

Introduction to Big Data & ML in Remote Sensing

Greg Tsagkatakis

CSD - UOC

ICS – FORTH

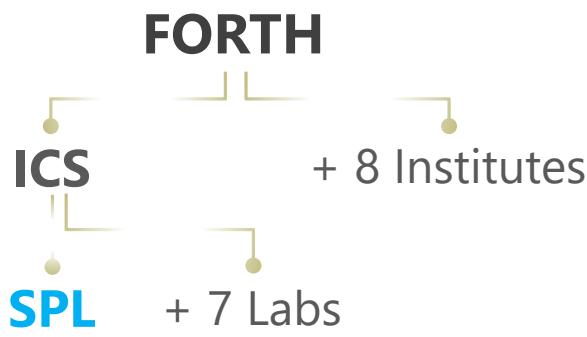
GREG@ICS.FORTH.GR

Who am I?

- BSc & MSc Electronics and Computer Engineering at the University of Crete, Greece (2005, 2007)
- Ph.D. Imaging Science, Rochester Institute of Technology, NY (2011)
- Research associate with Signal Processing Lab at the Foundation for Research and Technology – Hellas (FORTH)
- Marie-Curie fellowship with Prof. Mahta Moghaddam, Electrical and Computer Engineering, USC
- Assistant Professor at the Computer Science Department, University of Crete, Greece
- Research interest: (deep) machine learning for big scientific data analysis



FORTH



⌚ 2006



4 Researchers/Academics
(permanent)

⌚ 1 Postdoctoral Researchers
_wallet 12 Postgraduate Students
👤 2 Research Engineers

> 9M€ (PI/co-PI in EU & national
projects)

Research Institutes



Collaborators



Panos Tsakalides
Head of the Signal
Processing lab



Mahta Moghaddam
Head of Microwave
Systems, Sensors, and
Imaging Lab



Jean-Luc Starck
Head of the CosmoStat lab
French Alternative Energies
and Atomic Energy
Commission



Funding

CALCHAS

*Computational Intelligence for Multi-
Source Remote Sensing Data Analytics*



TITAN

ARTIFICIAL INTELLIGENCE
IN ASTROPHYSICS



Overview

Introduction

- Motivation
- Introduction to Deep Learning (DL)

Discriminative models

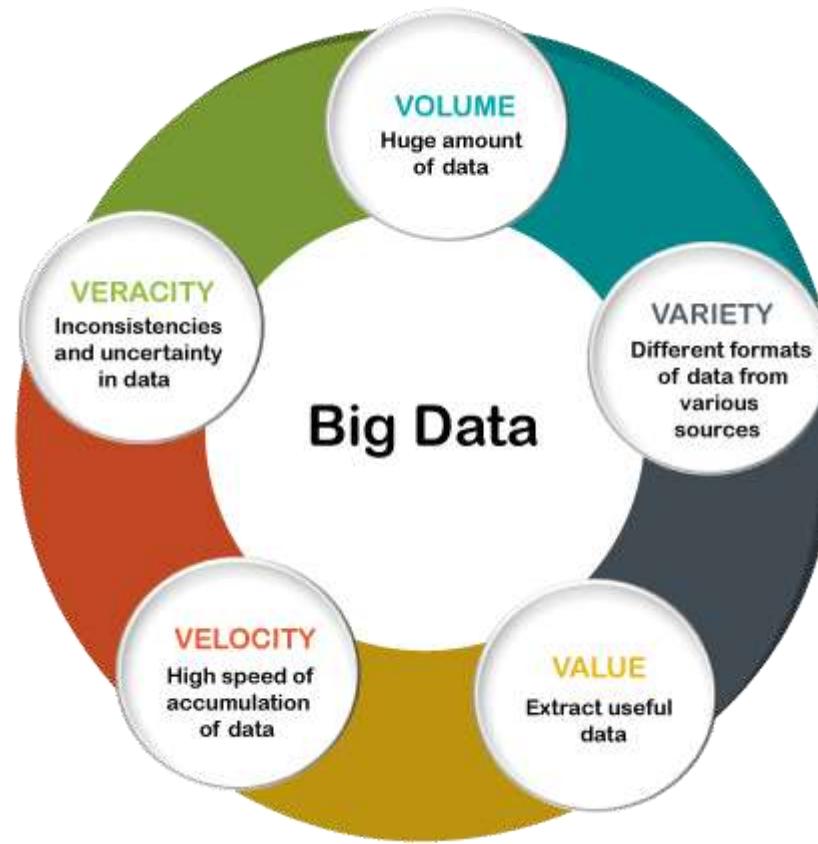
- Classification
- Hands-on

Generative models and inverse problems

- Enhancement
- Hands-on

Other topics

Big Data



The Big Data era

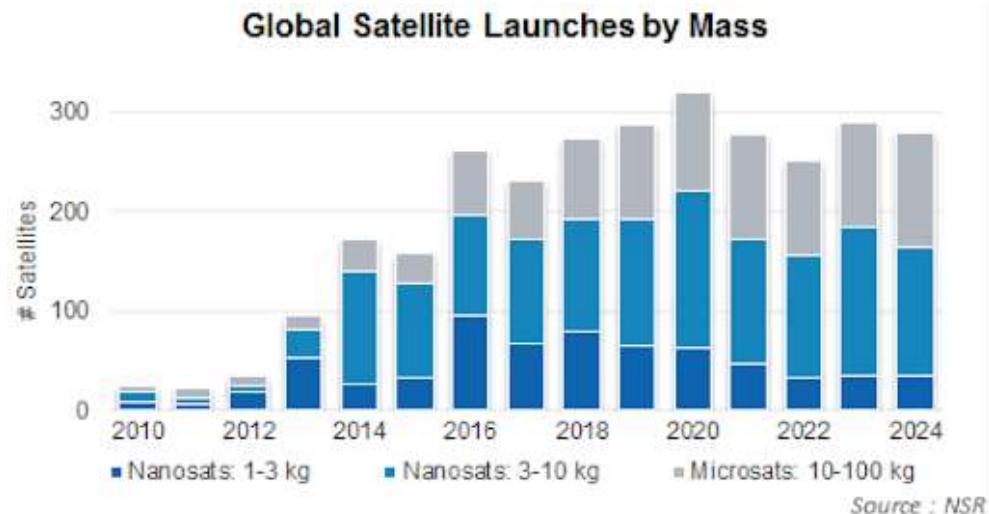
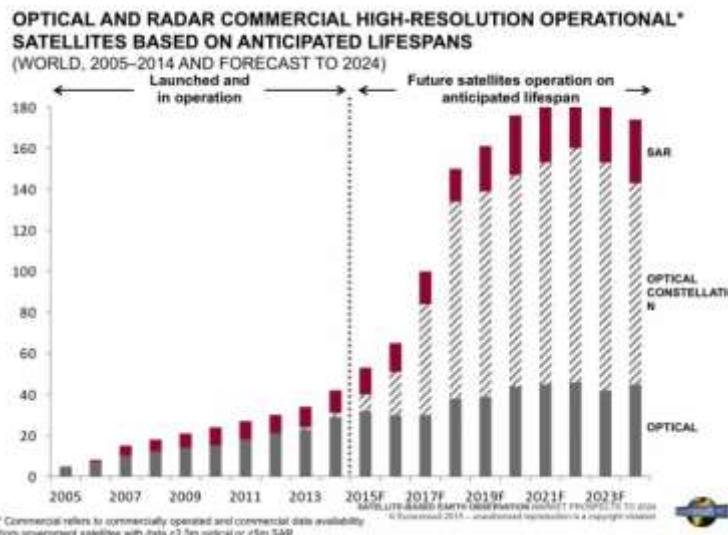


Big Data in Astrophysics

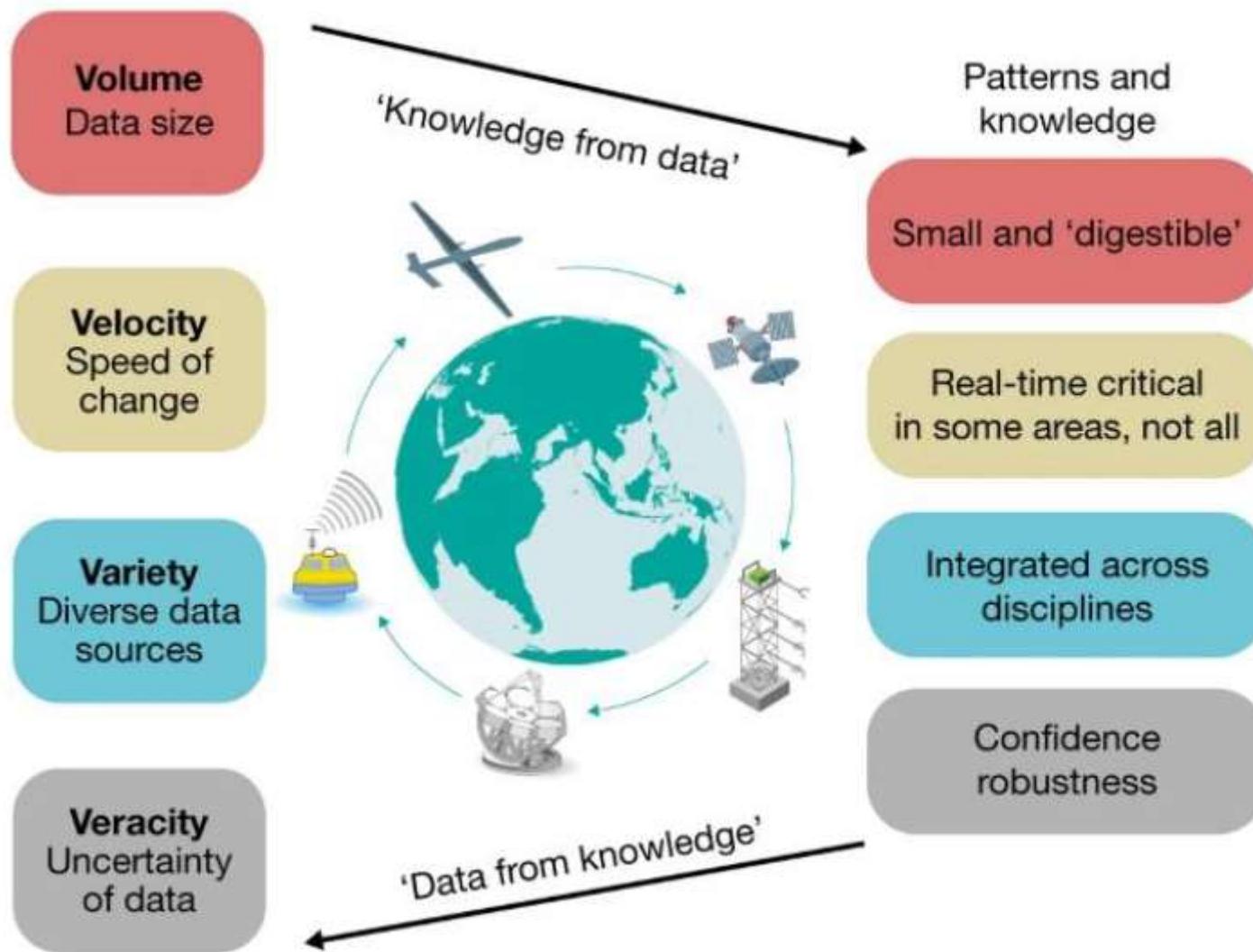
Sky Survey Project	Volume	Velocity	Variety
The Palomar Digital Sky Survey	3 PB		
Sloan Digital Sky Survey (SDSS)	50 TB	200 GB per day	Images, redshifts
Large Synoptic Survey Telescope (LSST)	~ 200 PB	10 TB per day	Images, catalogs
Square Kilometer Array (SKA)	~ 4.6 EB	150 TB per day	Images, redshifts



Big Data in Earth Observation



Observed and
simulated 'big data'

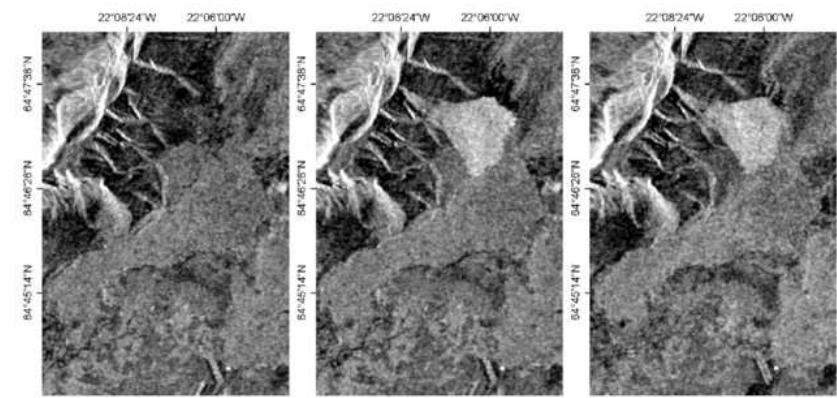
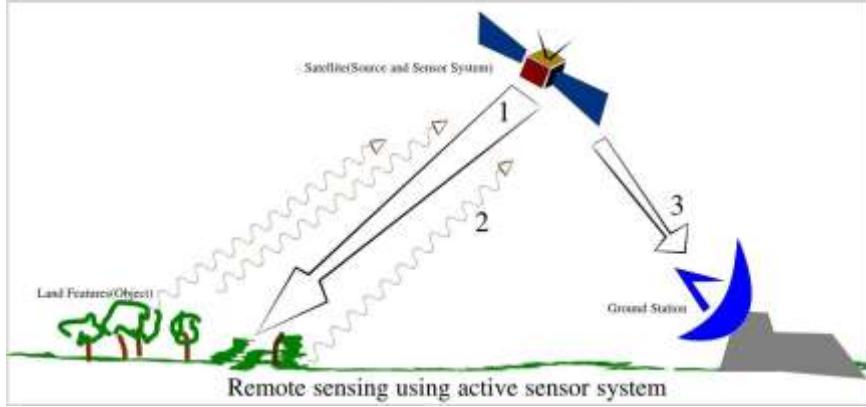
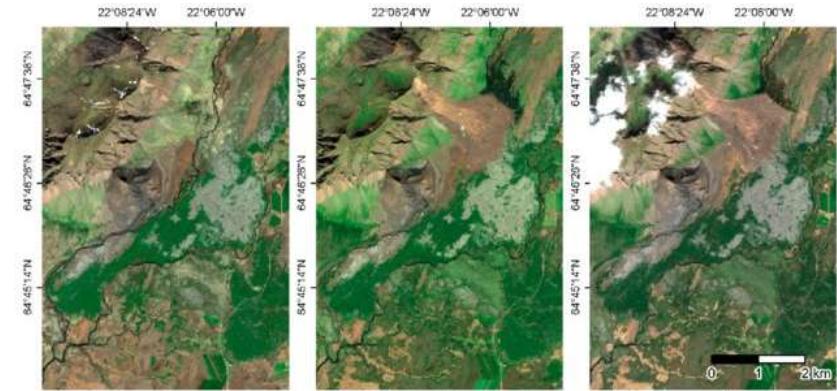
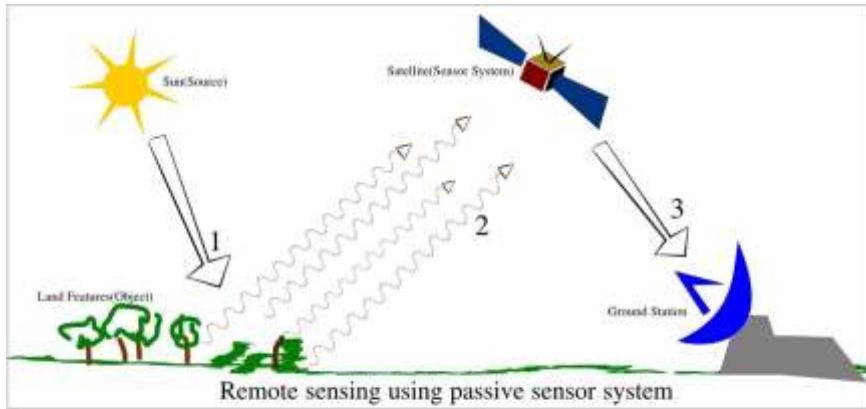


Reichstein et al. (2019), <https://doi.org/10.1038/s41586-019-0912-1>

Types of Earth Observation data

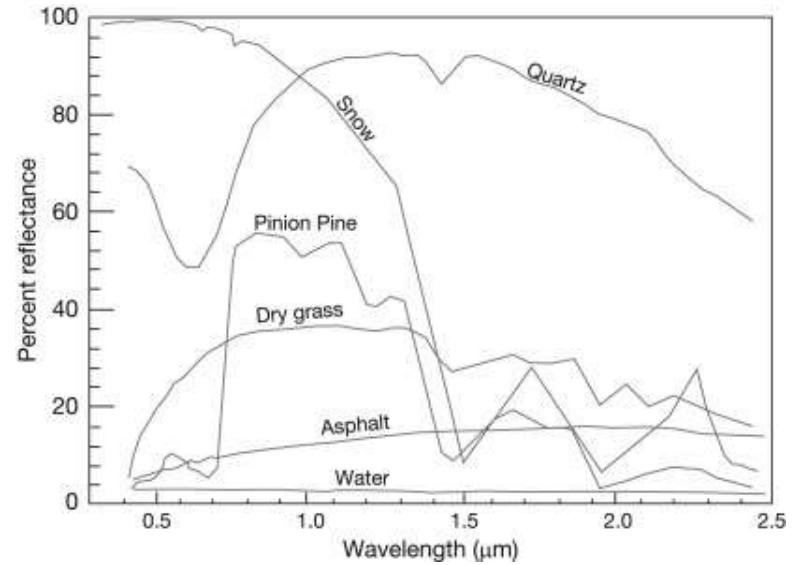
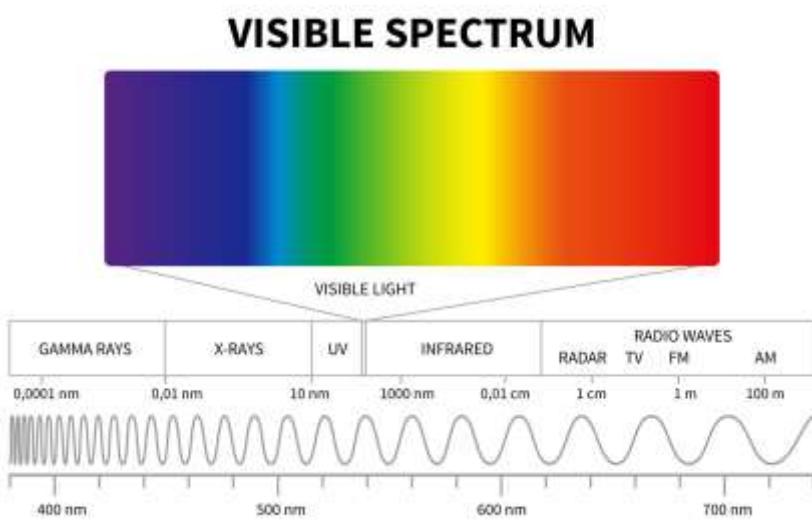
- Active vs Passive imaging
- Color -> Multispectral -> Hyperspectral
- Global coverage vs tasked
- Private vs public/institutional
- Historical vs modern

Active vs Passive sensing



https://en.wikipedia.org/wiki/Remote_sensing

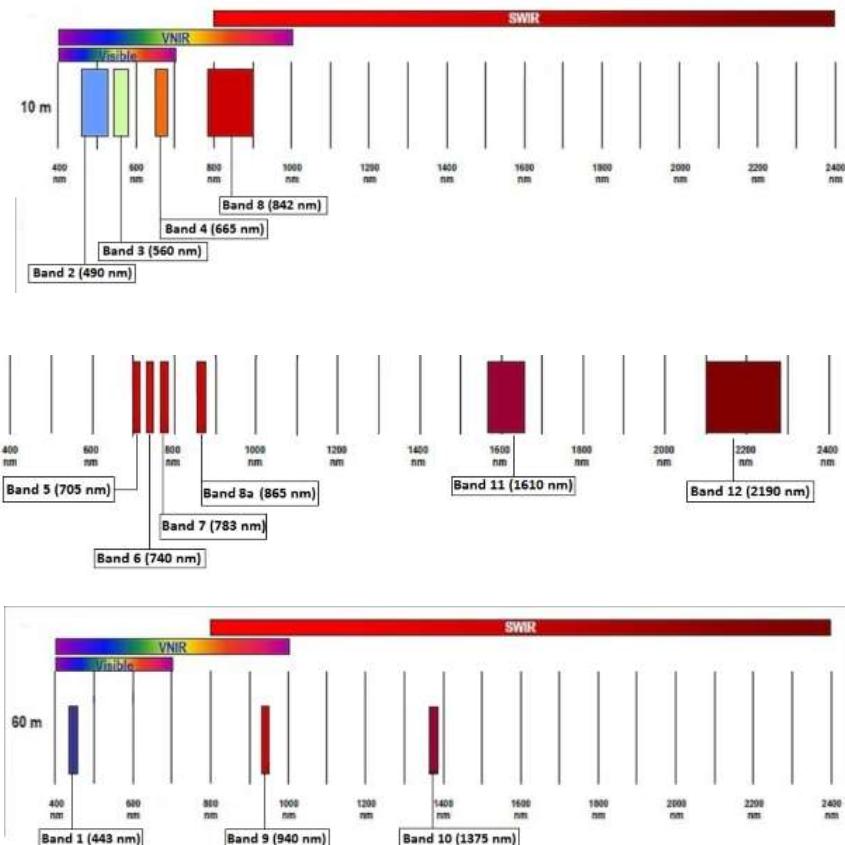
EM spectrum



ESA Sentinel 2

- Optical (Multispectral) imaging
- Global coverage
- Open-access

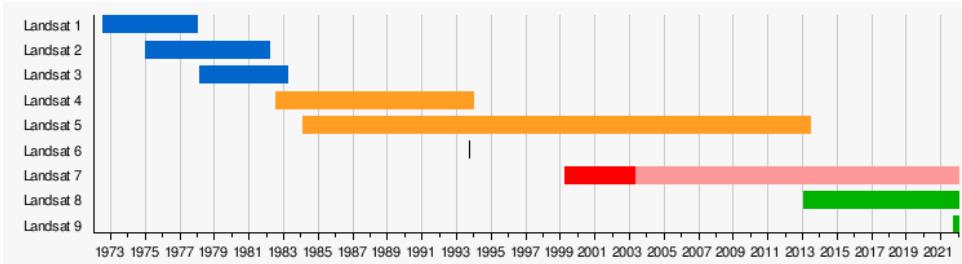
Spatial Resolution	10/20/60m
Spectral Resolution	13 bands
Swath	290km
Revisit	5-days



NASA Landsat 7/8/9

- Optical (Multispectral) imaging
- Global coverage
- Open-access

Spatial Resolution	15/30/100m
Spectral Resolution	11 bands
Swath	185km
Revisit	16-days



Private

An overview of the **Pléiades Neo mission**

What

Pléiades Neo is a constellation of two identical, very high-resolution (VHR) satellites, developed as a follow-on to the Pléiades optical satellites.

