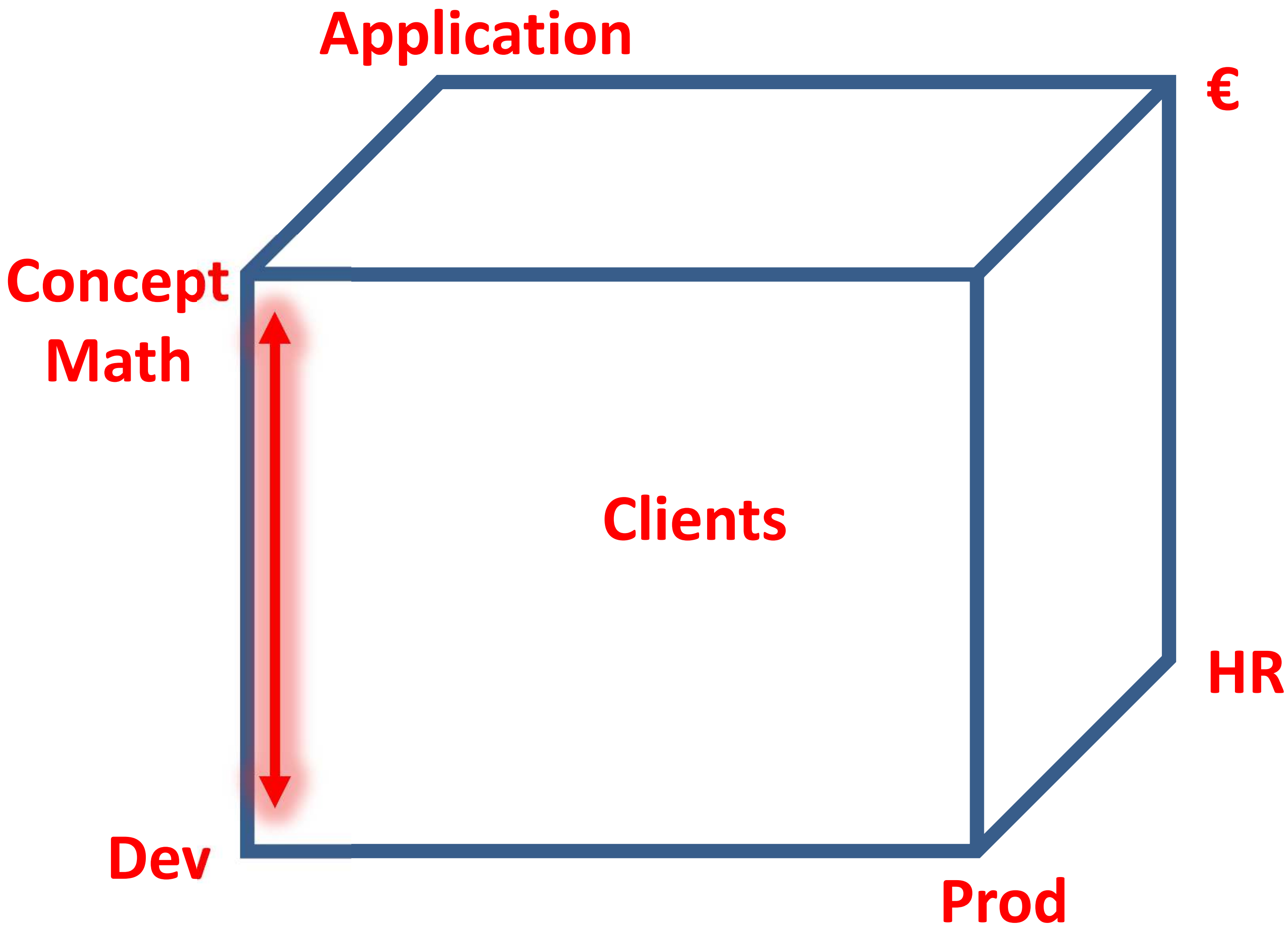
Optimisation & Machine Learning

Pr. Bijan Mohammadi
Université de Montpellier


Application
€
Concept
Math
Clients
HR
Dev
Prod

Optimisation & Machine Learning

1. Sans contrainte

-Gradient (SD, N, QN, SGD, ADAM,…)
-Sans gradient : région de confiance
-Sans gradient : stochastique, génétique, essaim

