calanques**”



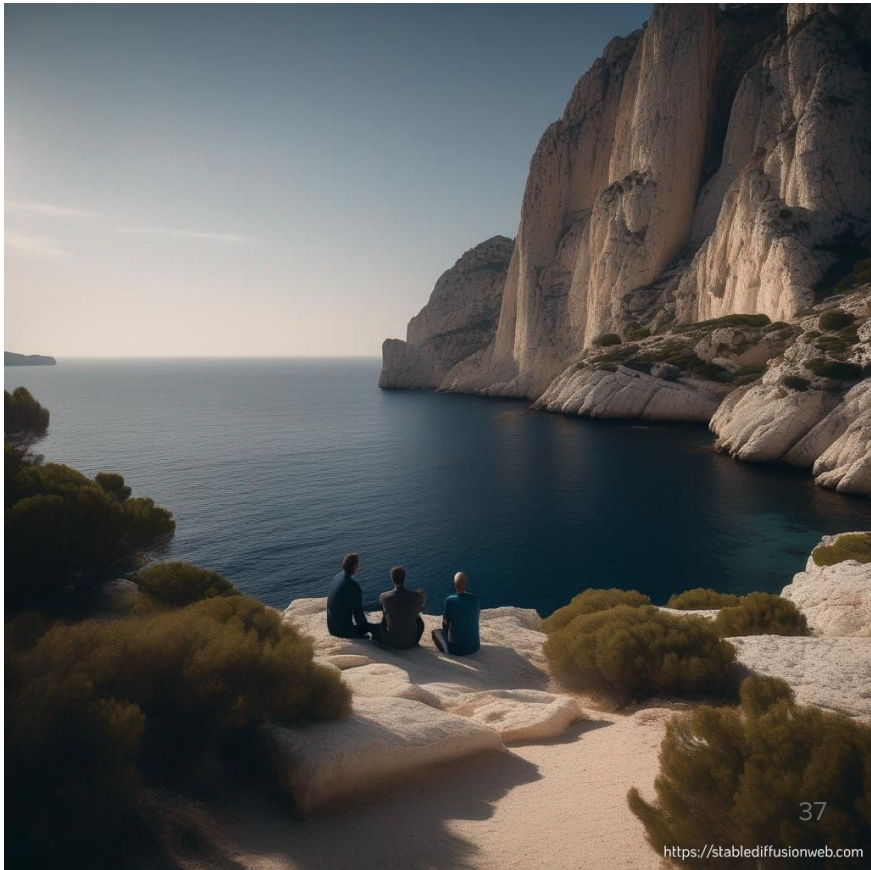
https://stablediffusionweb.com

Generative models

Stable diffusion:

image generation from prompt

“**a guy giving a seminar to scientists in the calanques**”



https://stablediffusionweb.com

Biases, ethics, fairness, privacy

Models reproduce biases in the data

Model can take shortcuts

Inforce fairness when training

Inforce privacy when training

Robustness

Adversarial examples

classified as

Stop Sign

+

=

classified as

Max Speed 100

“pig” (91%)

+ 0.005 x

noise (NOT random)

=

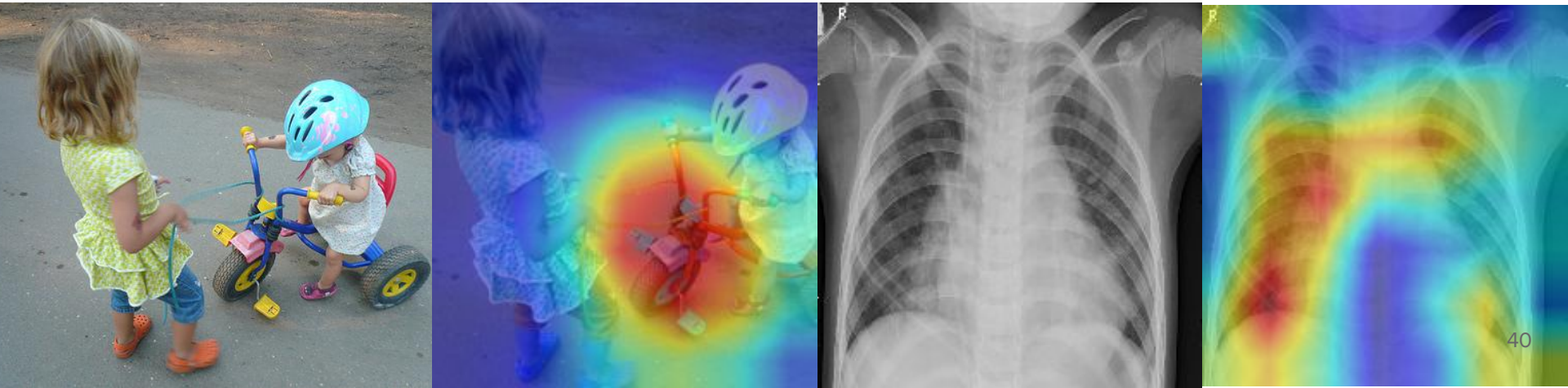
“airliner” (99%)

XAI

eXplainable AI: post-hoc interpretability vs transparency

Attributions, local vs global.

Saliency maps as an explanation for image classification.



Environmental impact of deep learning

Power consumption of GPUs

Training BERT = flight NYC to San Francisco.

Frugal models, light models (training, inference)

“Energy and Policy Considerations for Deep Learning in NLP”

https://arxiv.org/pdf/1906.02243.pdf

Questions

Ressources

Thanks to Cécile Capponi, Francois-Xavier Dupé and Yannis Avrithis.