The satellites bring unprecedented capability to provide imagery with a high level of detail, including more visibility of small objects, such as vehicles and road markings.

When

Launched on

28 APR
2021

16 AUG
2021

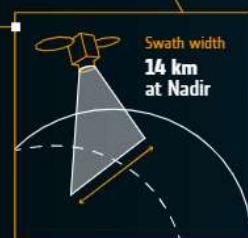
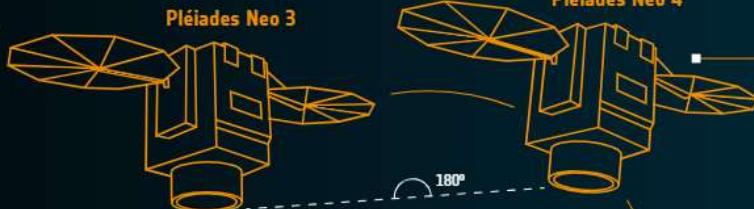
Pléiades Neo 3 Pléiades Neo 4

The two satellites launched in 2021, phased at 180° from each other. Their expected lifetime is ten years

Instruments

The sensors carried onboard the satellites have a spatial resolution up to 30 cm (panchromatic) and 1.2 m multispectral imagery (4 or 6 bands), offering mono, stereo and tri-stereo acquisitions

1 million km²
per day
(half a million per satellite)



Built by

The constellation is funded, owned, manufactured and operated by Airbus Defence & Space



Objectives

The constellation has several remote sensing applications for commercial, institutional and governmental use, including:



- Defence, security and crisis management
- Maritime monitoring – such as oceans protection
- Forestry
- Agriculture
- Urban planning – such as mapping, civil engineering and mobility
- Environment

Innovation

Pléiades Neo can be tasked at any time of the day and up to 15 minutes in advance. It has a revisit capability to any point on the globe at least twice daily

Data access

ESA offers access to Pléiades Neo data and to the OneAtlas Living Library subscription via the TPM programme earth.esa.int/eogateway/catalog/pleiades-neo-full-archive-and-tasking

MODIS

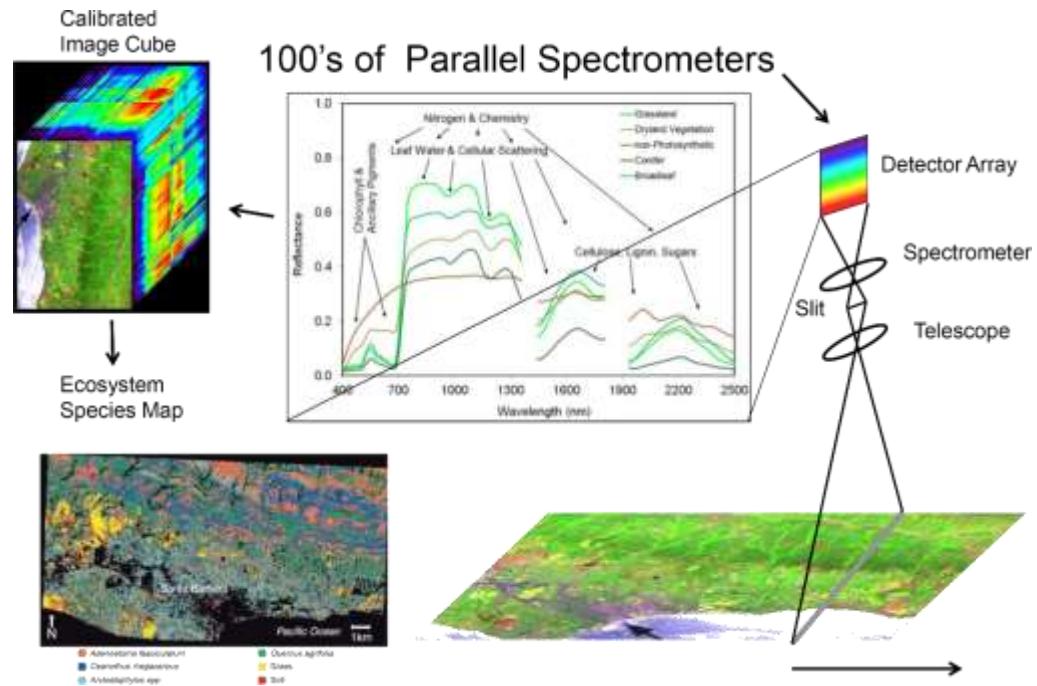
Terra & Aqua Moderate Resolution Imaging Spectroradiometer

Orbit:	705 km, 10:30am descending node (Terra) or 1:30p.m. ascending node (Aqua), sun-synchronous, near-polar, circular
Visit	1-2 days
Swath Dimensions:	2330km (cross track) by 10km (along track at nadir)
Spatial Resolution:	250m (bands 1-2) 500m (bands 3-7) 1000m (bands 8-36)
Wavebands:	36 bands: 1-19 from 405 to 2155nm 20-36 from 3.66 to 14.28 microns
Design Life:	6 years
Duration:	Operational



AVIRIS & AVIRIS-NG

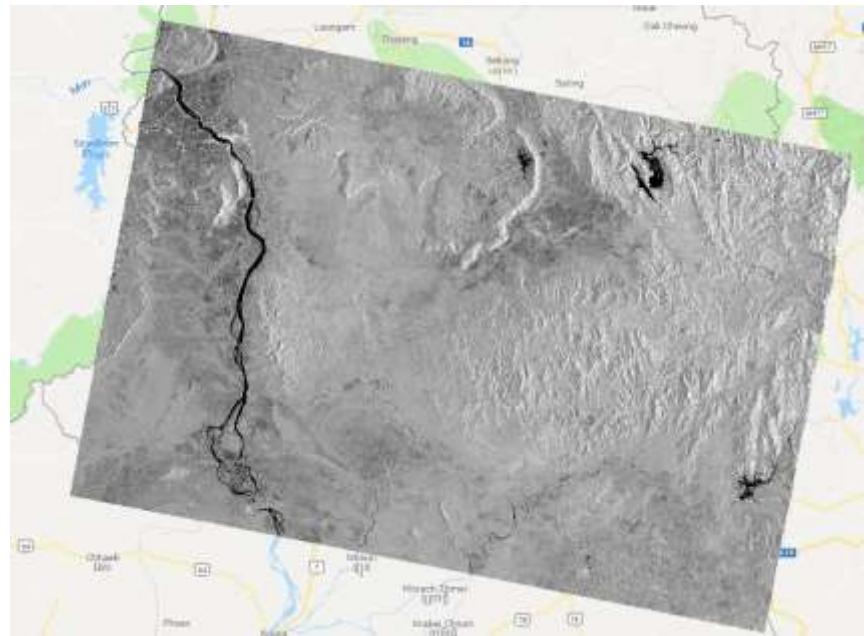
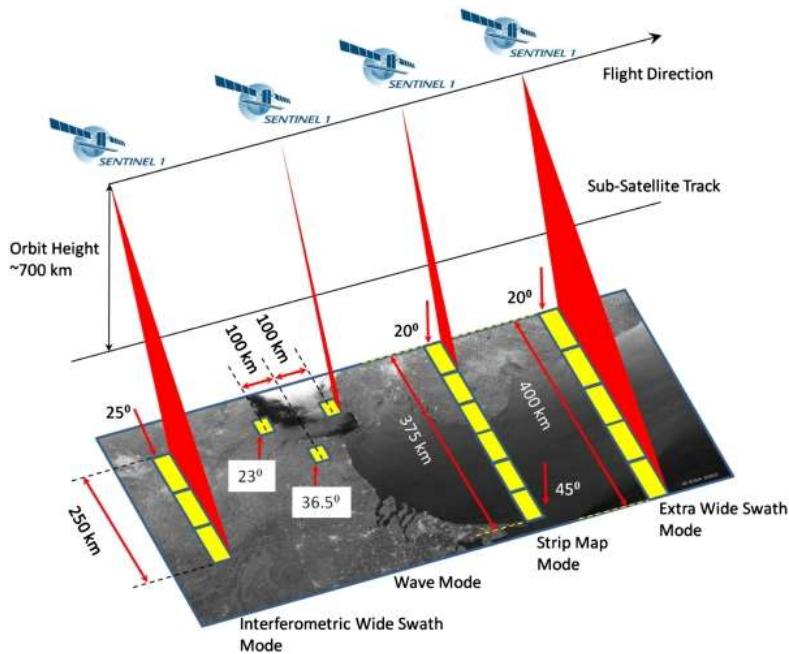
Airborne Visible/Infrared Imaging Spectrometer



ESA Sentinel-1

ESA Sentinel 1 A/B

- C-band radar (5.405 GHz, 4 polarizations)
- 6 days revisit



NASA SMAP mission

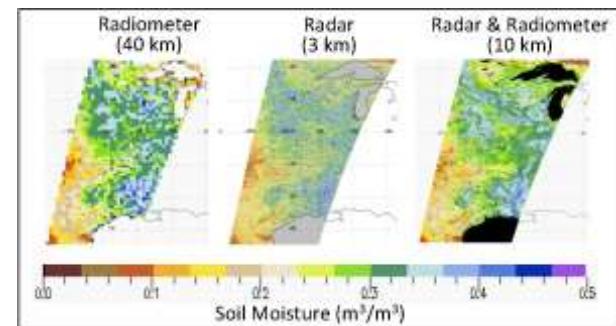
Soil Moisture Active Passive (SMAP) satellite

Objective: global soil moisture every 2-3 days

Operational since April 2015

Instruments @ L-band (1.2 GHz , 1.41 GHz) :

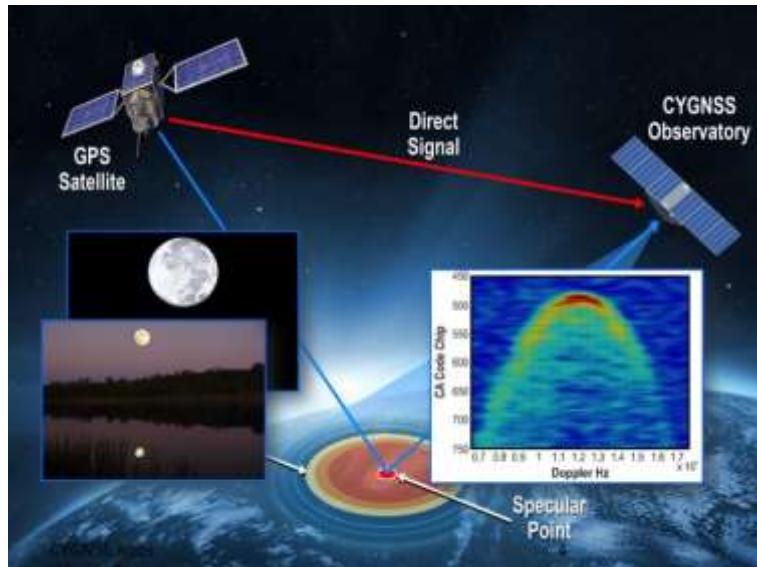
- Radar (active): spatial resolution 3km, high SM uncertainty
- Radiometer (passive): spatial resolution 36km, low SM uncertainty
- Combined: 9km, accuracy $0.04 \text{ cm}^3/\text{cm}^3$



Challenge: Radar failed after 3 months of operation

NASA CYGNSS

- Cyclone Global Navigation Satellite System: 8 satellites in LEO (500 km altitude) at 35° orbit inclination
- Primary objective: Sensing sea level wind speed in tropical cyclones
- Application in Surface Soil Moisture

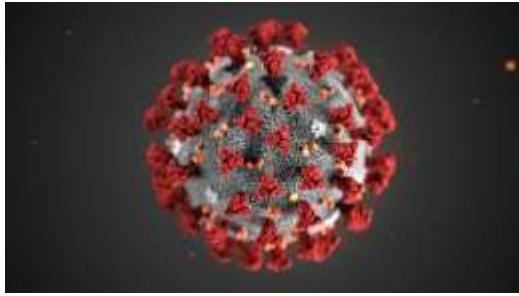


<https://spaceflight101.com/cygnss/>

Advances in Machine Learning

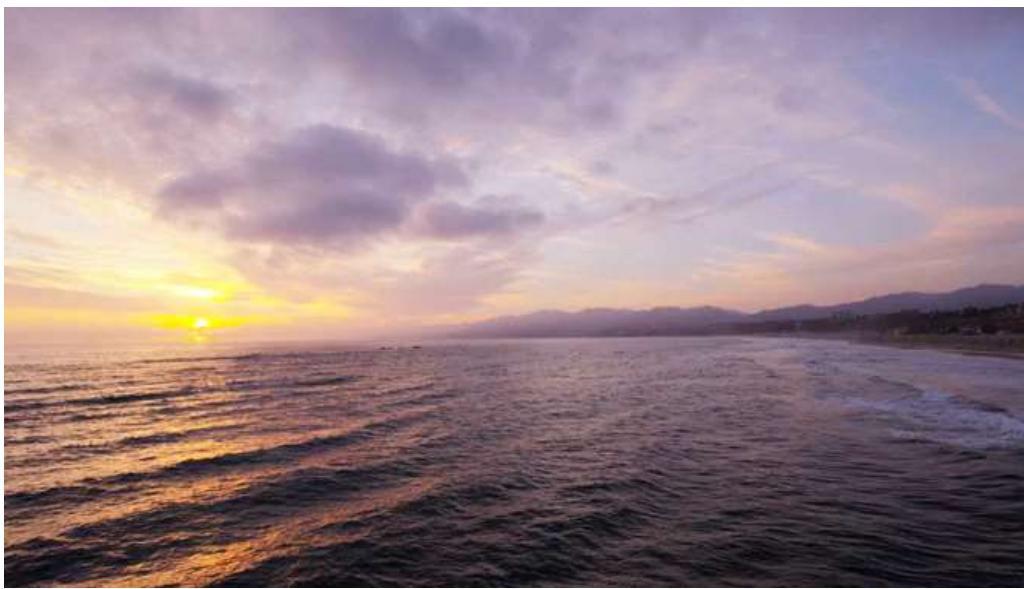


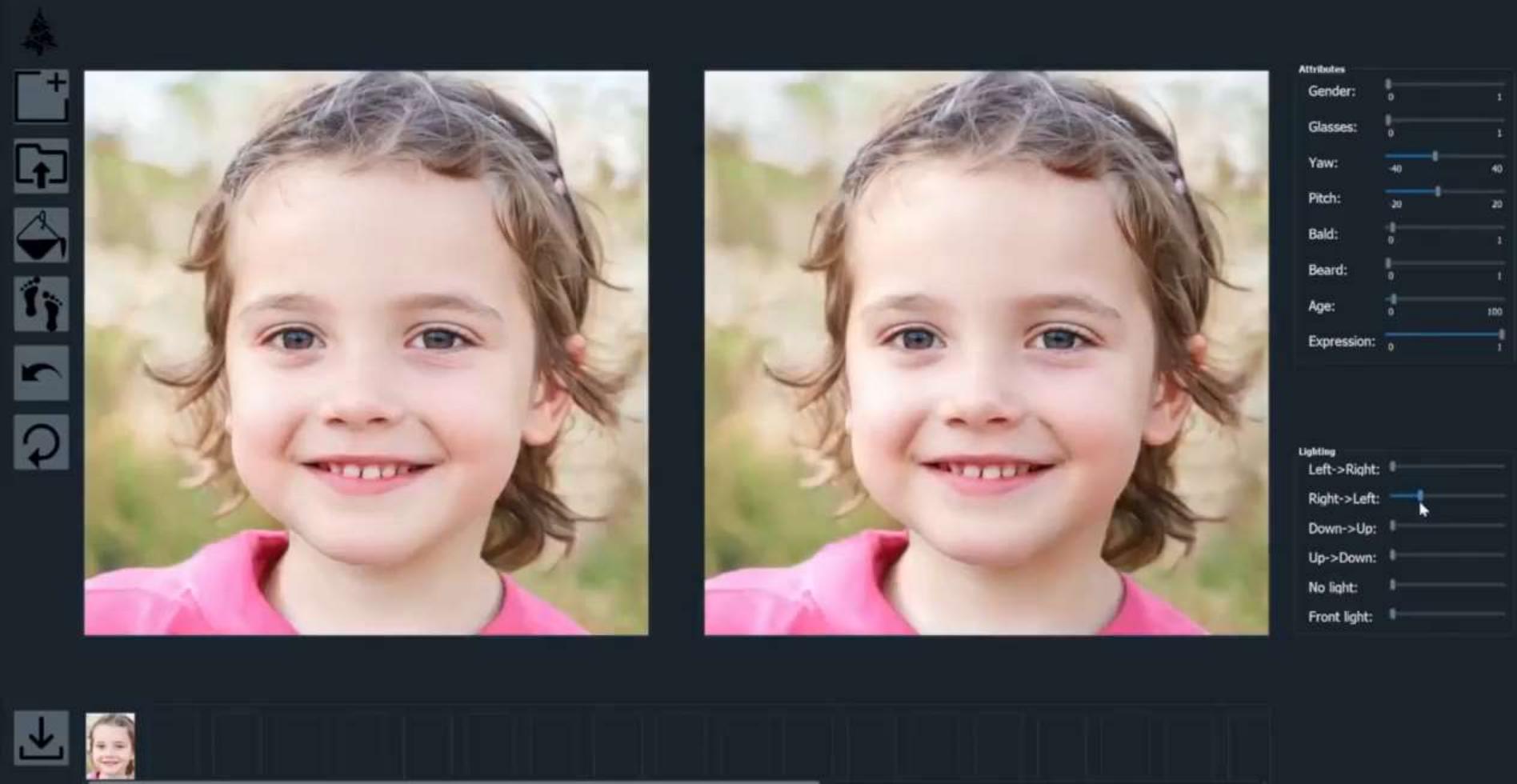
Artificial intelligence-created medicine to be used on humans for first time, BBC news, 30/1/2020



Coronavirus: Can AI Make A Difference?
Forbes, 2/2/2020







Source: [Abdal et al. 2020]

2023



D Make a song about a programmer and someone who isn't a programmer.