2. Avec contraintes

- Egalité / Inégalité
Multiplicateurs de Lagrange, Lagrangien, KKT, Uzawa

3. Multicritère, Pareto

4. Optimisation robuste, sur intervalle, multipoint

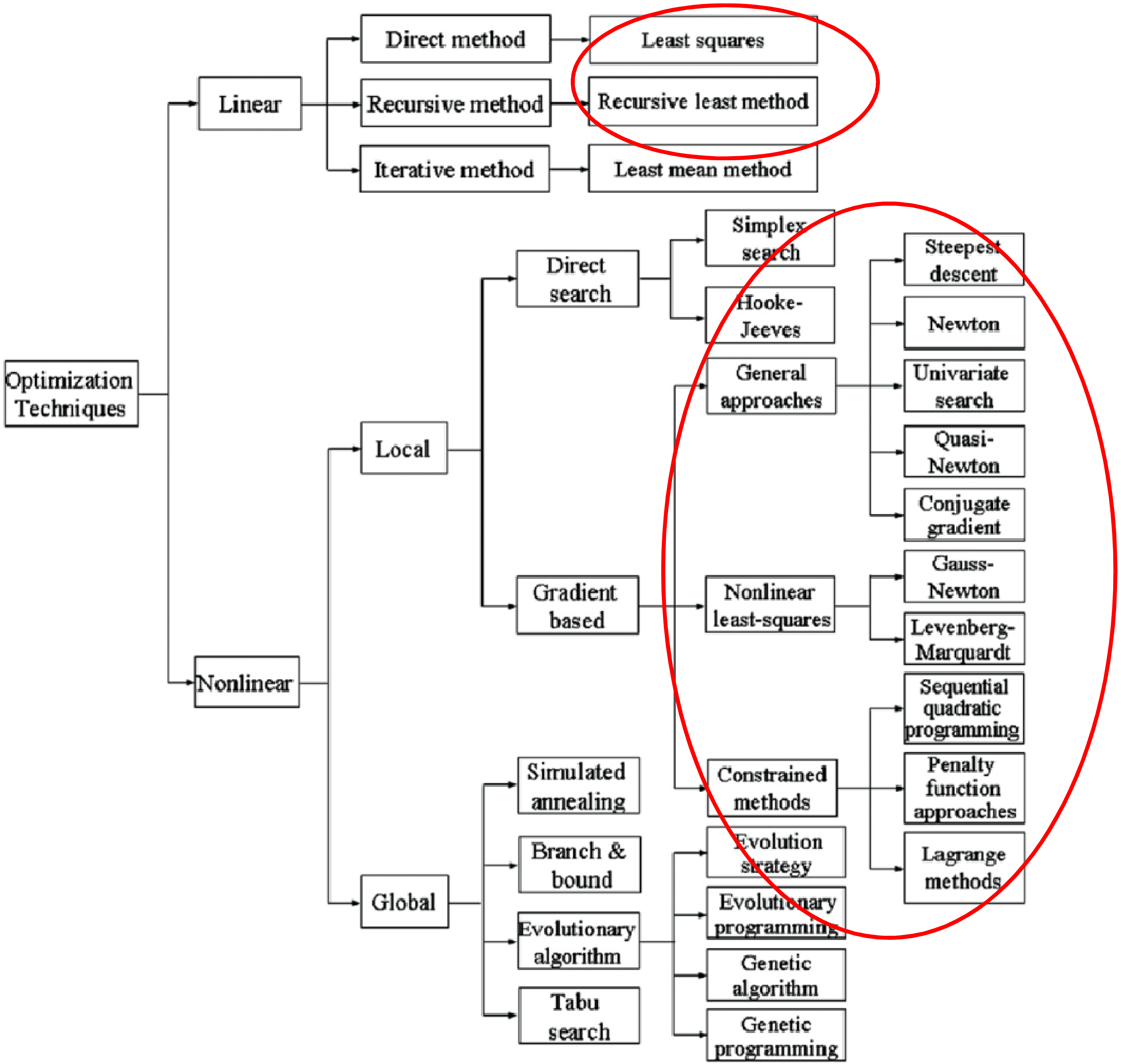
5. Optimisation globale

-Systèmes dynamiques d'ordre
-Stochastique (Génétique, recuit simulé, Tabu, Essaim, …)