MOOC Andrew Ng (Stanford)

https://www.youtube.com/watch?v=6QRpDLj8huE&t=201s

https://www.youtube.com/watch?v=h0e2HAPTGF4&t=361s

https://ericdatascience.wordpress.com/python-machine-learning-randomforest_p2-parameter-tuning/

https://desh2608.github.io/2018-07-27-deep-learning-theory-2/

UNETR: Transformers for 3D Medical Image Segmentation (WACV 2022)

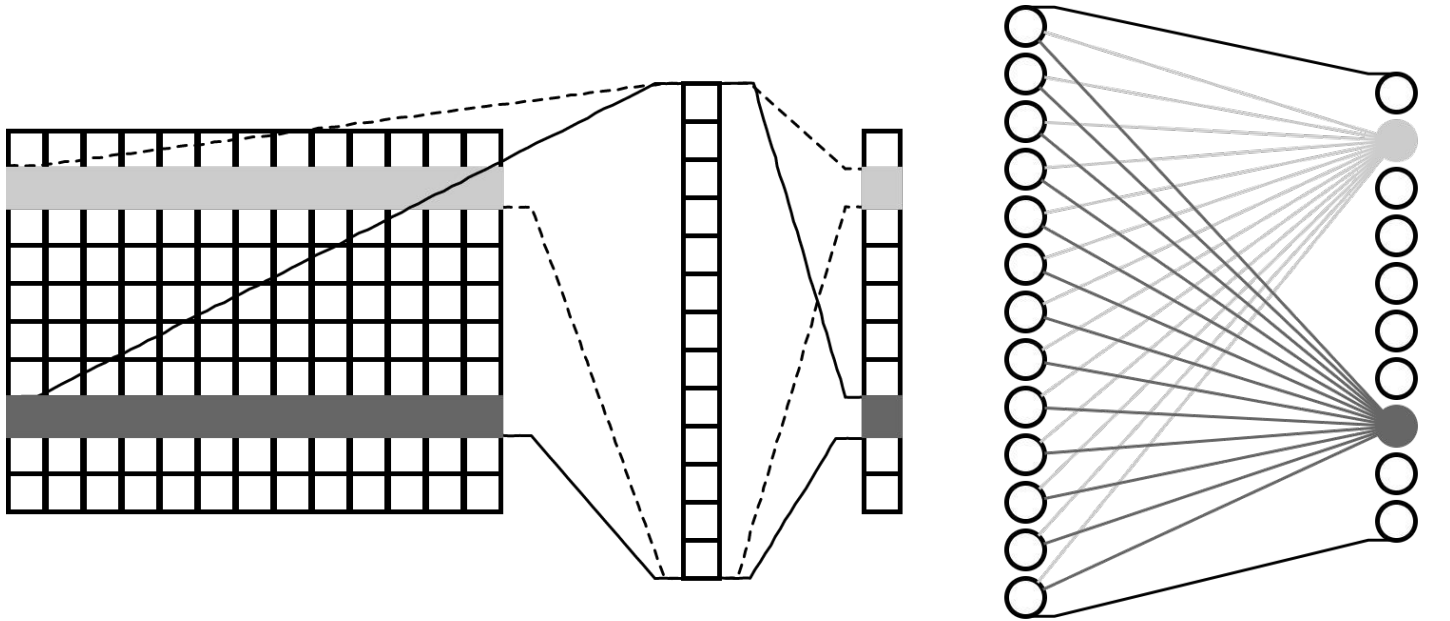
https://stablediffusionweb.com/#ai-image-generator

https://www.researchgate.net/figure/Adversarial-examples-for-traffic-signs-picture-by-Chen-and-Wu-71_fig1_369368588

https://networkpages.nl/ai-thinks-my-dog-is-a-pig-want-to-know-why/

Building blocks of deep architectures

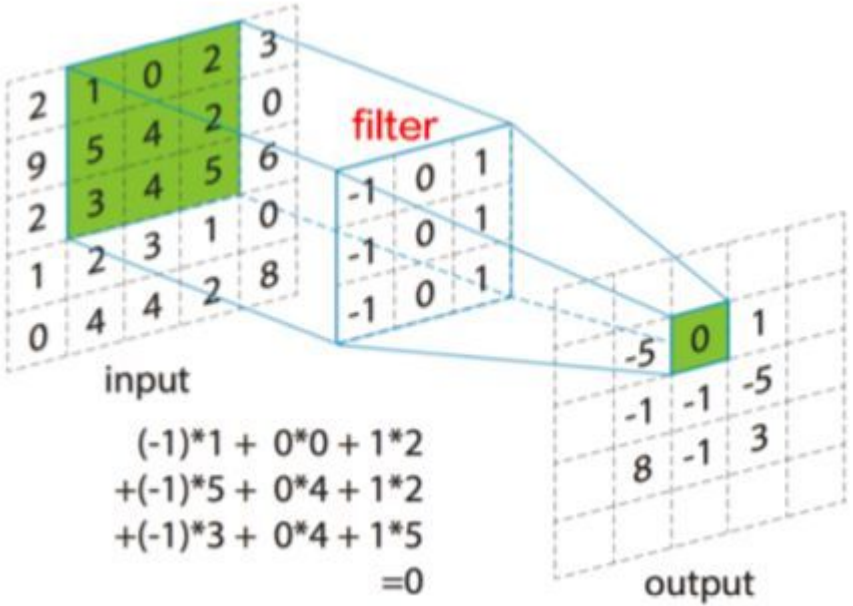
Dense (or fully connected) layers: f(x) = W x + b



Building blocks of deep architectures

Convolutional layers (1D, 2D, 3D)

Stride, padding



Building blocks of deep architectures

Pooling: max pool, average pool, global.

Normalizations: batch norm

Non-linearities: ReLU, tanh, sigmoid

Dropout