Here is a sample song about a programmer and someone who isn't a programmer:

Verse 1:

Programmer, programmer, sitting at the screen
Typing code all day, living in a dream
Creating apps and websites, making them just right
A master of technology, always up to date

Chorus:

Programmer, programmer, so smart and so bright
Always coding away, day and night

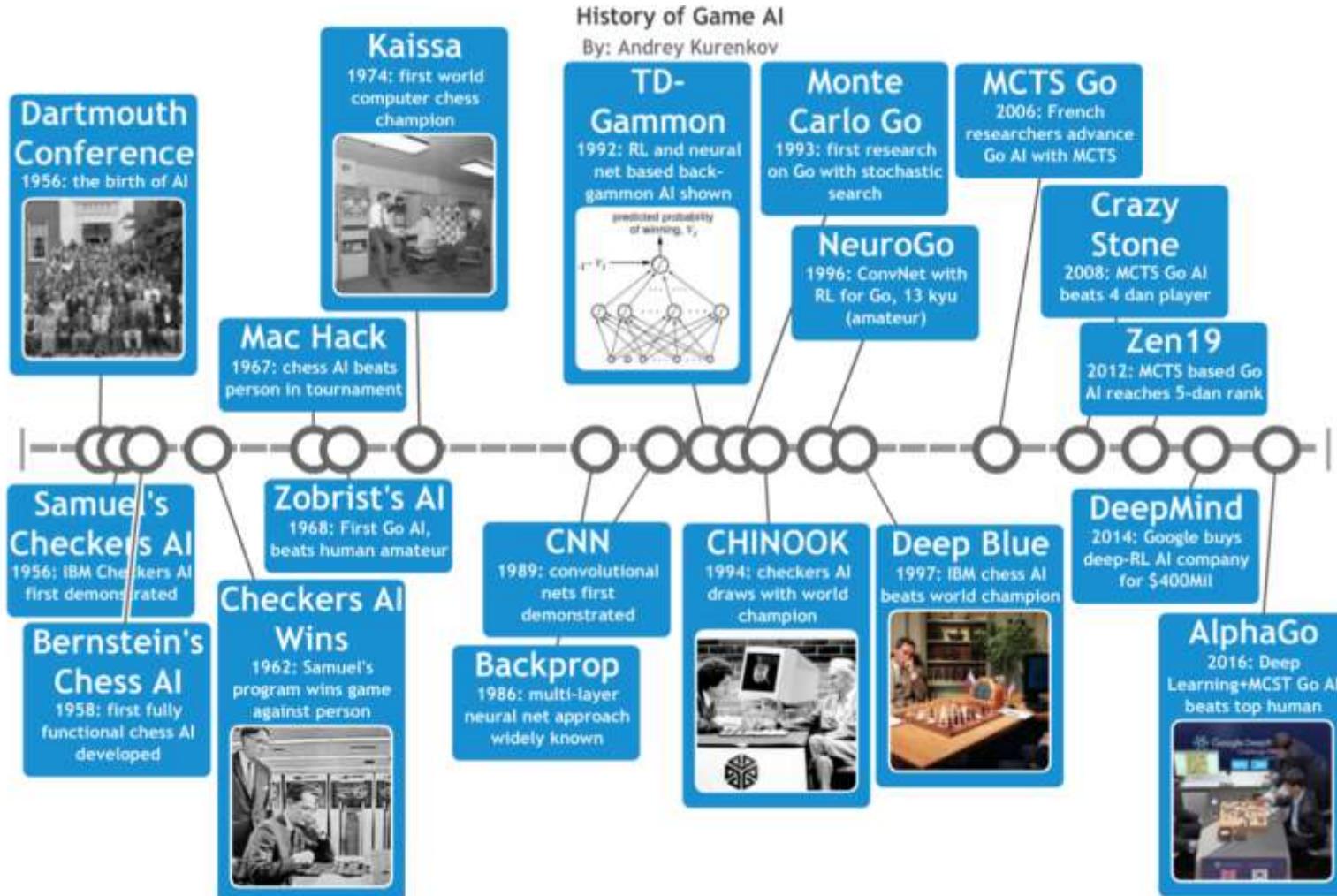
Verse 2:

Non-programmer, non-programmer, not a clue
About the world of code, what it can



“An astronaut riding a horse in a photorealistic style”

Brief history of ML/AI



Types of ML

Supervised learning

- Classification
- Regression

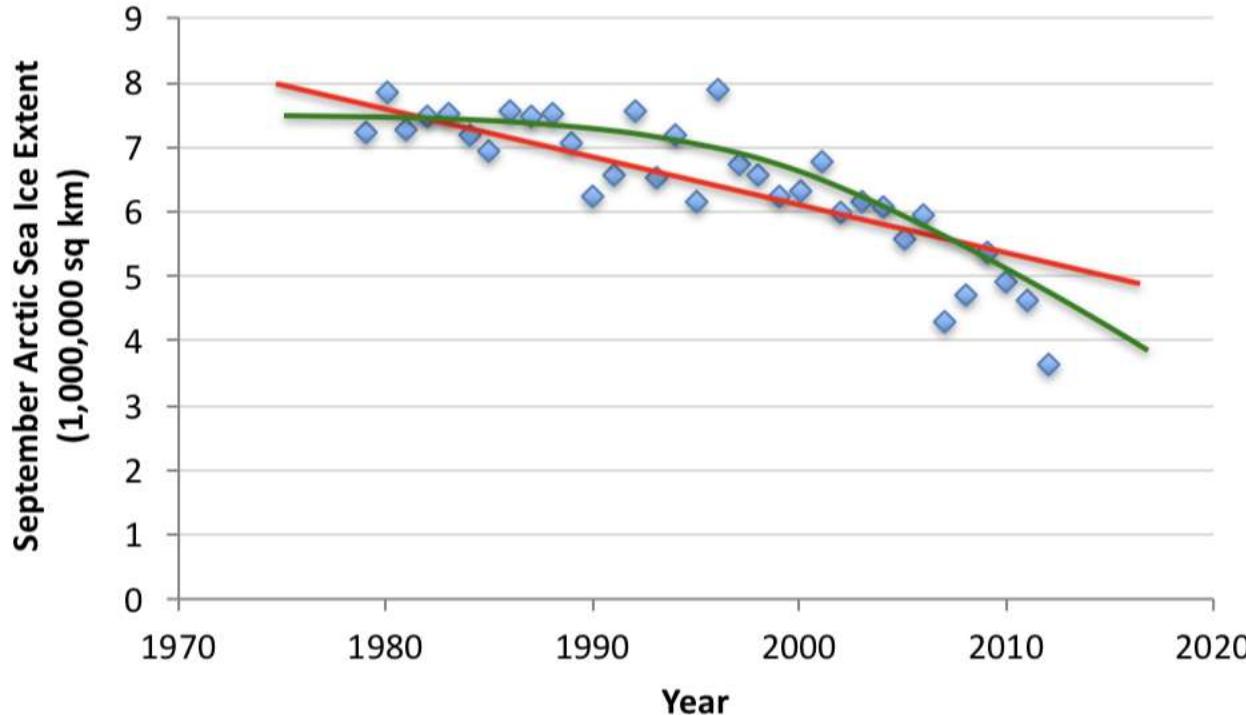
Unsupervised learning

- Clustering
- Generative models
- Inverse problems

Reinforcement learning

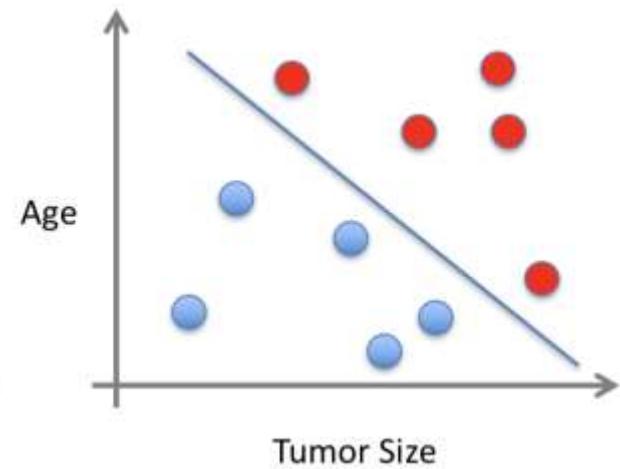
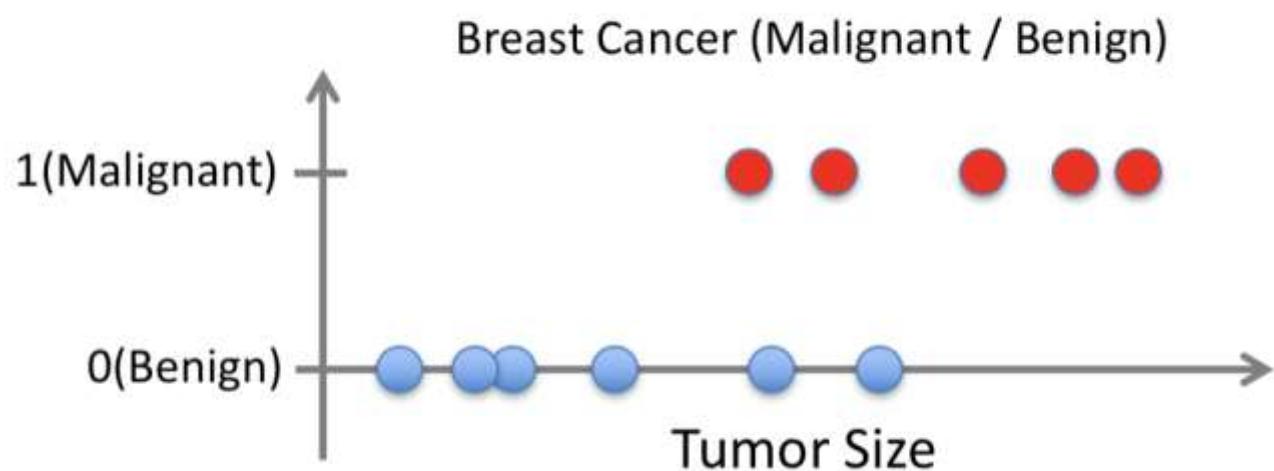
Supervised Learning: Regression

Given $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$: Learn a function $f(x)$ to predict real-valued y given x



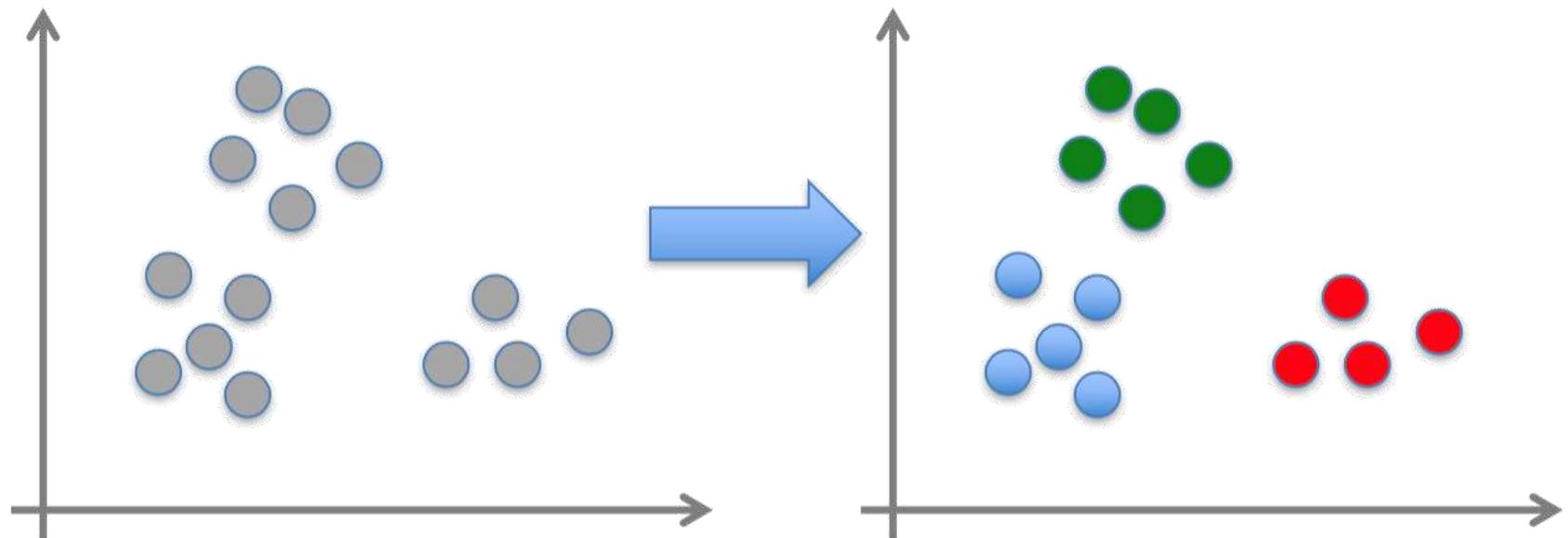
Supervised Learning: Classification

Given $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$: Learn a function $f(x)$ to predict categorical y given x



Unsupervised Learning

- Given x_1, x_2, \dots, x_n (without labels)
- Output hidden structure behind the x 's
 - E.g., clustering

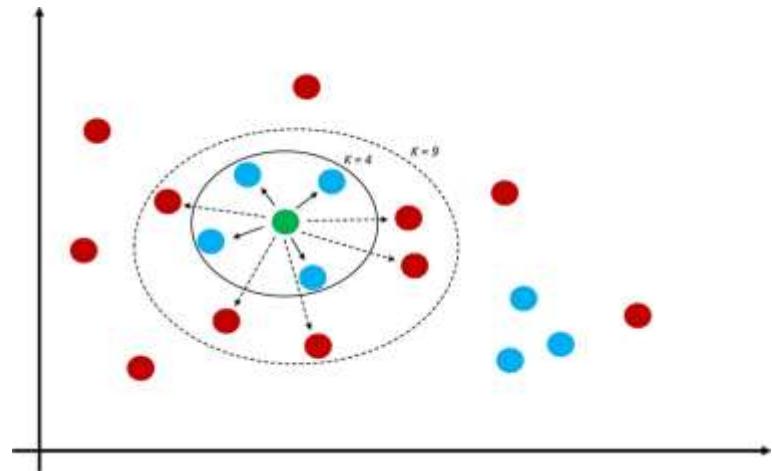


KNN classification

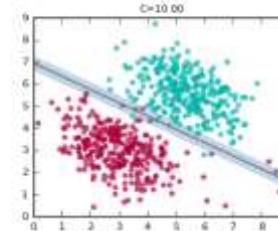
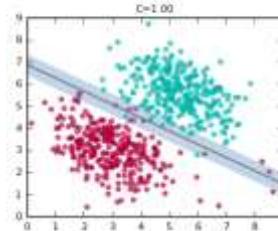
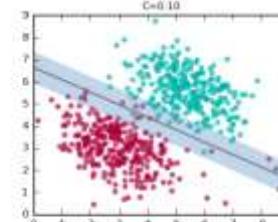
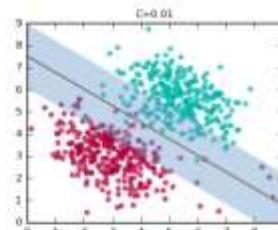
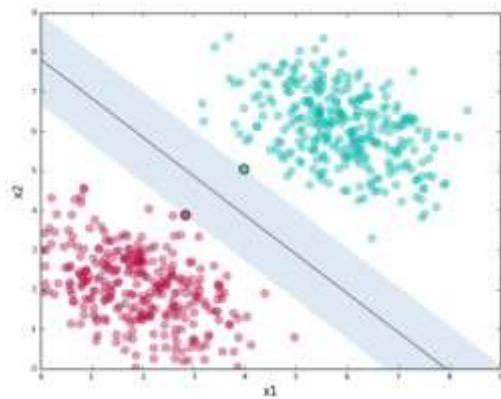
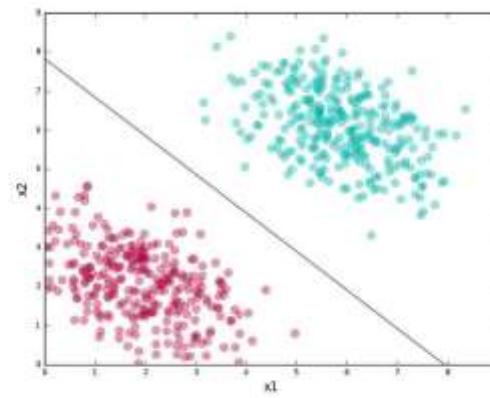
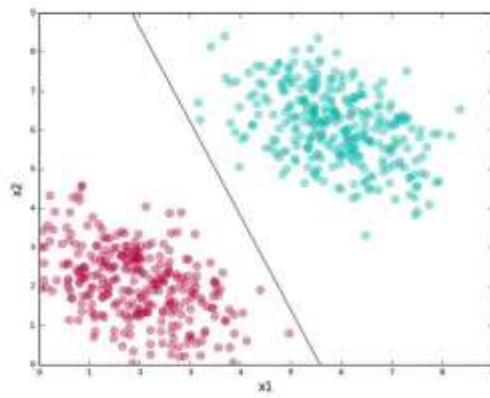
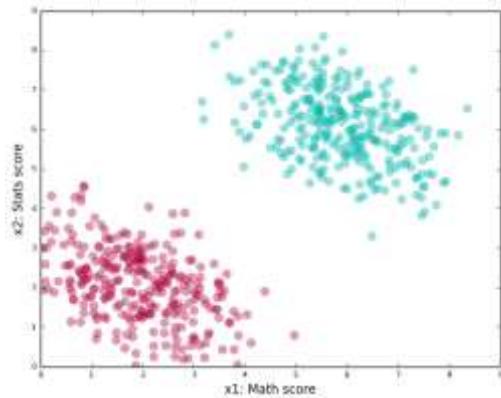
Use neighbors to infer class

Challenges

- Preprocessing & Feature extraction
- Impact of dimensionality
- Sensitivity to the value of k
- Real-time application

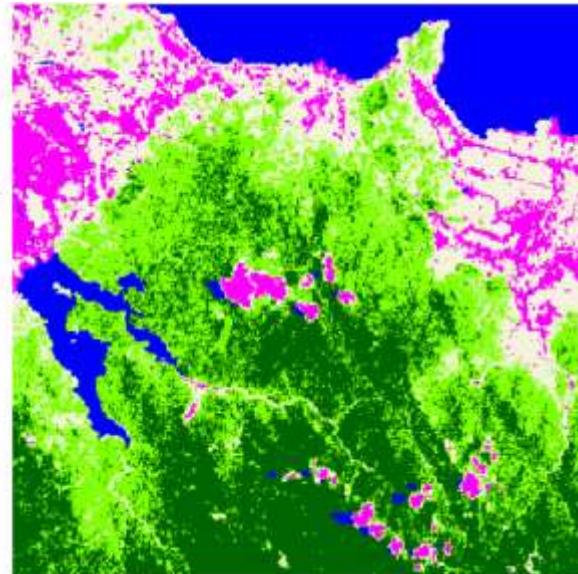
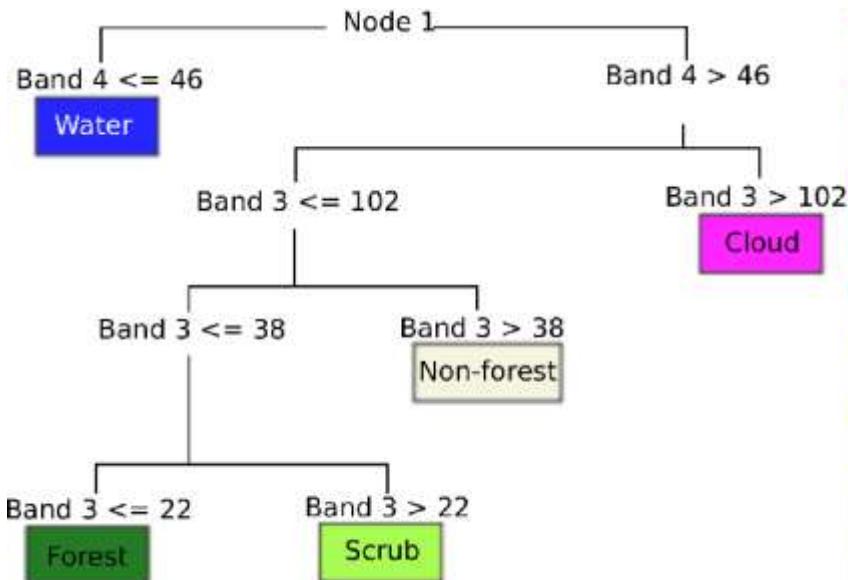