5. Apprentissage supervisé (classification / régression)

-Décomposition en Valeurs Singulières (Singular Value Decomposition)
-Modèles linéaires (Linear Models)
-KNN (Nearest Neighbors)
-Séparateurs à Vaste Marge (Support Vector Machines)
-Forêts aléatoires (Random Forests)
-réseaux de neurones (Neural Networks)

6 Apprentissage non-supervisé (clustering k-means)



Optimization Techniques
Linear
Nonlinear
Direct method
Recursive method
Iterative method
Least squares
Recursive least method
Least mean method
Local
Global
Direct search
Gradient based
Simplex search
Hooke-Jeeves
General approaches
Nonlinear least-squares
Constrained methods
Steepest descent
Newton
Univariate search
Quasi-Newton
Conjugate gradient
Gauss-Newton
Levenberg-Marquardt
Sequential quadratic programming
Penalty function approaches
Lagrange methods
Simulated annealing
Branch & bound
Evolutionary algorithm
Tabu search
Evolution strategy
Evolutionary programming
Genetic algorithm
Genetic programming

Supervised Learning today in short

-Black-boxes & Requires know-how (esp. Deep NN)

-Requires lots of quality data

→Problem when data is rare, confidential & expensive to get

→Needs human intervention (*mechanical turks / crowdsourcing*)

Impossible for industrial regression pbs

Human Intelligence Tasks (HITs)

~300K click workers in France

Myth of Sisyphus

-Does not perform well on industrial configurations (success in B2C)

-Needs quality data (but bias is the rule in industry : machine wear, …)

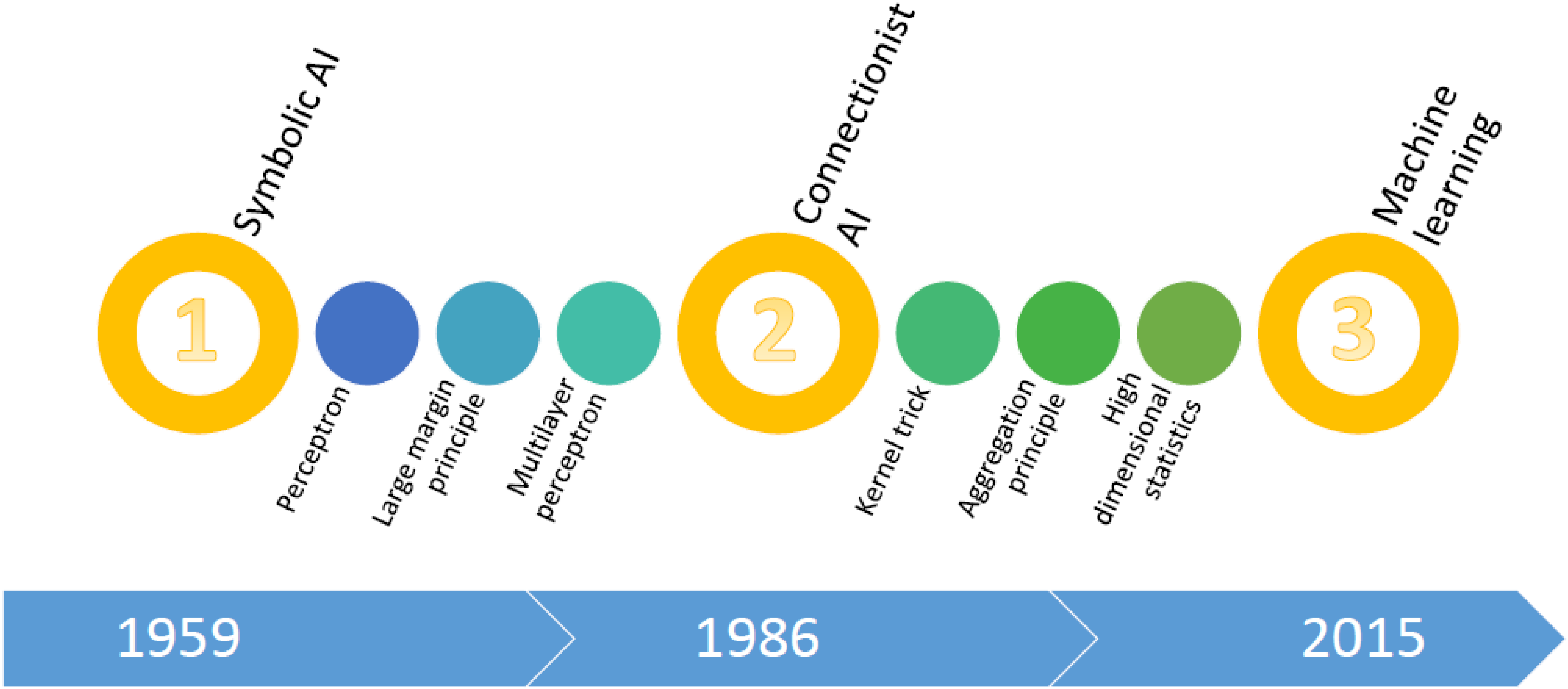
-Cannot be used in certification because stochastic/ensemble (e.g. RF, SGD)