SVM

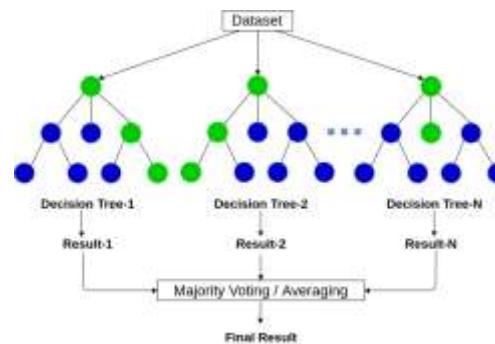


<https://medium.com/cube-dev/support-vector-machines-tutorial-c1618e635e93>

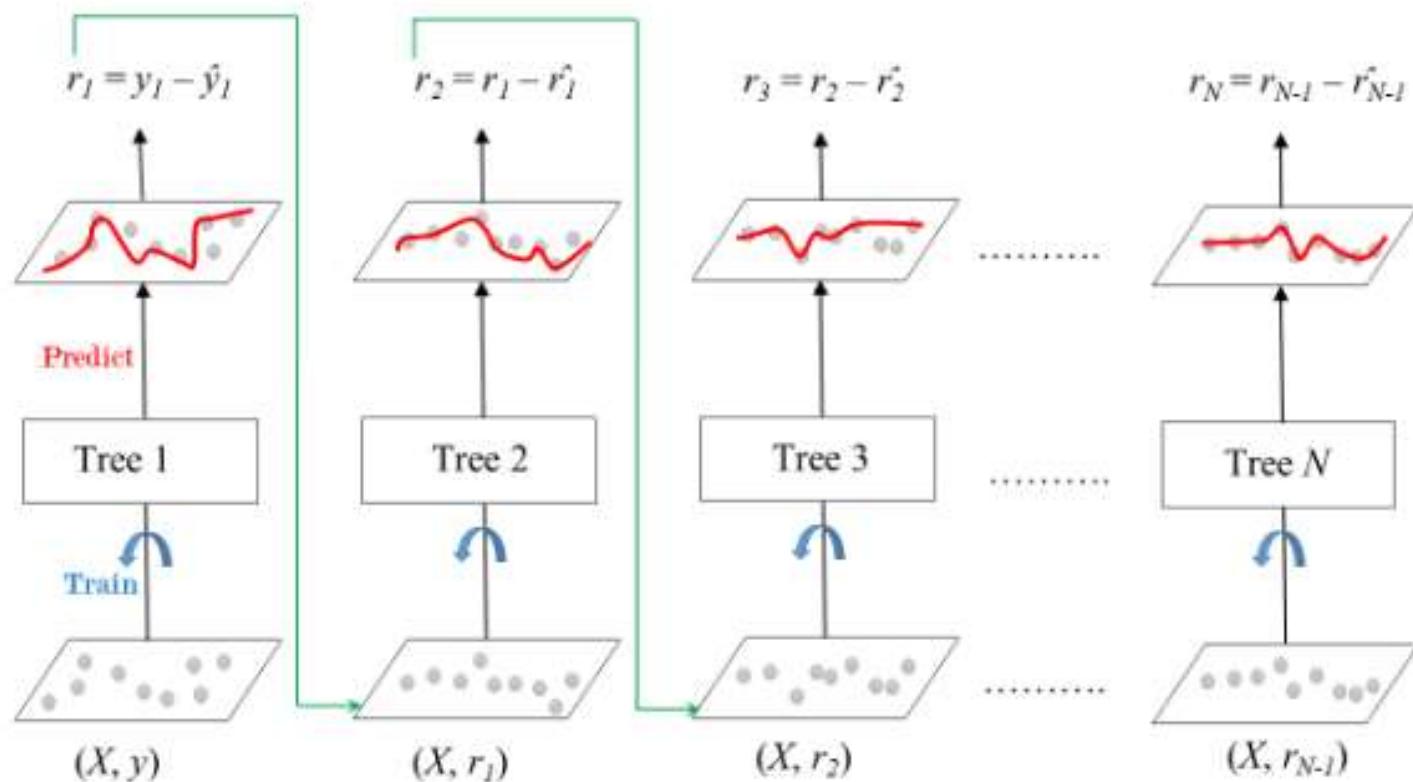
Trees & Forests



Random Forests : An algorithm for image classification and generation of continuous fields data sets

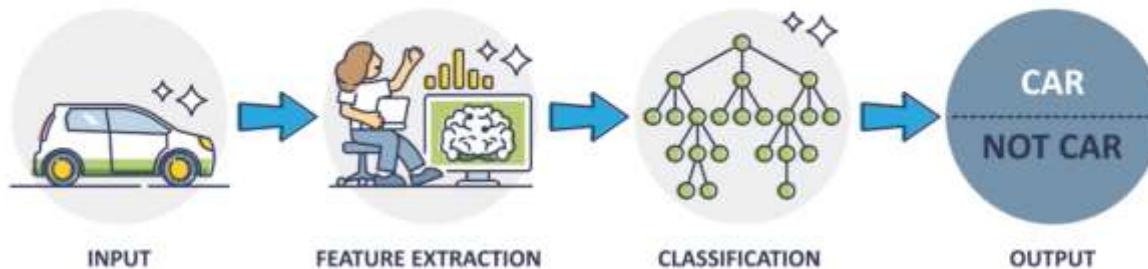


Gradient boosting



Deep Learning

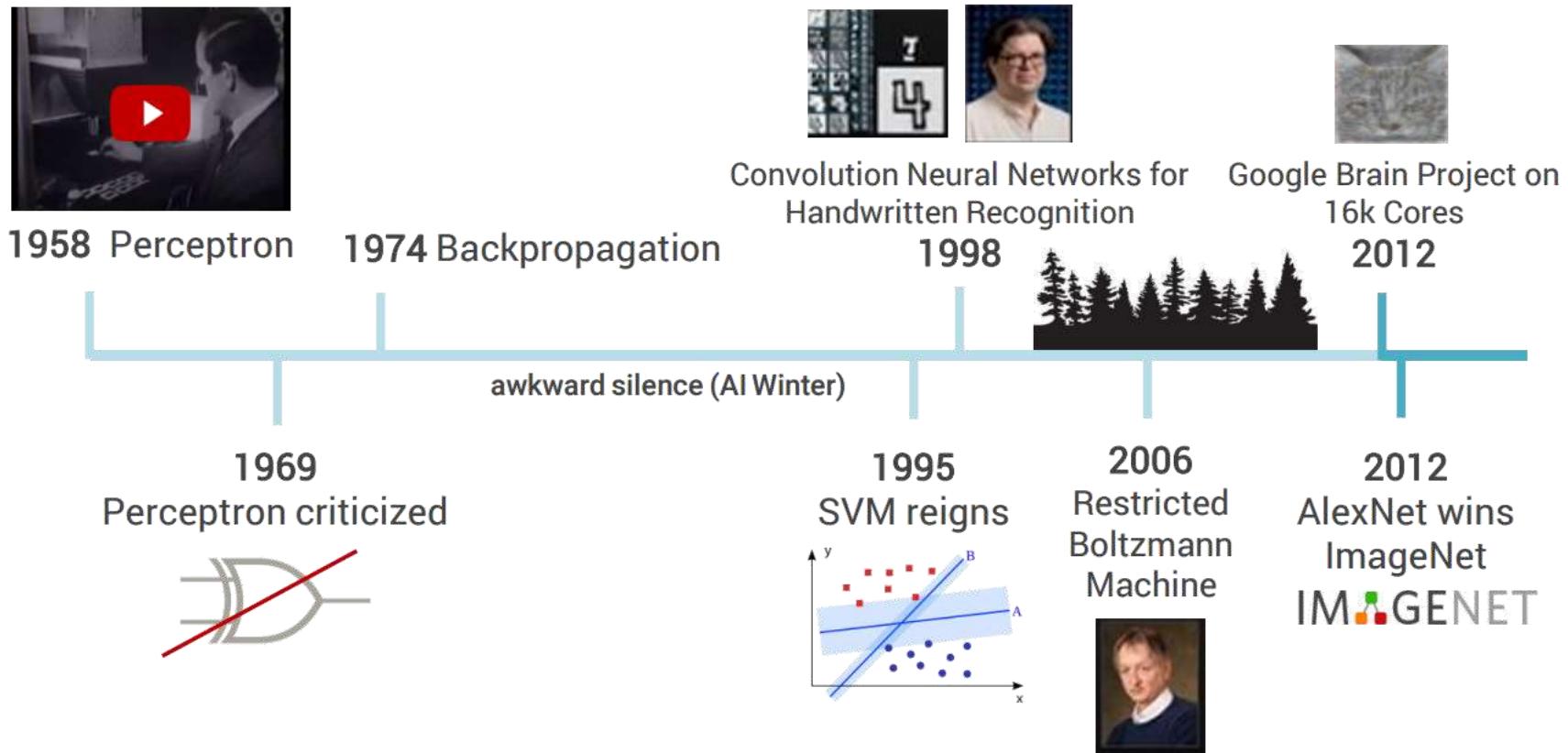
MACHINE LEARNING



DEEP LEARNING



Brief history of DL

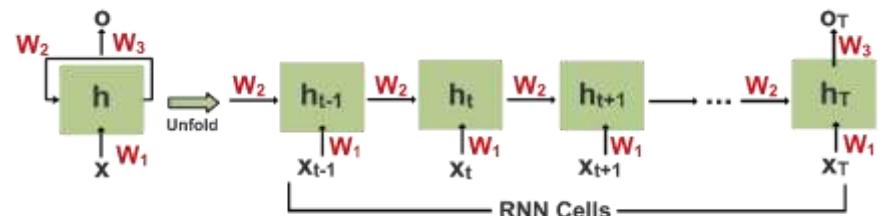


State-of-the-art

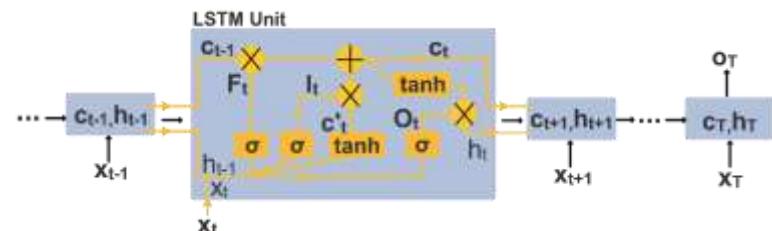
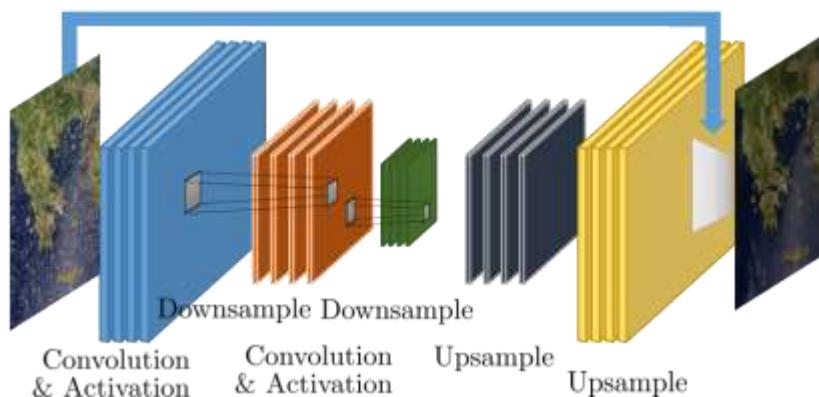
CNNs



RNN/LSTMs



Residual



Transformers



Applications in RS

Land Use/Land Cover Classification	Infrastructure Monitoring
Object Detection	Oceanographic Features Analysis
Change Detection	Species Distribution Modelling
Weather and Climate Prediction	Iceberg Detection
Agricultural Crop Yield Prediction	Forest Fire Detection
Disaster Damage Assessment	Sea Pollution Monitoring
Vegetation and Biomass Estimation	Earthquake Damage Assessment
Urban Planning and Development	Flood Mapping and Assessment
Surface Water Mapping	Atmosphere Gas Concentration
Soil and Land Degradation Assessment	Marine Traffic Monitoring

Scene classification - Aerial



$$x \in \mathbb{R}^{m \times m \times 3} \rightarrow y \in \{0, 1, 2, \dots k\}$$

AID: A Benchmark Dataset for Performance Evaluation of Aerial Scene Classification

EUROSAT

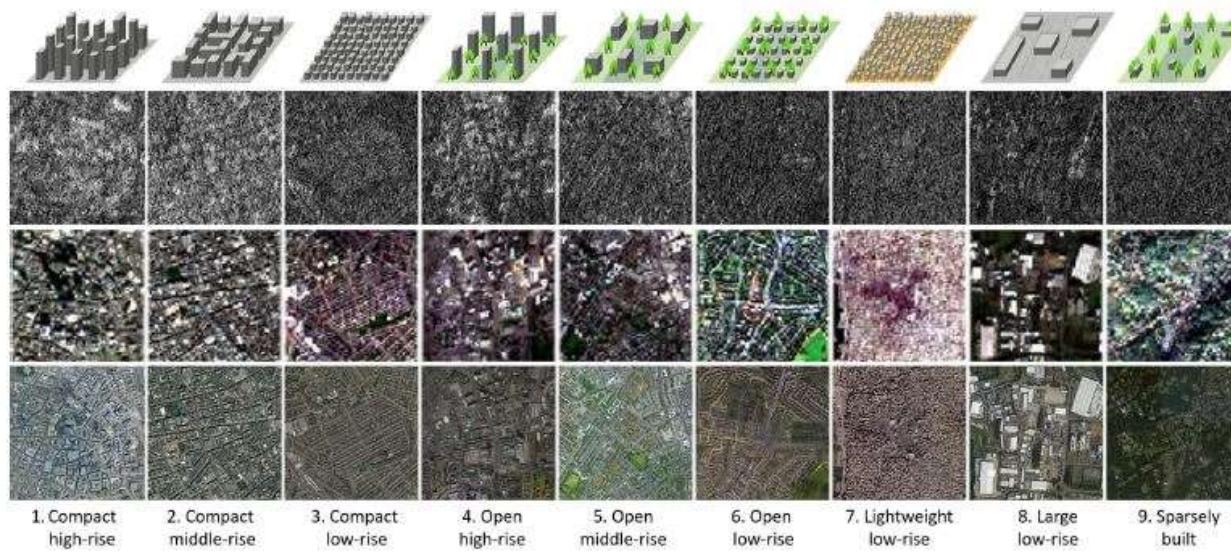
- A land use / land cover classification on Sentinel-2 satellite images.
- Each sample is 64x64 pixels and is associated with one out of 10 classes.



$$x \in \mathbb{R}^{m \times m \times 12} \rightarrow y \in \{0, 1, 2, \dots k\}$$

So2SAT LCZ42

- Co-registered SAR (Sentinel 1) and multispectral (Sentinel 2) patches
- Local climate zones (LCZ) over 42 cities across different regions.
- Each sample is 32x32 pixels and is associated with one out of 17 classes



$$x_1 \in \mathbb{R}^{m \times m \times 12}, x_2 \in \mathbb{R}^{m \times m \times 2} \rightarrow y \in \{0, 1, 2, \dots k\}$$

BigEarthNet

- 590,326 Sentinel-2 image patches
- Each patch is associated with multiple labels (43 imbalanced labels).
- Each example is 120x120 pixels at 10 m (atmospherically corrected).
- Annotation based on L3-hierarchy of Corine Land Cover (CLC) 2018.



permanently irrigated land,
sclerophyllous vegetation,
beaches, dunes, sands,
estuaries, sea and ocean



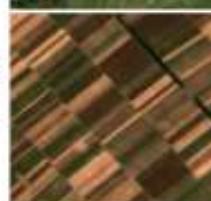
permanently irrigated land,
vineyards, beaches, dunes,
sands, water courses



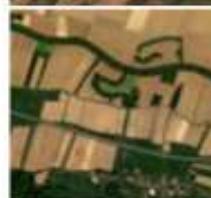
coniferous forest, mixed
forest, water bodies



non-irrigated arable land,
fruit trees and berry
plantations, agro-forestry
areas, transitional
woodland/shrub



non-irrigated arable land



discontinuous urban fabric,
non-irrigated arable land,
land principally occupied
by agriculture,
broad-leaved forest

$$x \in \mathbb{R}^{m \times m \times 12} \rightarrow y \in \{0\}, \{0, 1\}, \dots, \{1, 2\}, \dots, \{1, 7, k\}$$

Object detection



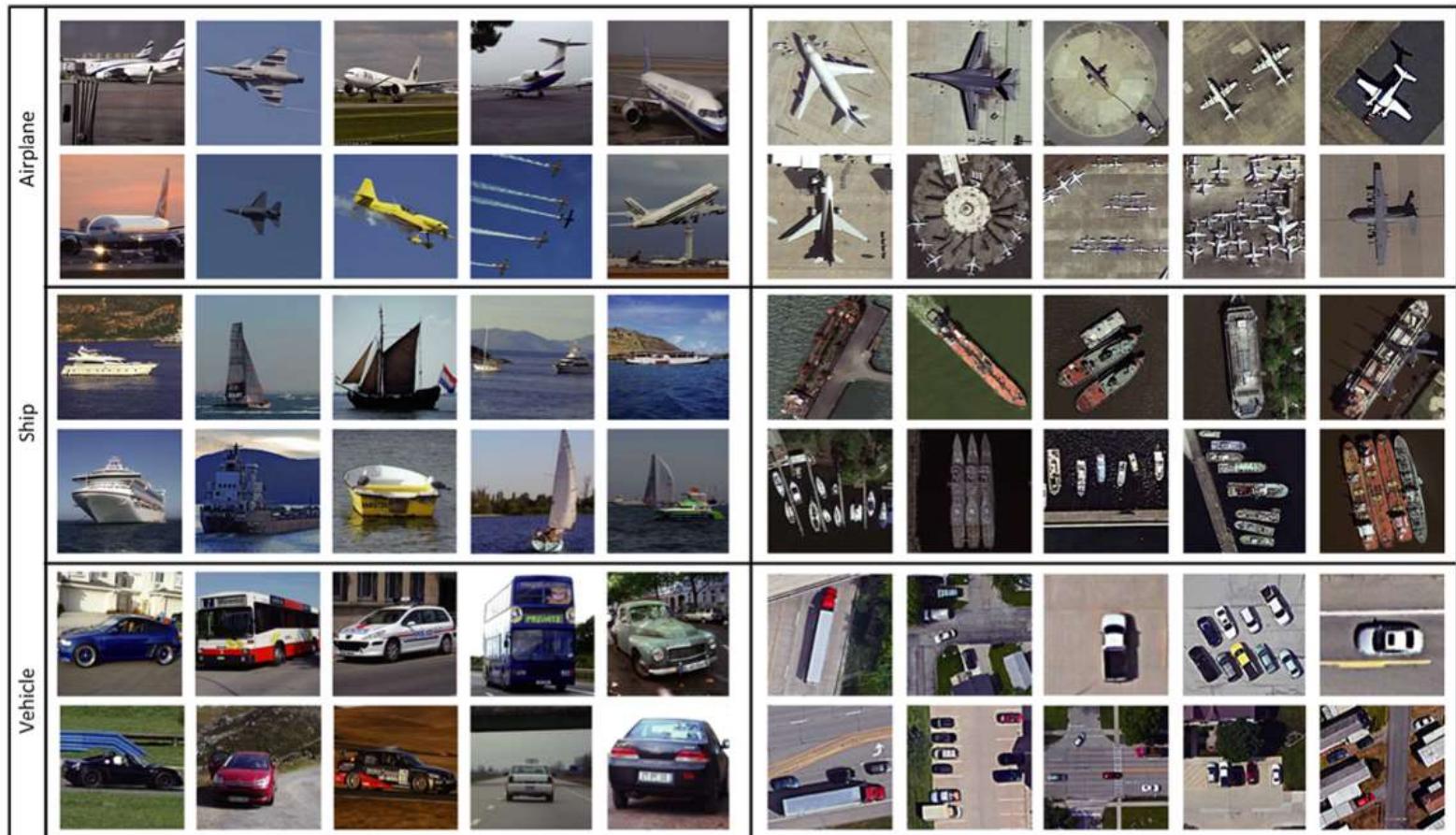
(a) Horizontal object detection.



(b) Rotated object detection.

$$x \in \mathbb{R}^{m \times m \times 12} \rightarrow \{BB_1, C_1\}, \dots, \{BB_n, C_n\}$$

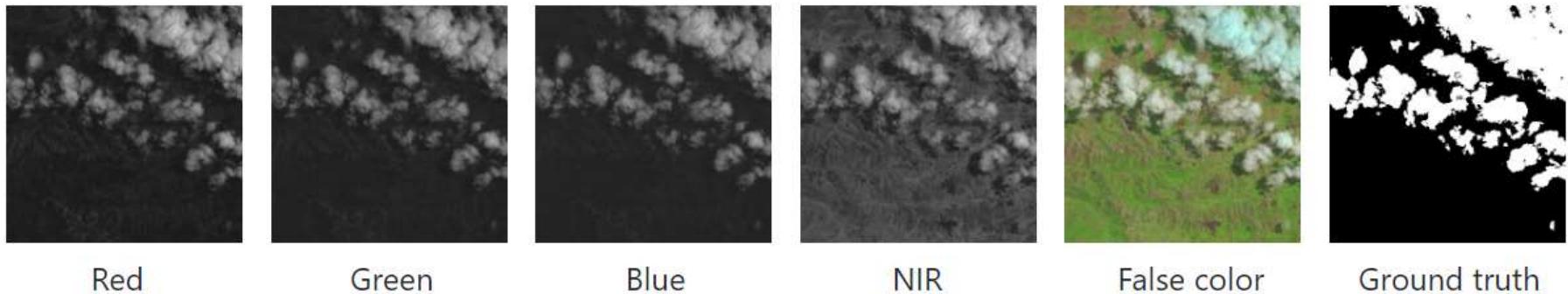
Object detection



Object detection in optical remote sensing images: A survey and a new benchmark

Cloud detection

- 38 Landsat 8 scene images
- 8400 patches for training and 9201 patches for testing.



Red

Green

Blue

NIR

False color

Ground truth

Object/target recognition

- Snow cover



- Oil spills



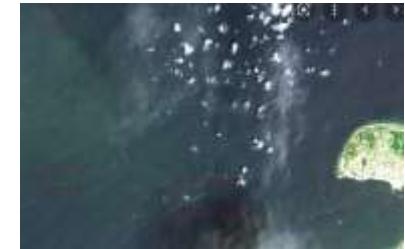
- Solar PV parks



- Active fire



- Floating plastic litter
(distribution)



- Volcanic eruption



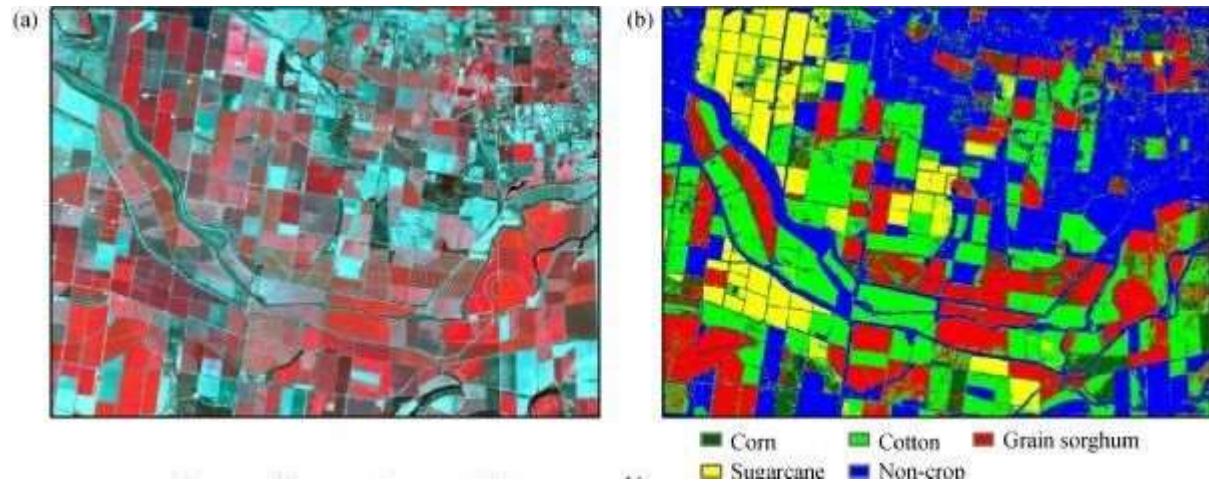
Image segmentation



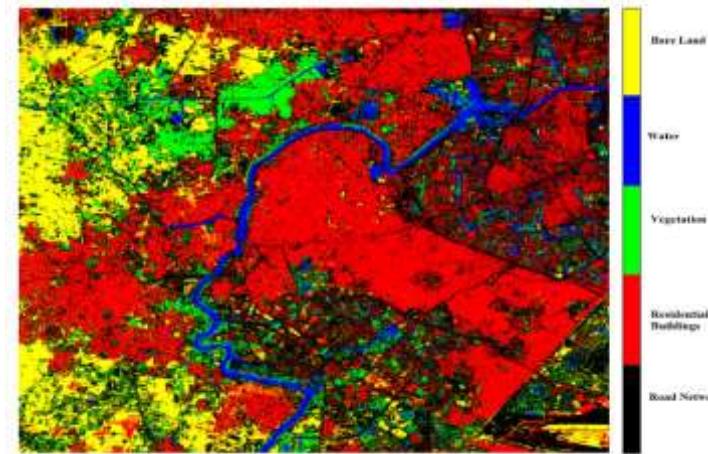
$$x \in \mathbb{R}^{m \times m \times 3} \rightarrow y \in \mathbb{Z}_k^{m \times m}$$

Land classification

Crop type



Land cover



Change detection

Given two instances -> find changes

Challenges

- Atmospheric conditions
- Resolution
- Modality
- Lack of ground truth



Image denoising

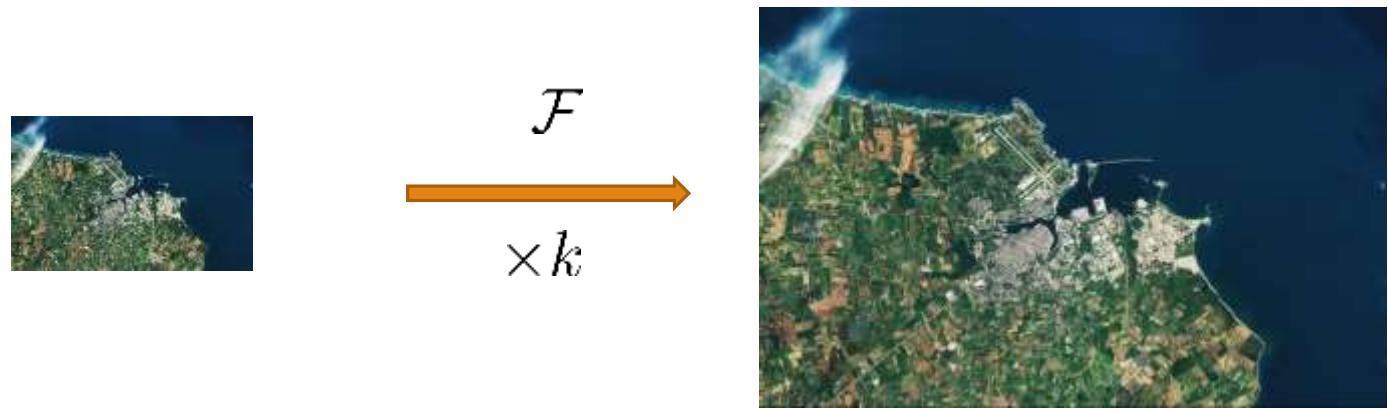


$$\xrightarrow{\mathcal{F}} +n$$



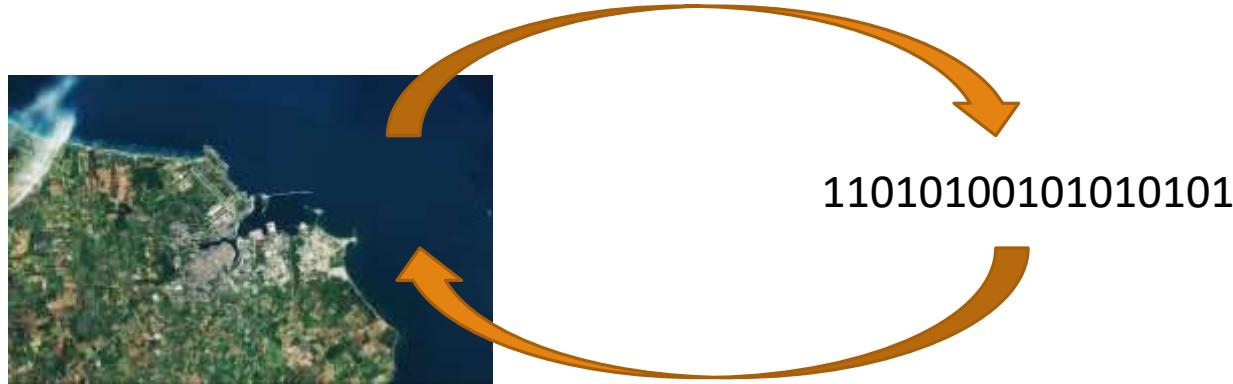
$$x \in \mathbb{R}^{m \times m \times b} \rightarrow y \in \mathbb{R}^{m \times m \times b}$$

Image Super-Resolution



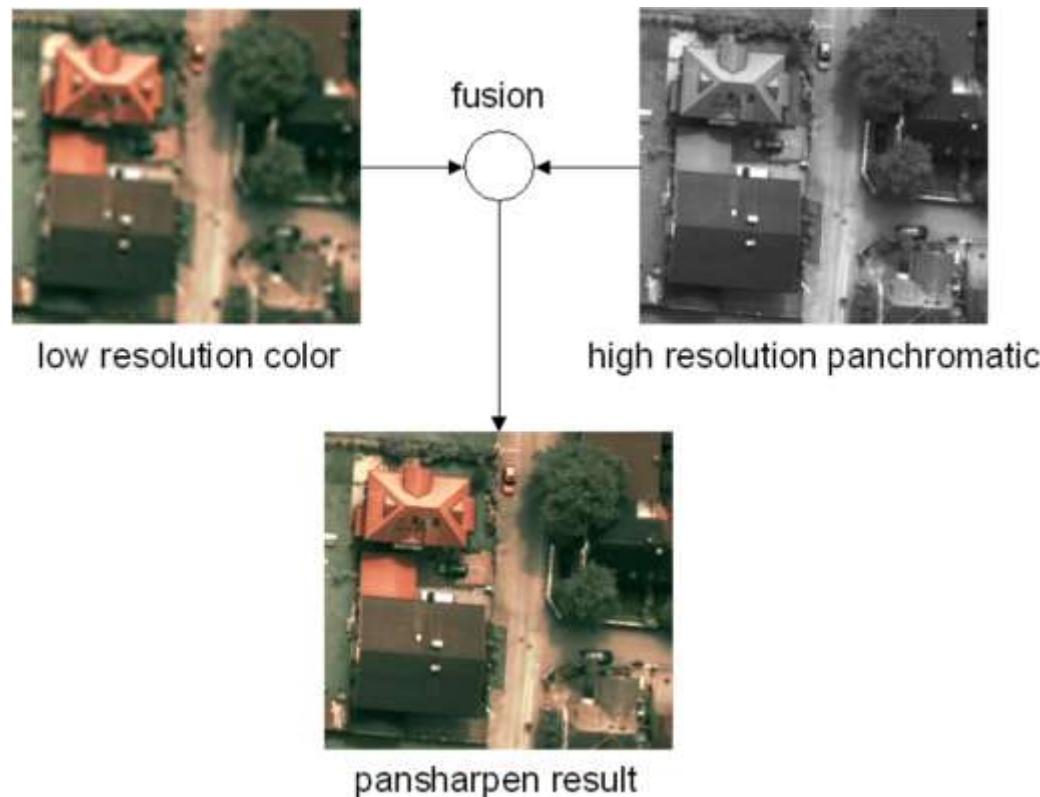
$$x \in \mathbb{R}^{m \times m \times b} \rightarrow y \in \mathbb{R}^{km \times km \times b}$$

Image compression



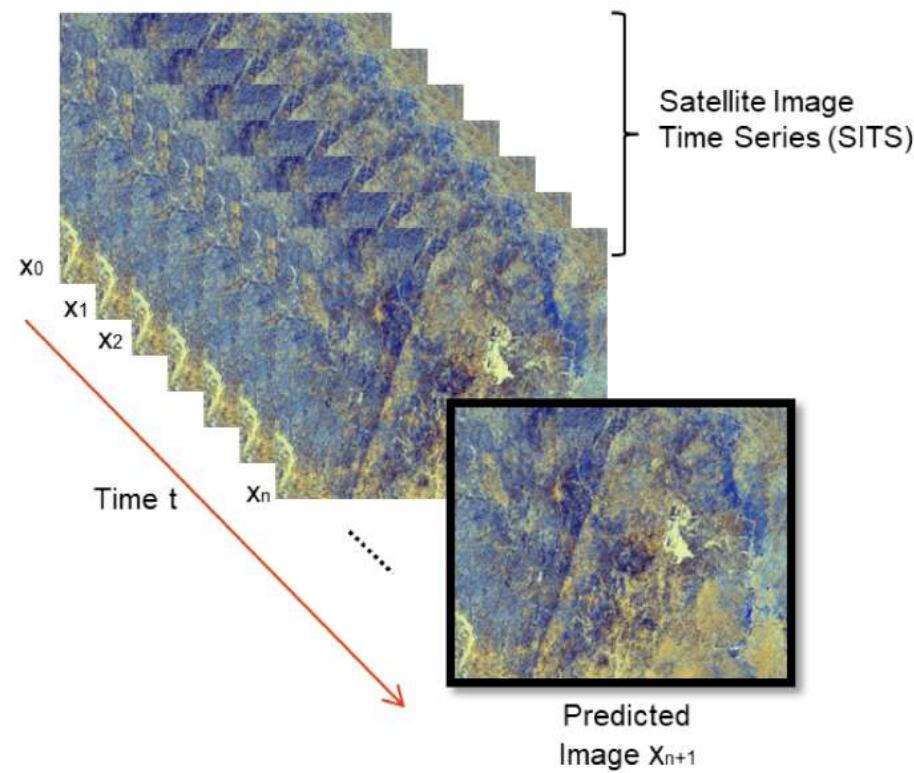
$$x \in \mathbb{R}^{m \times m \times b} \rightarrow y \in \{000111010101101010101 \dots\}$$

Fusion



$$\{x_1 \in \mathbb{R}^{m \times m \times b}, x_2 \in \mathbb{R}^{n \times n \times 1}\} \rightarrow y \in \mathbb{R}^{n \times n \times b}$$

Forecasting



$$x \in \mathbb{R}^{m \times m \times t} \rightarrow y \in \mathbb{R}^{m \times m}$$

Data sources



In-situ

+ High quality

+ High temporal resolution

- Localized



Remote Sensing

+ Global coverage

+ Moderate temporal
resolution

- Coarse-resolution

Numerical models

+ Capture physical laws

+ Flexible spatial/temporal
resolution

- Expensive



Copernicus data

The screenshot shows the Copernicus Open Access Hub interface. At the top, there are logos for the European Union, ESA, and Copernicus. The title "Copernicus Open Access Hub" is displayed. A search bar with placeholder text "Insert search criteria..." is present. A red circle highlights a message: "Your cart contains 49 products. Display 1 to 25 of 49 products." Below this, there are three product items listed:

- S3A SYNERGY** S3A_SY_2_SYN_20190209T214033_20190209T214333_201...
Download URL: [https://scihub.copernicus.eu/demohub/odata/v1/Products\('b9b4](https://scihub.copernicus.eu/demohub/odata/v1/Products('b9b4)
Mission: Sentinel-3 Instrument: SYNERGY Sensing Date: 2019-02-09T21:40:33.
- S3A SLSTR** S3A_SL_1_RBT_20190209T214833_20190209T214833_201...
Download URL: [https://scihub.copernicus.eu/demohub/odata/v1/Products\('61a5](https://scihub.copernicus.eu/demohub/odata/v1/Products('61a5)
Mission: Sentinel-3 Instrument: SLSTR Sensing Date: 2019-02-09T22:45:32.613

On the right side of the interface is a map of Europe and parts of Asia and Africa, showing various locations with green boxes indicating download points. Several blue ovals with text labels point to specific features:

- Remove single product from Cart**: Points to the trash can icon in the product card's control panel.
- Single product download**: Points to the download icon in the product card's control panel.
- Clear Cart**: Points to the "Clear Cart" button at the bottom of the cart area.
- Bulk Download**: Points to the red-bordered "Clear Cart" button, which also contains a trash can and a download icon.

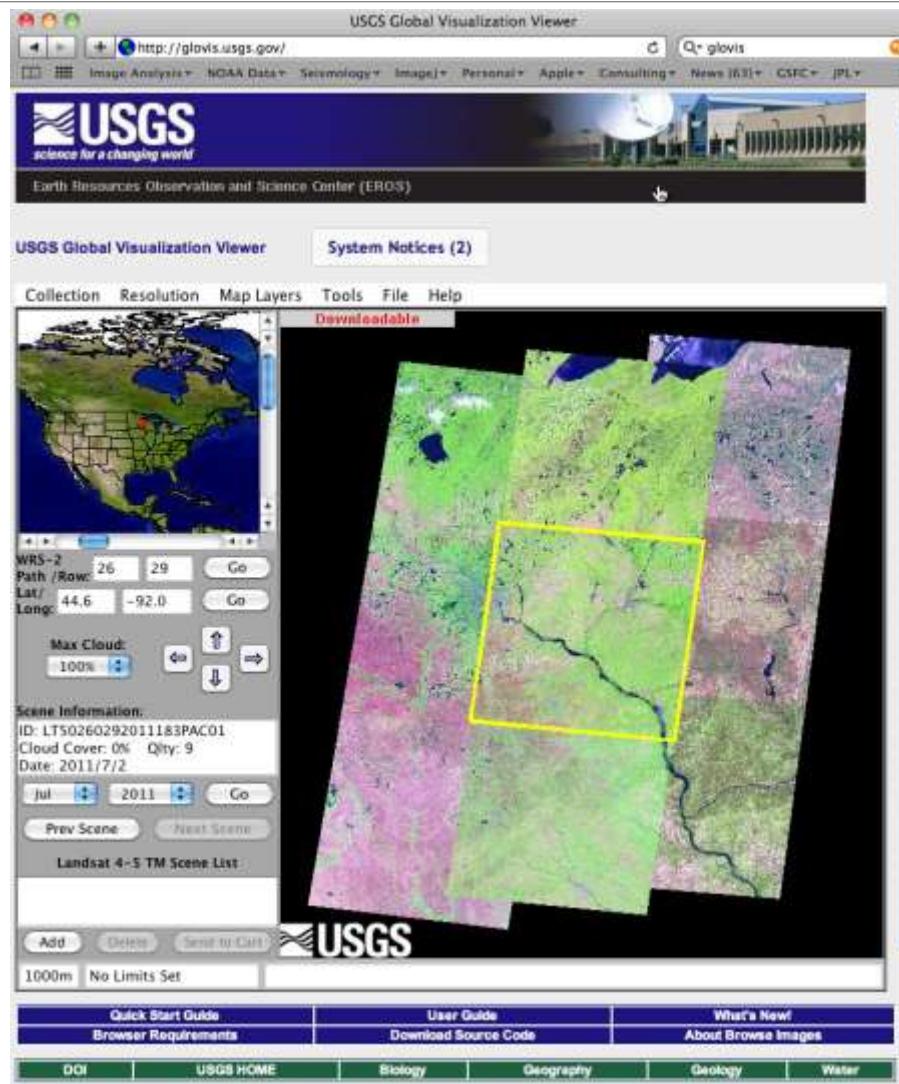
Alaska Satellite Facility

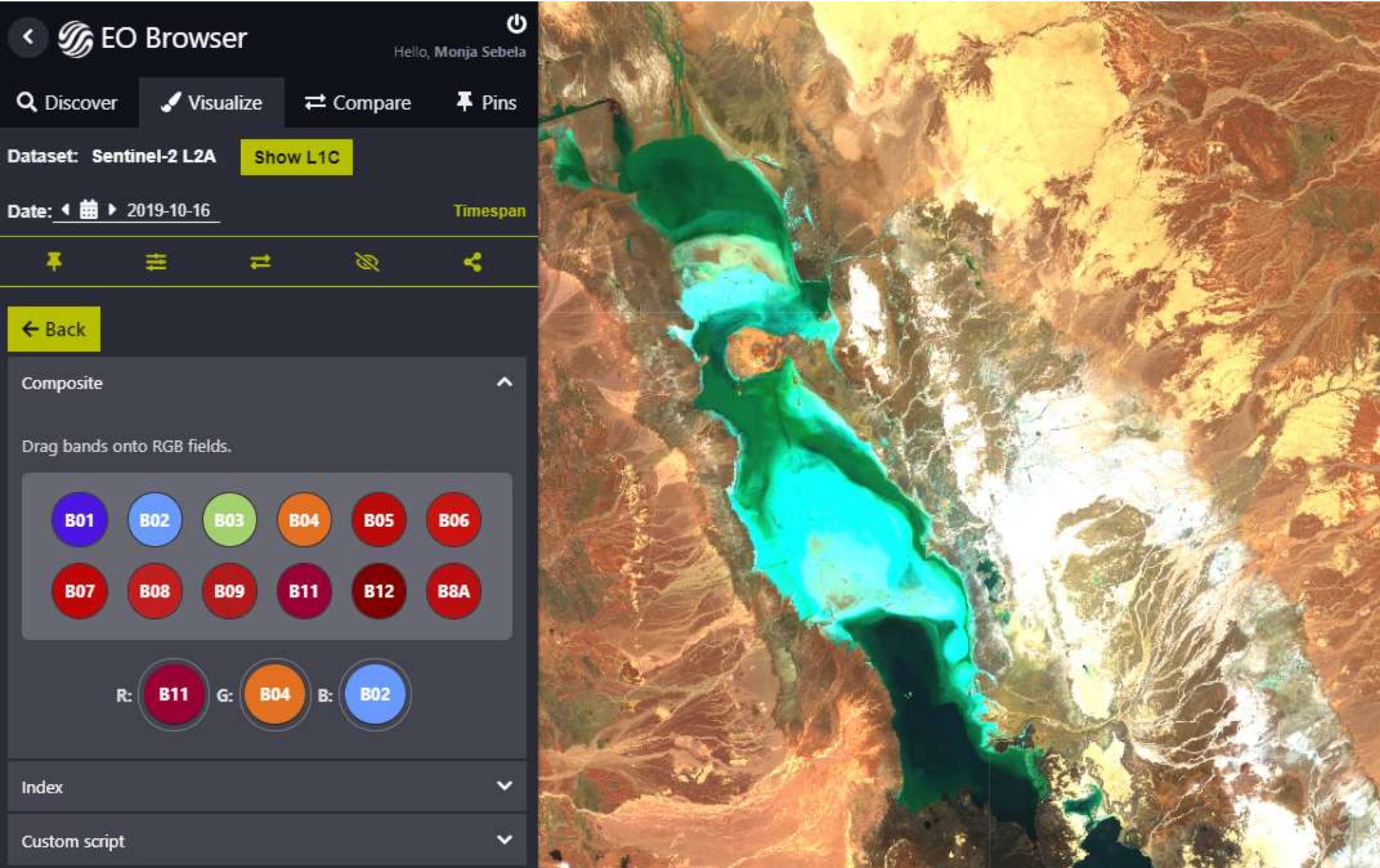
The screenshot shows the ASF Data Search interface. At the top, there's a search bar with "Geographic Search" selected, "Dataset" set to "Sentinel-1", and a search result of "250 of 3,829 Files". Below the search bar is a map of the Great Plains region, specifically Wyoming, Nebraska, and Iowa. A yellow rectangle highlights a specific area near Sioux Falls, and a red rectangle highlights another area near Lincoln. A blue rectangle highlights an area near Des Moines. A tooltip on the right side of the map states: "This website uses cookies to ensure you get the best experience on our website. [Learn More](#)". Below the map, there's a list of "42 Scenes (250 of 3,829 Files)" with the following details:

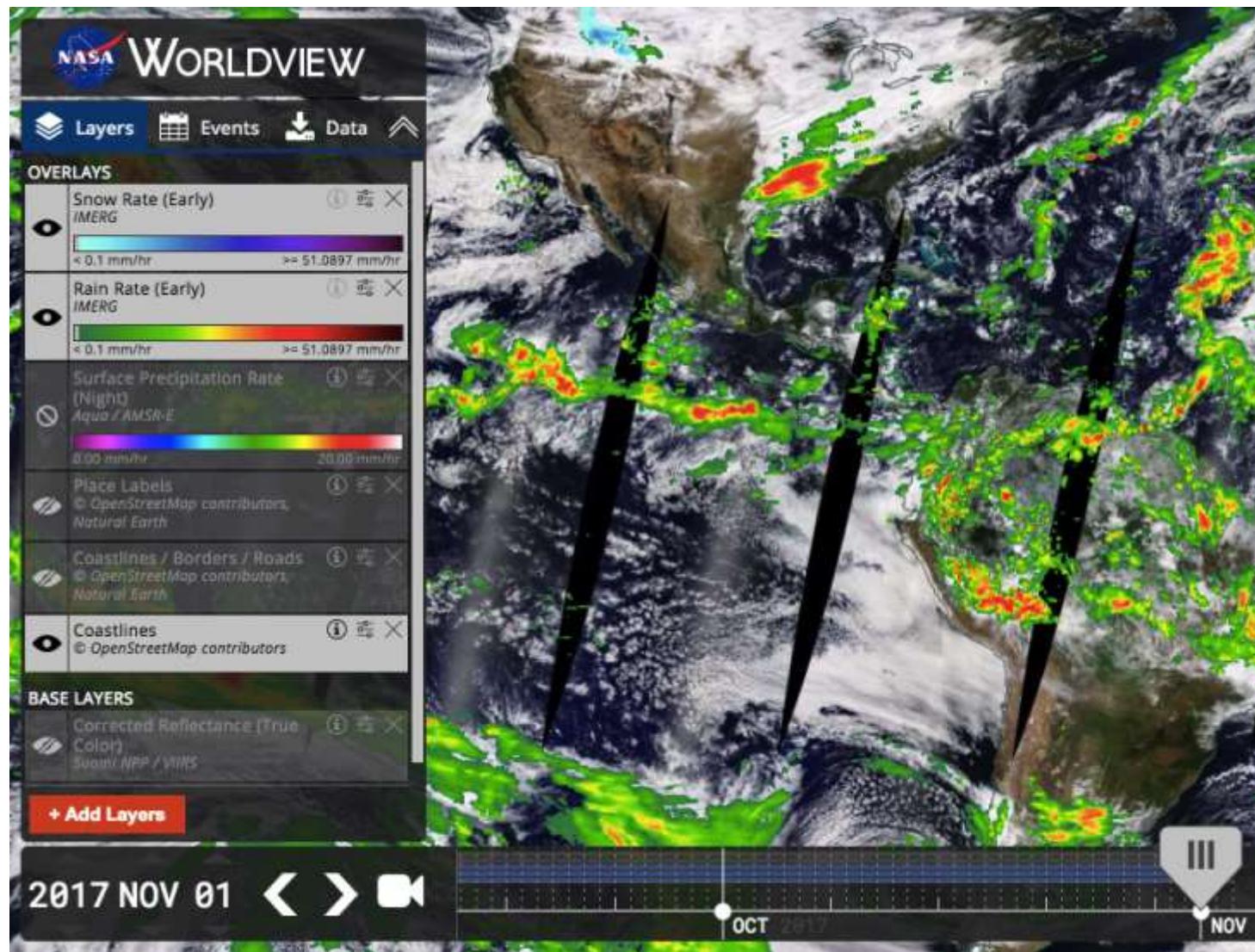
Scene ID	Date	Action Buttons	File Count
S1A_IW_GRDH_1SDV_20221220T002214_20221220T002239_046408_058F3A_9057	December 20 2022 00:22:14Z	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	0/6
S1A_IW_GRDH_1SDV_20221208T002215_20221208T002239_046408_058F3A_9057	December 08 2022 00:22:15Z	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	0/6
S1A_IW_GRDH_1SDV_20221203T001420_20221203T001440_046408_058F3A_9057	December 03 2022 00:14:20Z	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	0/6
S1A_IW_GRDH_1SDV_20221203T001355_20221203T001415_046408_058F3A_9057	December 03 2022 00:13:55Z	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	0/6

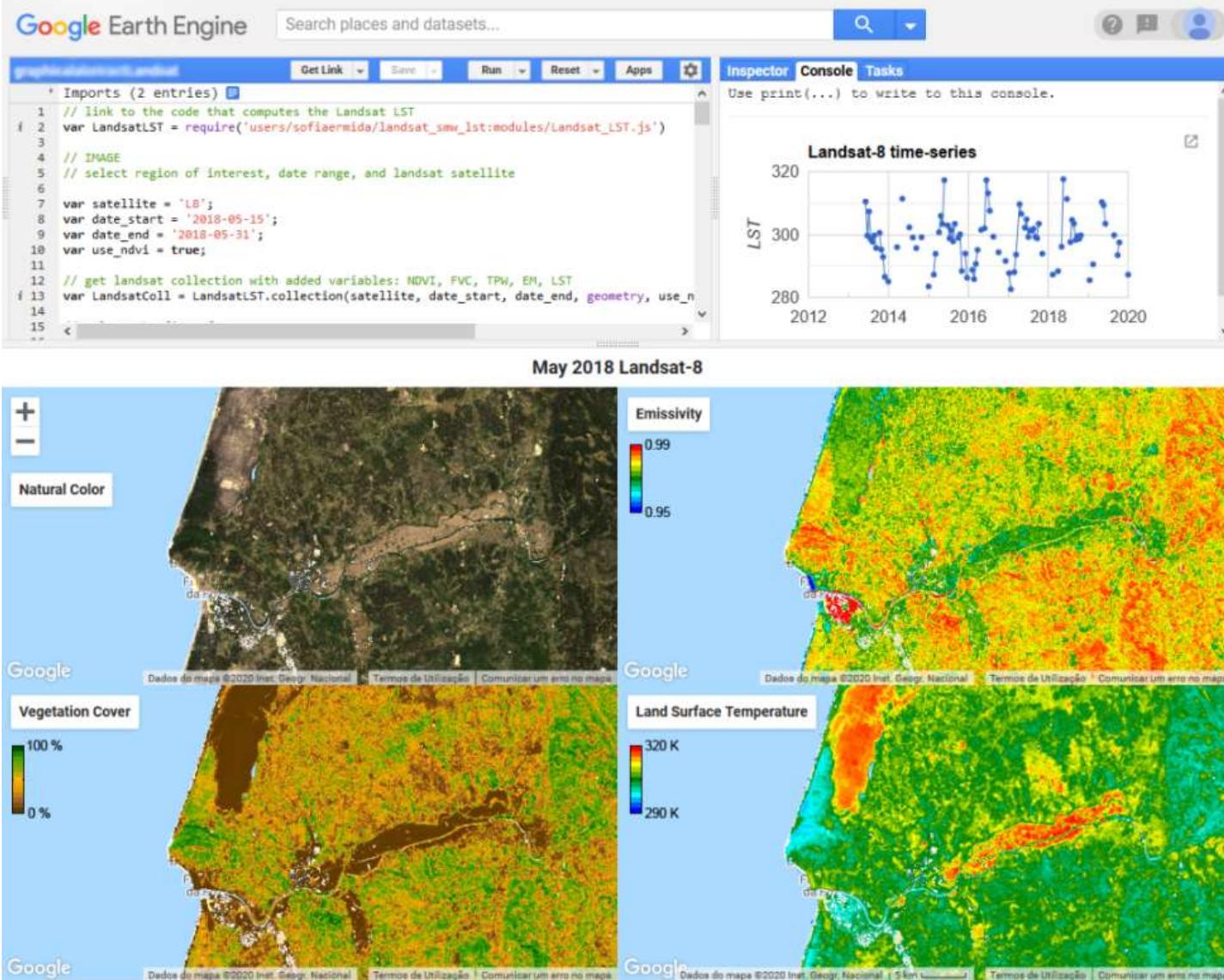
On the right side of the interface, there's a "Scene Detail / 6 Files" panel for the first scene listed. It shows the file name "S1A_IW_GRDH_1SDV_20221220T002214_20221220T002239_046408_058F3A_9057", the date "December 20 2022 00:22:14Z", and the mode "Sentinel-1 • C-Band". It also includes a note: "Accessing this data requires you to log in." Below this are the "Flight Direction", "Polarization", "Absolute Orbit", and "Data courtesy of ESA" information. At the bottom of the panel are buttons for "SEARCH:", "Baseline", "SBAS", and "More Like This".

USGS









Data assimilation

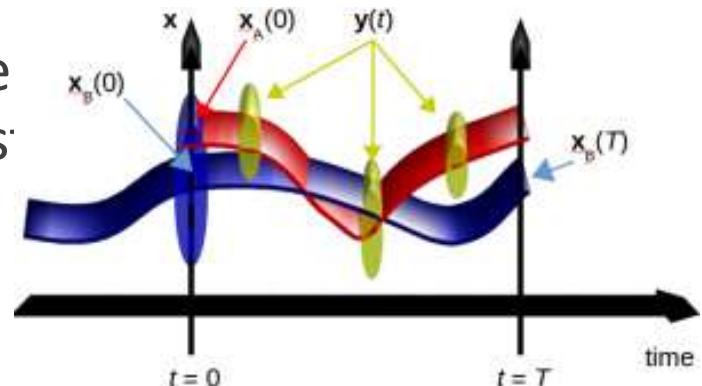
Challenges

- Free running model will drift from reality (modelling approximations, unknown processes acting, and uncertain initial conditions).

Approach

Data Assimilation: fusion of different types of information to estimate possible states of a system.

- Sparse observations
- Measurements error
- Indirect sensing

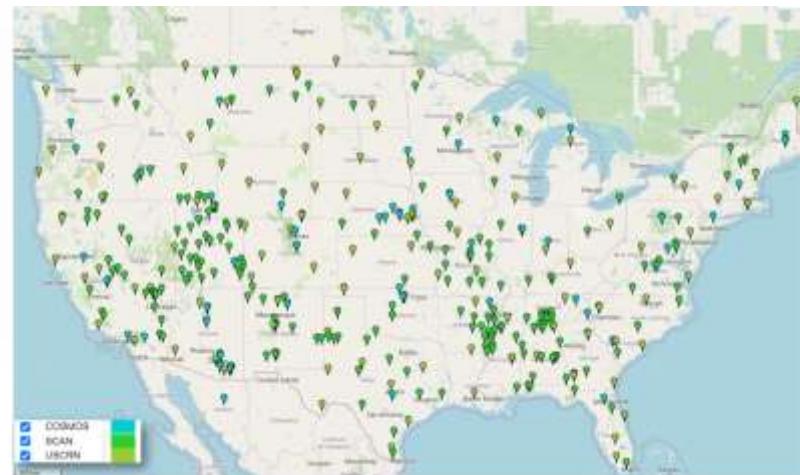


<https://research.reading.ac.uk/met-darc/aboutus/what-is-data-assimilation/>

In-situ sensor network

The International Soil Moisture Network (ISMN)

Variable name	Abbreviation	Units	Measurement depth* (m)	Variable with depth?	No. of time series (stations)
Soil moisture	sm	$\text{m}^3 \text{ m}^{-3}$	0.00–2.10	Y	10 610 (2822)
Soil suction	su	kPa	0.04–0.75	Y	73 (18)
Soil temperature	ts	°C	0.00–2.03	Y	8 113 (1629)
Air temperature	ta	°C	2.00–12.00	Y	1 292 (1234)
Surface temperature	tsf	°C	0.00–0.00	N	126 (126)
Precipitation	p	mm	0.00–2.00	N	759 (700)
Snow depth	sd	mm	0.00	N	562 (555)
Snow water equivalent	sweq	mm	0.00	N	507 (427)



- Harmonization
- Quality Control
- Sparse locations

Supervised Learning

Data
Labels



Model
Prediction



← Spiral



← Elliptical

Exploiting prior knowledge

- Expert users
- Crowdsourcing
- Other instruments
- Time



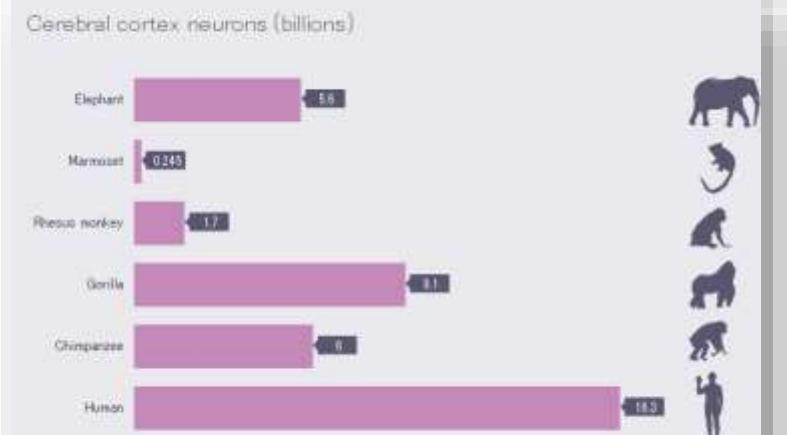
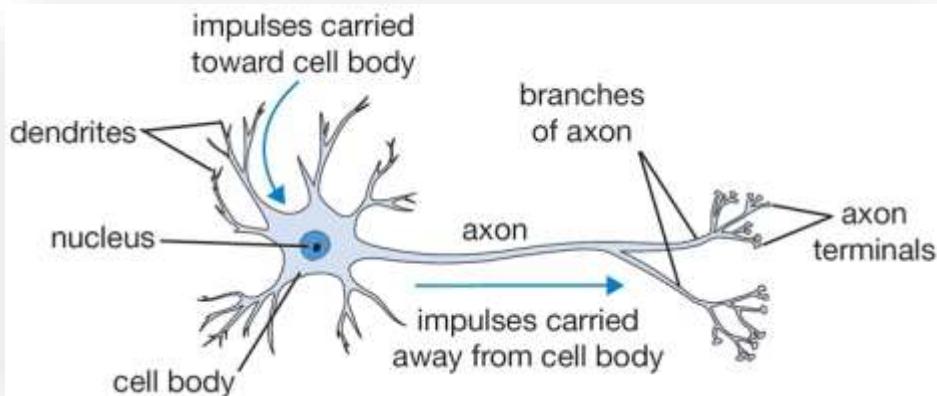
→ ?

Artificial Neural Networks (ANN)

aka Deep Learning

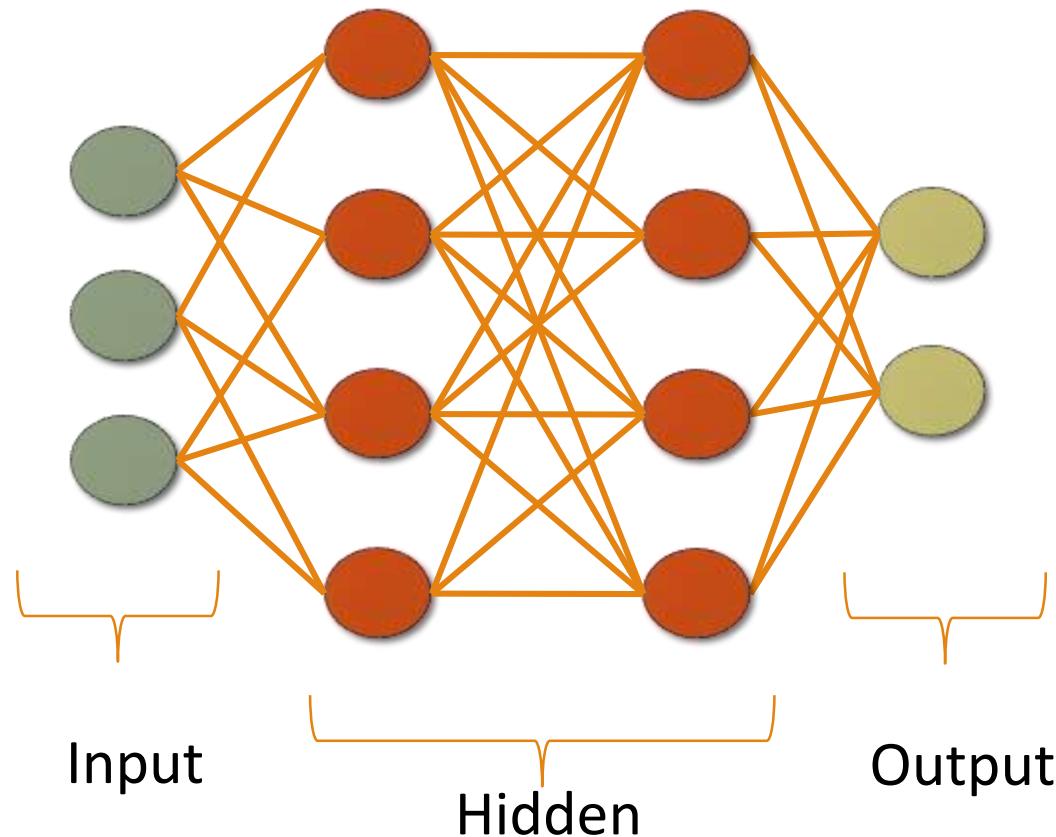
Inspiration from brain

- 86Billion neurons
- 10^{14} - 10^{15} synapses



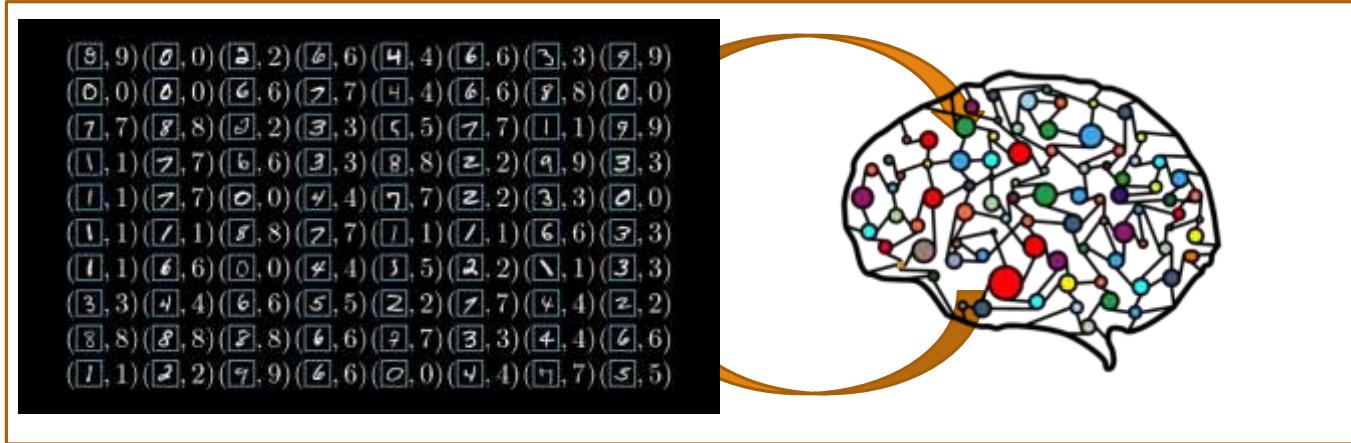
Key components of ANN

- Architecture (input/hidden/output layers)