-Most AI are for large to small dimension pbs

-Learning is on cloud (even testing)

-Energy consumption & Obesity issues

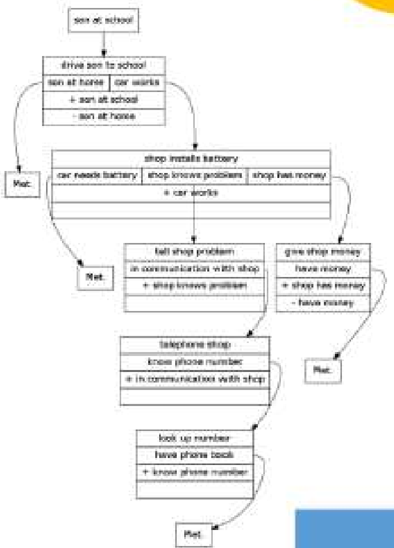
Three AI waves… and two AI winters



Coustesy N. Vayatis

Winters explained

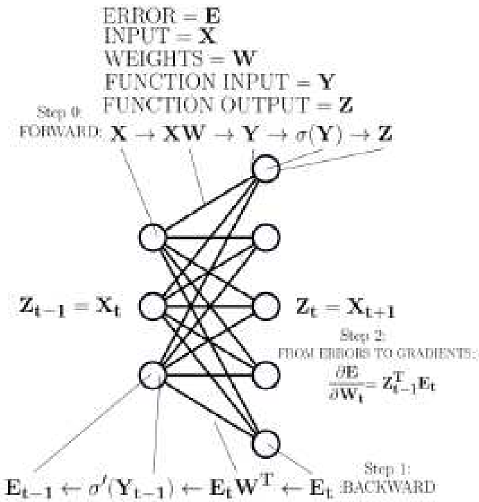
1 Symbolic AI



- Only able to capture explicit knowledge
- Lack of robustness
- Computational burden

2 Connectionist AI

- Lack of data
- Lack of interpretability
- Hard to maintain

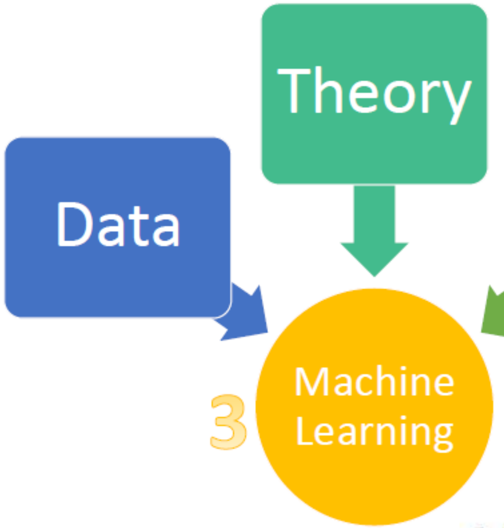


1959 → 1986

Coustesy N. Vayatis

-A child recognizes a cat after seeing one → AI on small data
-Human brain consumes 10W → AI < 1W