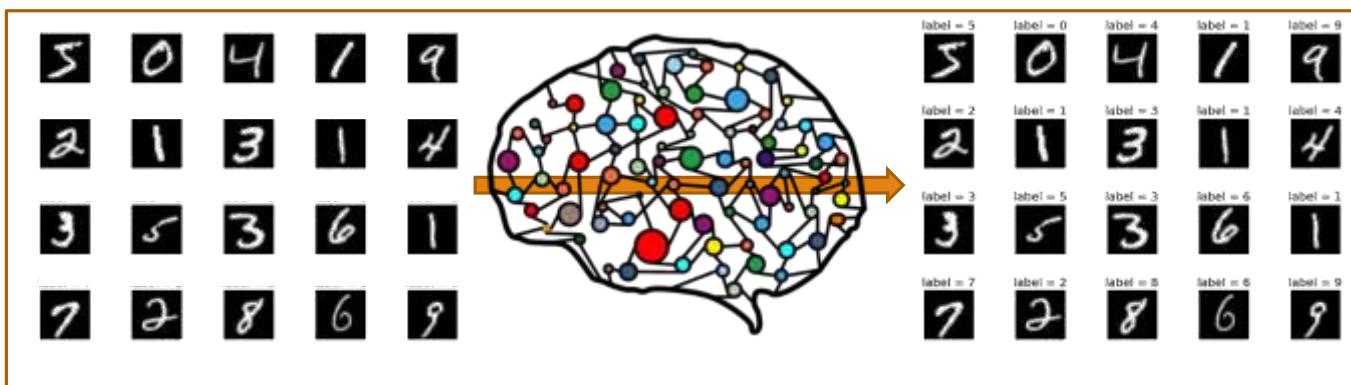


Key processing in ANN

Training

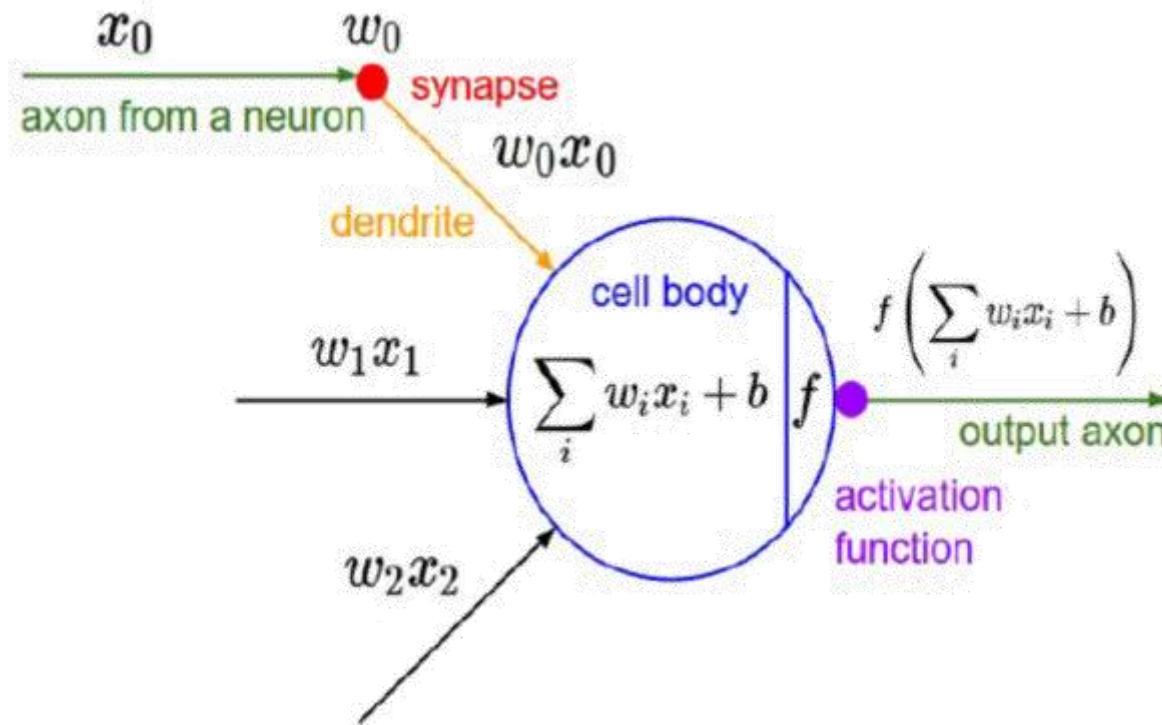


Inference (Test)



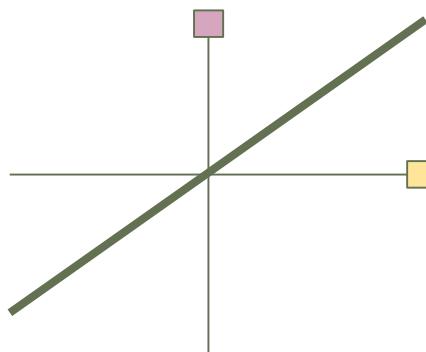
Key components of ANN

- Architecture (input/hidden/output layers)
- Weights

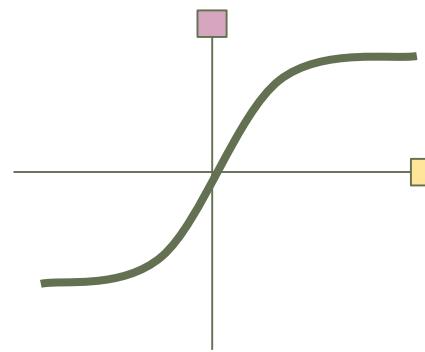


Key components of ANN

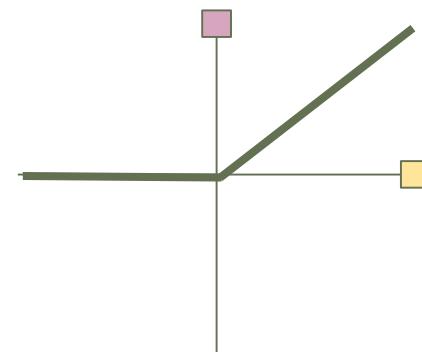
- Architecture (input/hidden/output layers)
- Weights
- Activations



LINEAR



**LOGISTIC /
SIGMOIDAL / TANH**



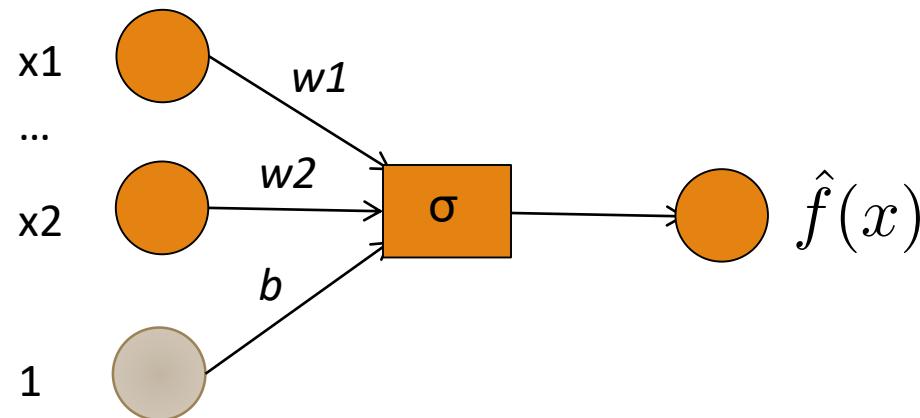
**RECTIFIED
LINEAR (ReLU)**

Perceptron: an early attempt

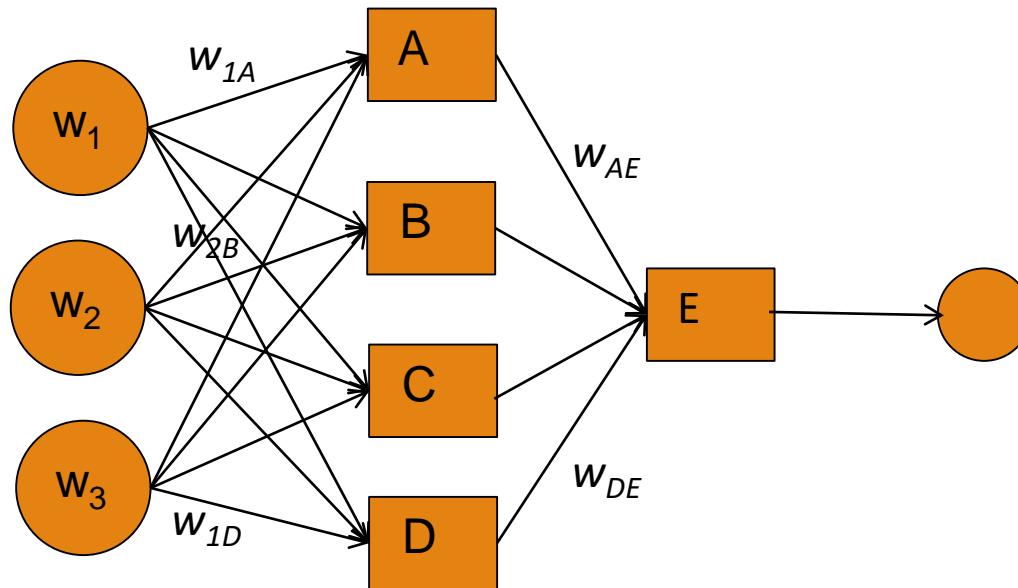
Activation function

$$\hat{f}(x) = \sigma(w \cdot x + b) \quad \sigma(y) = \begin{cases} 1, & y > 0 \\ 0, & o/w \end{cases}$$

Need to tune w and b



Multilayer perceptron



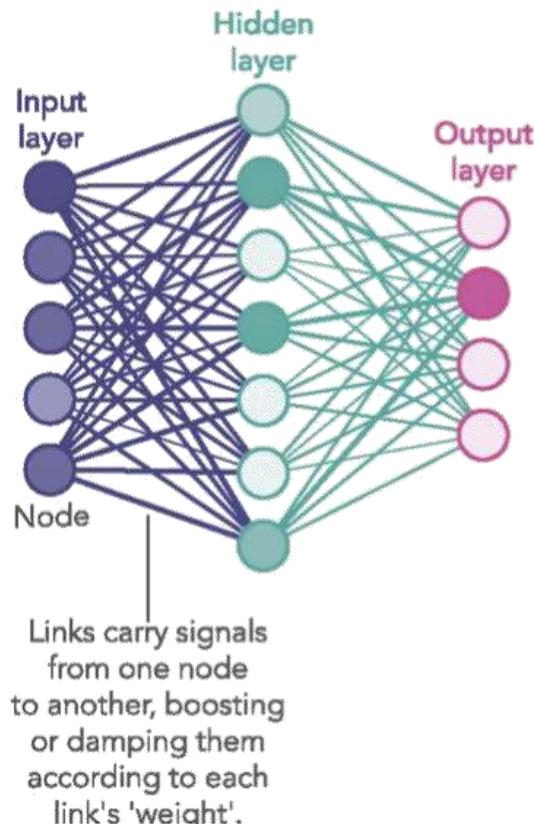
We just added a
neuron layer!

We just introduced
non-linearity!

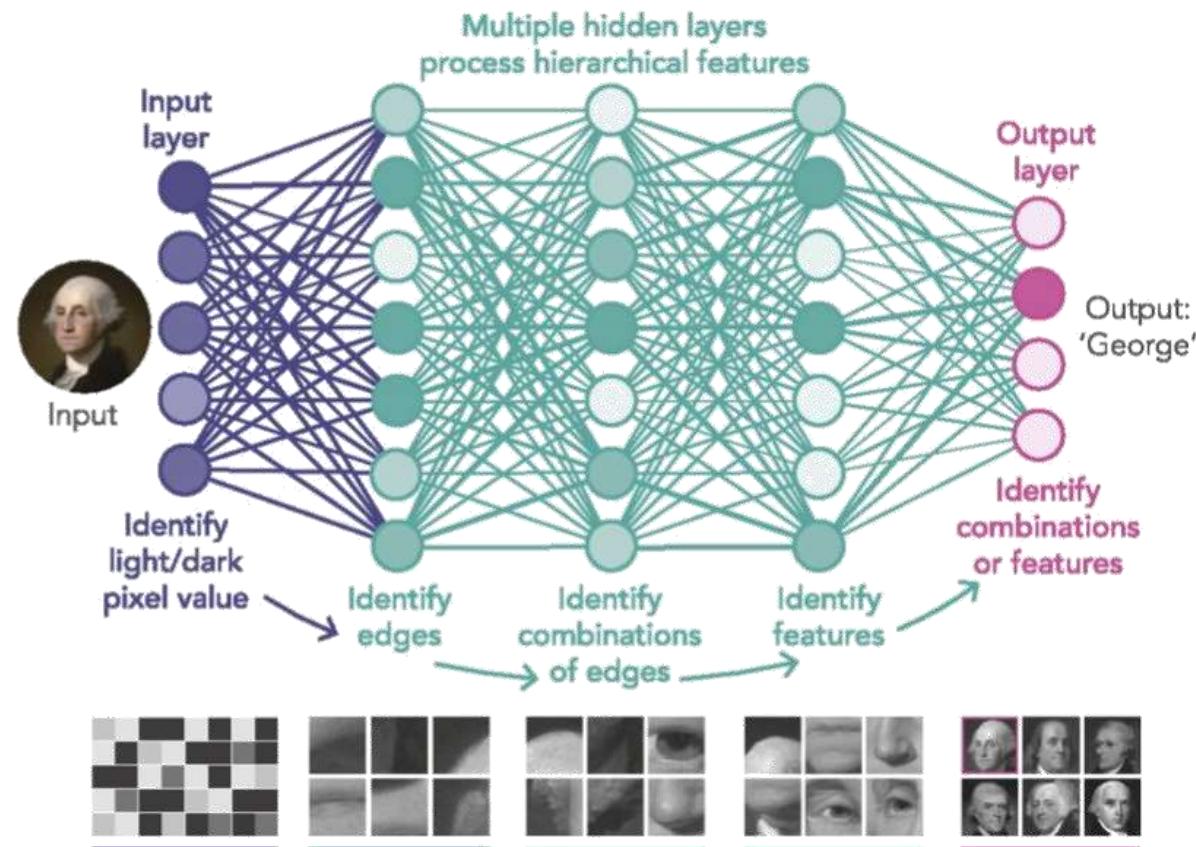
A neuron is of the form
 $\sigma(w \cdot x + b)$ where σ is
an *activation function*

Deep Neural Networks

1980S-ERA NEURAL NETWORK



DEEP LEARNING NEURAL NETWORK



Neural Networks

©2016 Fjodor van Veen - asimovinstitute.org

○ Backfed Input Cell

○ Input Cell

△ Noisy Input Cell

● Hidden Cell

● Probabilistic Hidden Cell

△ Spiking Hidden Cell

● Output Cell

● Match Input Output Cell

● Recurrent Cell

● Memory Cell

△ Different Memory Cell

● Kernel

● Convolution or Pool

Perceptron (P)



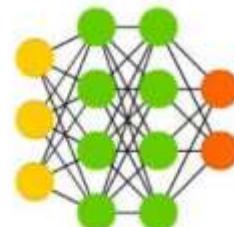
Feed Forward (FF)



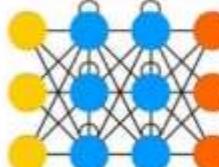
Radial Basis Network (RBF)



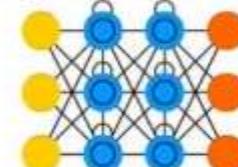
Deep Feed Forward (DFF)



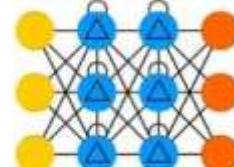
Recurrent Neural Network (RNN)



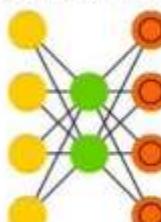
Long / Short Term Memory (LSTM)



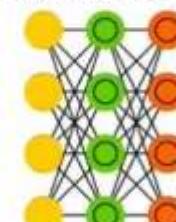
Gated Recurrent Unit (GRU)



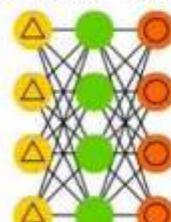
Auto Encoder (AE)



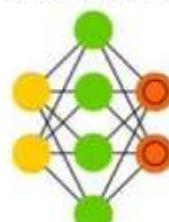
Variational AE (VAE)



Denoising AE (DAE)



Sparse AE (SAE)



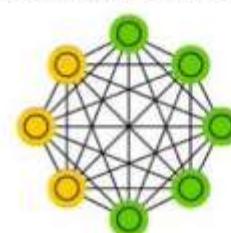
Markov Chain (MC)



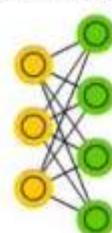
Hopfield Network (HN)



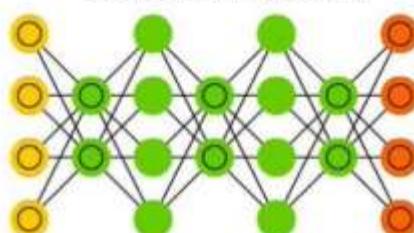
Boltzmann Machine (BM)



Restricted BM (RBM)



Deep Belief Network (DBN)



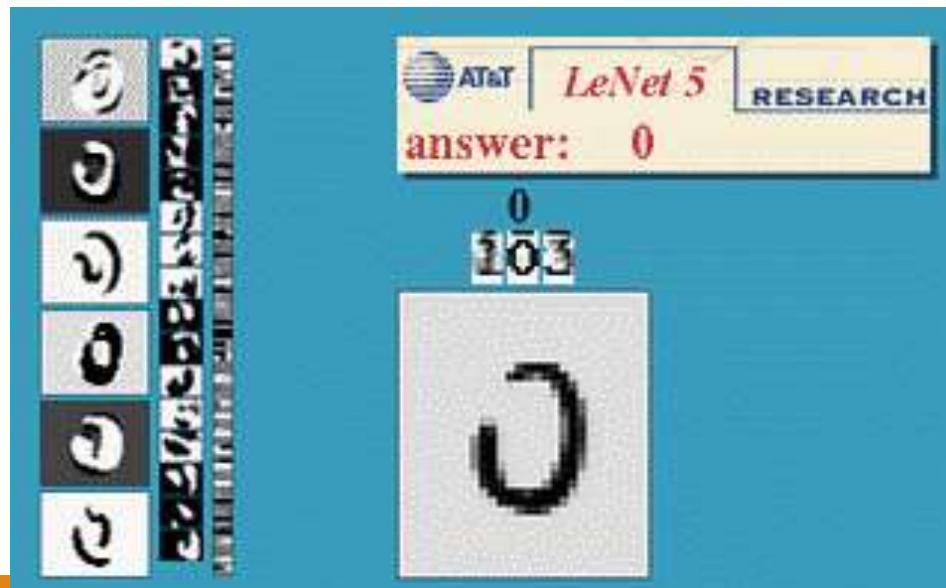
[Sasen Cain \(@spectralradius\)](#)

Training & Testing

Training: determine weights

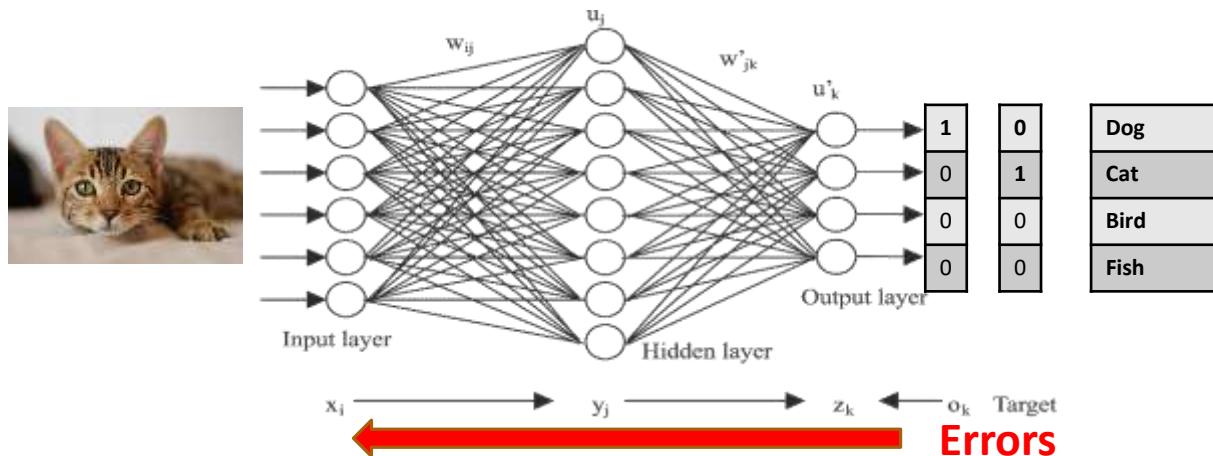
- Supervised: labeled training examples
- Unsupervised: no labels available
- Reinforcement: examples associated with rewards

Testing (Inference): apply weights to new examples



Training DNN

1. Get batch of data
2. Forward through the network -> compute loss
3. Backpropagate error
4. Update weights based on gradient



Training DNN

Input: x_i Target: y_i DNN: $f(\cdot; \theta)$

Forward pass $\hat{y}_i = f(x_i; \theta)$

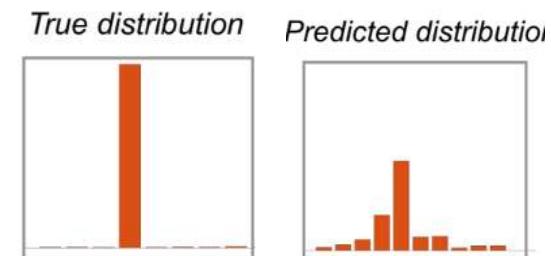
Compute loss

- Regression: Root Mean Squared Error (RMSE)

$$\mathcal{L}(\hat{y}, y) = \frac{1}{2N} \sum_i (y_i - f(x_i; \theta))^2$$

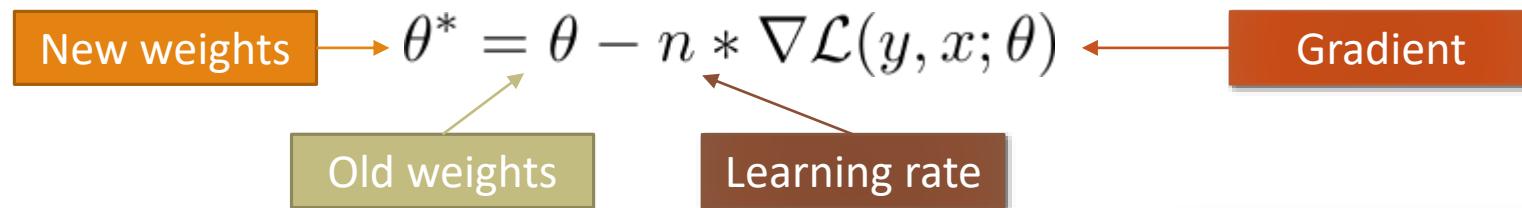
- Classification: Categorical Cross Entropy (CCE)

$$\mathcal{L}(\hat{y}, y) = -\frac{1}{N} \sum_i \sum_j y_{ij} \log \hat{y}_{ij}$$



Gradient Descent

► Minimize function J w.r.t. parameters θ

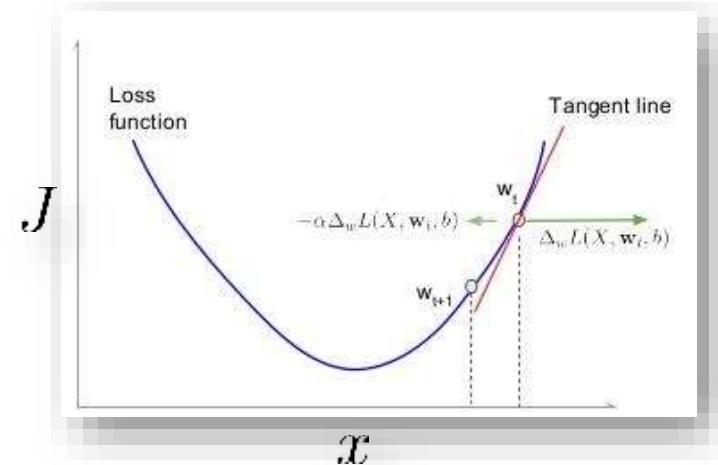


■ Gradient

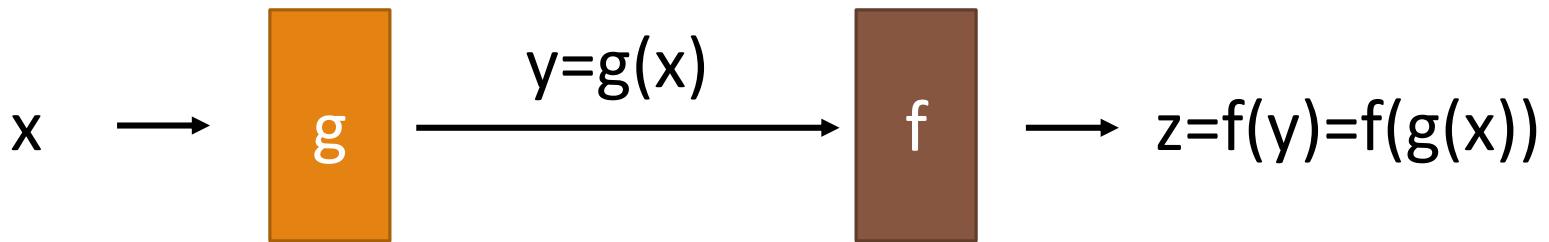
$$\nabla \mathcal{L}(x) = \left(\frac{\partial \mathcal{L}(x)}{\partial x_1}, \frac{\partial \mathcal{L}(x)}{\partial x_2}, \dots, \frac{\partial \mathcal{L}(x)}{\partial x_n} \right)$$

■ Stochastic Gradient Descend

$$\theta^* := \theta - n \sum_{i=1}^N \nabla \mathcal{L}(y_i, x_i; \theta)$$



Backpropagation



Chain rule:

- Single variable

$$\frac{dz}{dx} = \frac{dz}{dy} \frac{dy}{dx}.$$

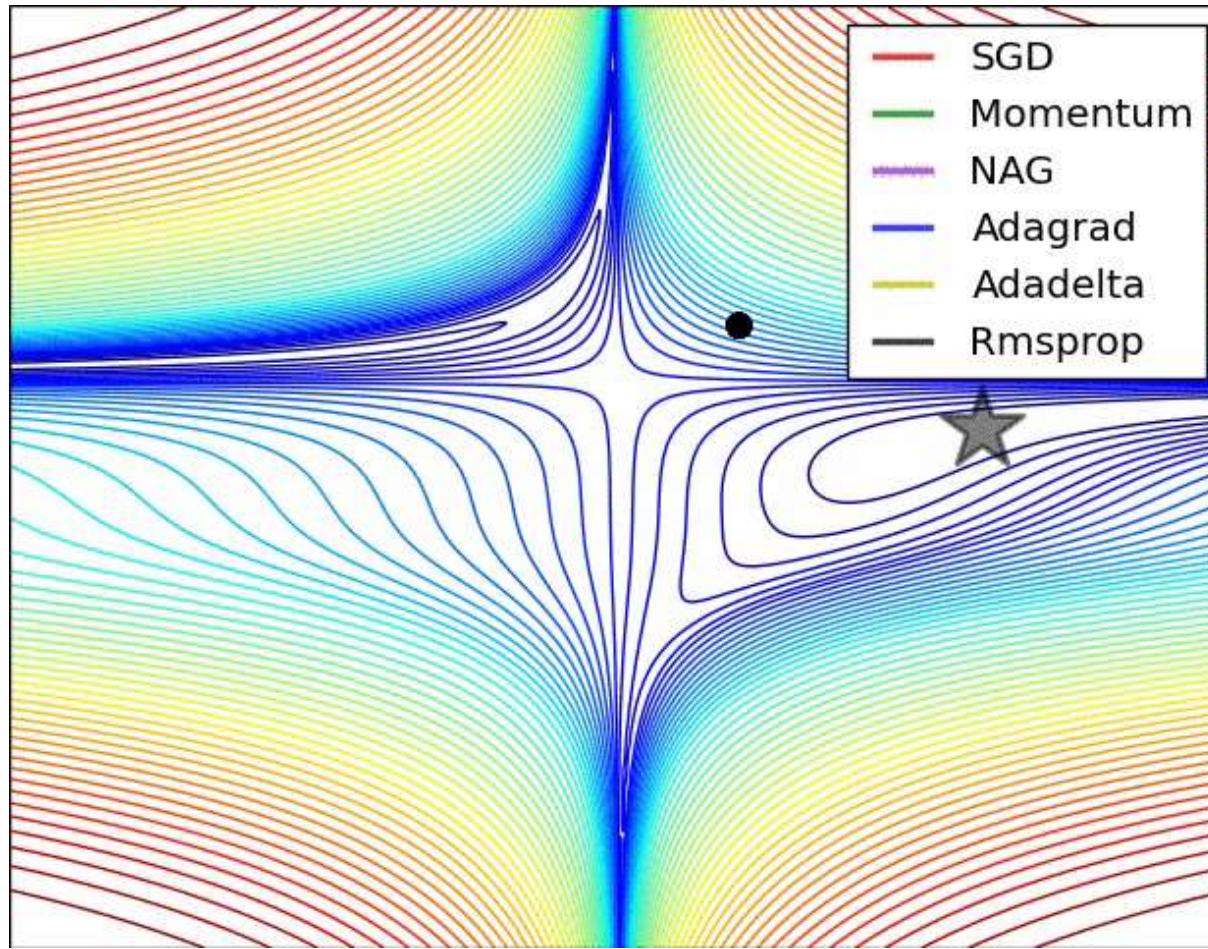
- Multiple variables

$$\frac{\partial z}{\partial x_i} = \sum_j \frac{\partial z}{\partial y_j} \frac{\partial y_j}{\partial x_i}.$$

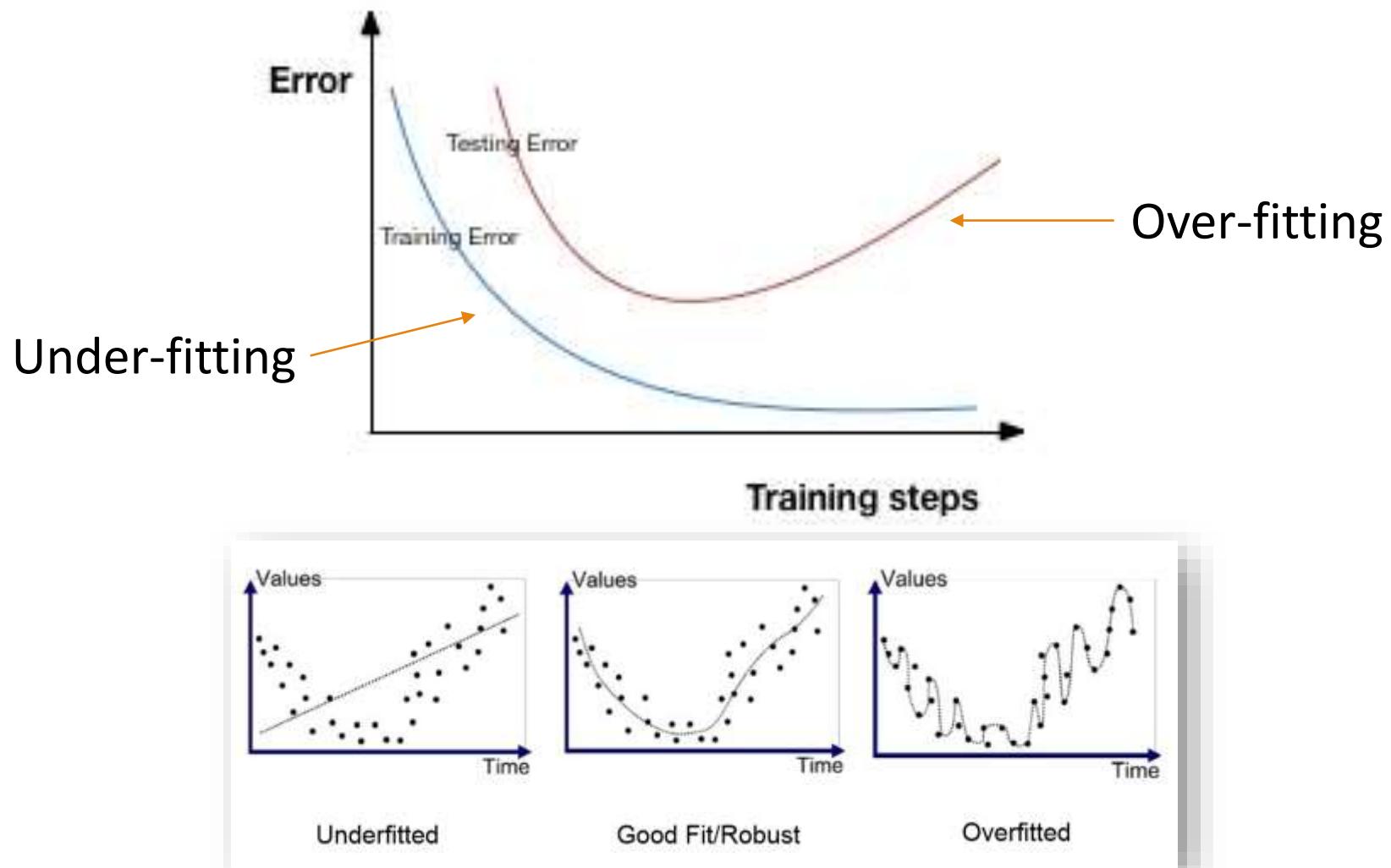
Optimization algorithms

Optimization algorithm	Core idea	Pros	Cons
SGD [140]	Computes the gradient of mini-batches iteratively and updates the parameters	<ul style="list-style-type: none">• Easy to implement	<ul style="list-style-type: none">• Setting a global learning rate required• Algorithm may get stuck on saddle points or local minima• Slow in terms of convergence• Unstable
Nesterov's momentum [125]	Introduces momentum to maintain the last gradient direction for the next update	<ul style="list-style-type: none">• Stable• Faster learning• Can escape local minima	<ul style="list-style-type: none">• Setting a learning rate needed
Adagrad [126]	Applies different learning rates to different parameters	<ul style="list-style-type: none">• Learning rate tailored to each parameter• Handle sparse gradients well	<ul style="list-style-type: none">• Still requires setting a global learning rate• Gradients sensitive to the regularizer• Learning rate becomes very slow in the late stages
Adadelta [141]	Improves Adagrad, by applying a self-adaptive learning rate	<ul style="list-style-type: none">• Does not rely on a global learning rate• Faster speed of convergence• Fewer hyper-parameters to adjust	<ul style="list-style-type: none">• May get stuck in a local minima at late training
RMSprop [140]	Employs root mean square as a constraint of the learning rate	<ul style="list-style-type: none">• Learning rate tailored to each parameter• Learning rate do not decrease dramatically at late training• Works well in RNN training	<ul style="list-style-type: none">• Still requires a global learning rate• Not good at handling sparse gradients
Adam [127]	Employs a momentum mechanism to store an exponentially decaying average of past gradients	<ul style="list-style-type: none">• Learning rate stailored to each parameter• Good at handling sparse gradients and non-stationary problems• Memory-efficient• Fast convergence	<ul style="list-style-type: none">• It may turn unstable during training

Visualization

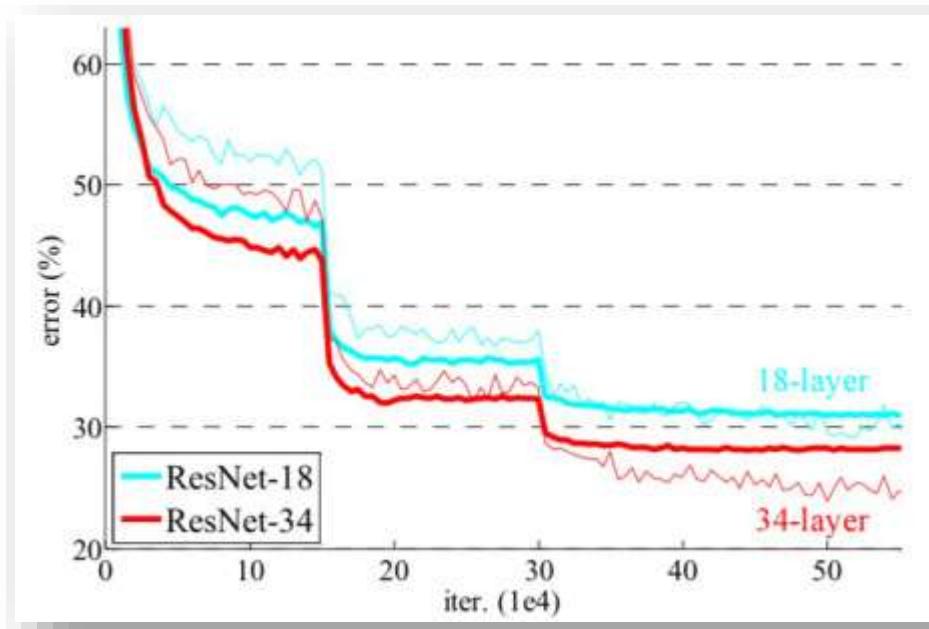
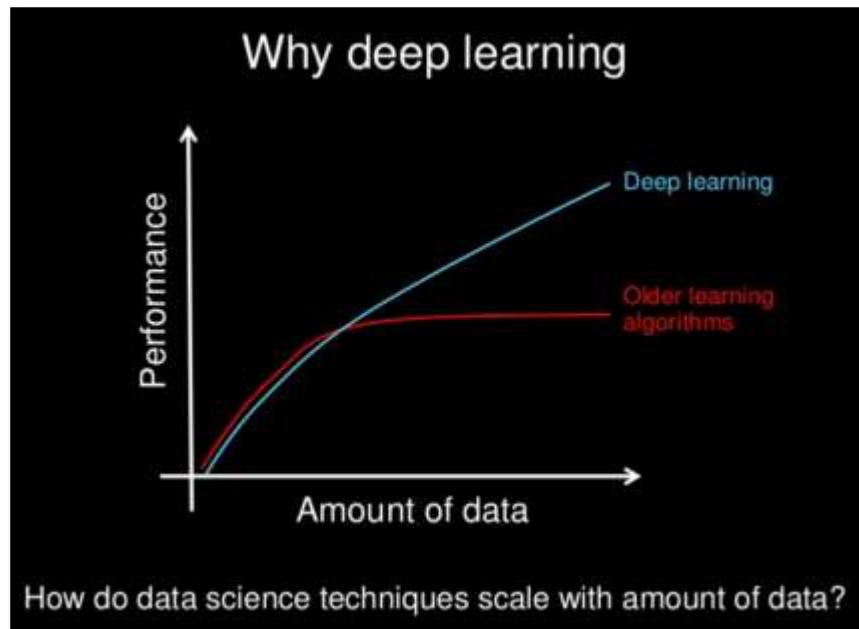


Training Characteristics



Why Today?

Lots of Data

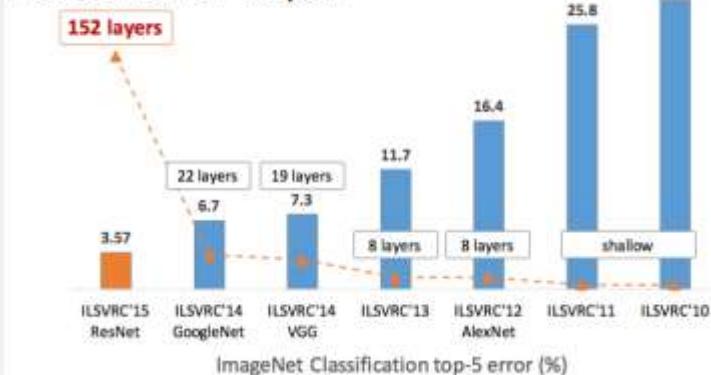


Why Today?

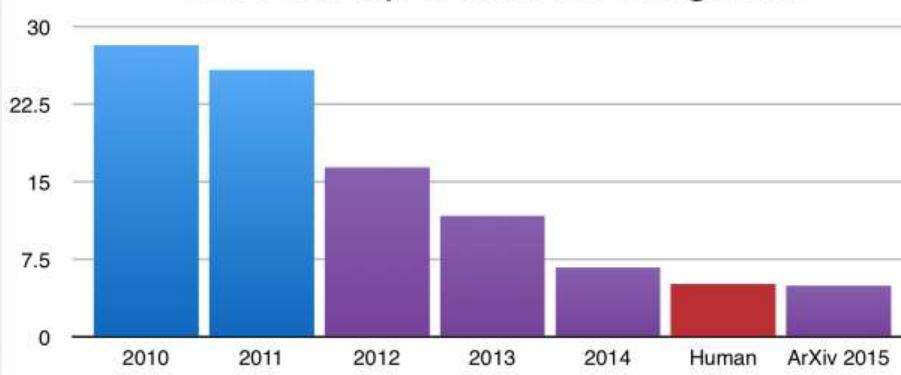
Lots of Data

Deeper Learning

Revolution of Depth



ILSVRC top-5 error on ImageNet

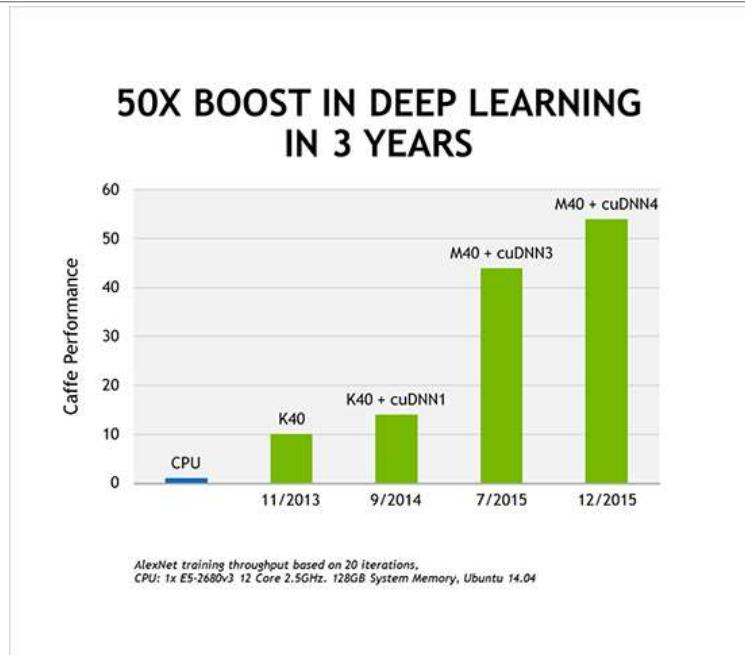


Why Today?

Lots of Data

Deep Learning

More Power



<https://blogs.nvidia.com/blog/2016/01/12/accelerating-ai-artificial-intelligence-gpus/>

<https://www.slothparadise.com/what-is-cloud-computing/>