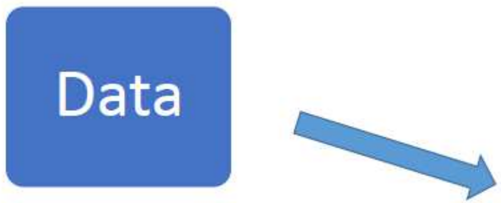
What is different now

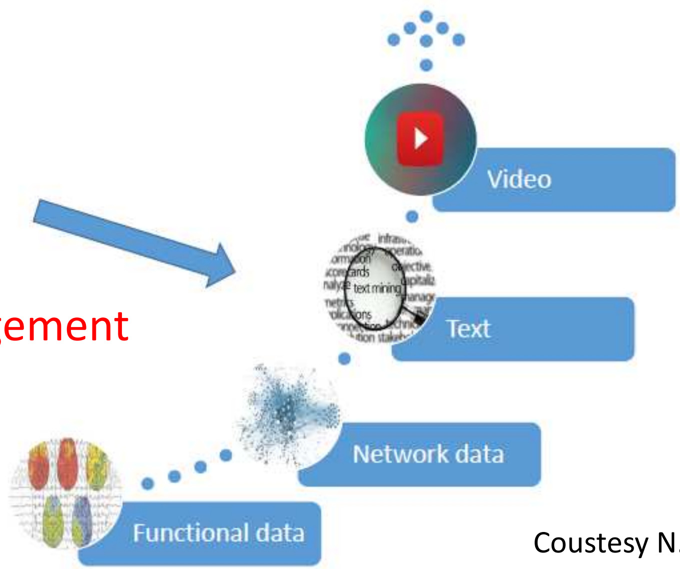


+ cash flow

All is data

Success mainly in B2C





IT for data management

Imputation
Data cleaning
Physical meanings

Coustesy N. Vayatis

DEEP LEARNING PLATFORMS

TensorFlow : Google

PyTorch : Facebook

Keras : Deep Learning libraries and interfaces in Python

Scikit-learn : Inria, Google, +7

Caffe : Berkeley AI Research (BAIR)

DL4j: Skymind

cuDNN: Nvidia

MatConvNet: Mathworks

…

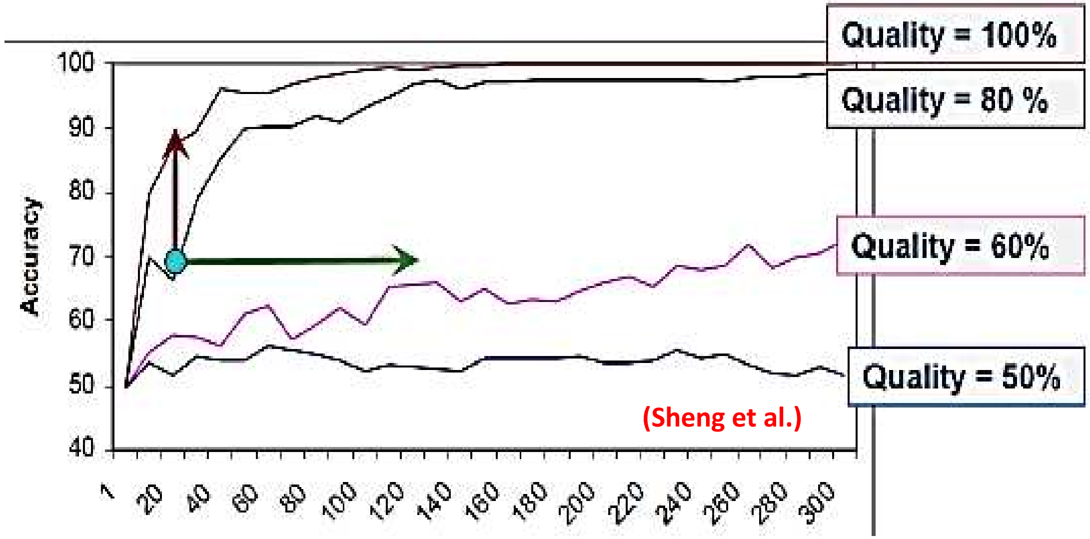
Energy consumption distribution
in
cloud-based embedded AI

IoT	Network	Servers
1/3	1/3	1/3

Edge

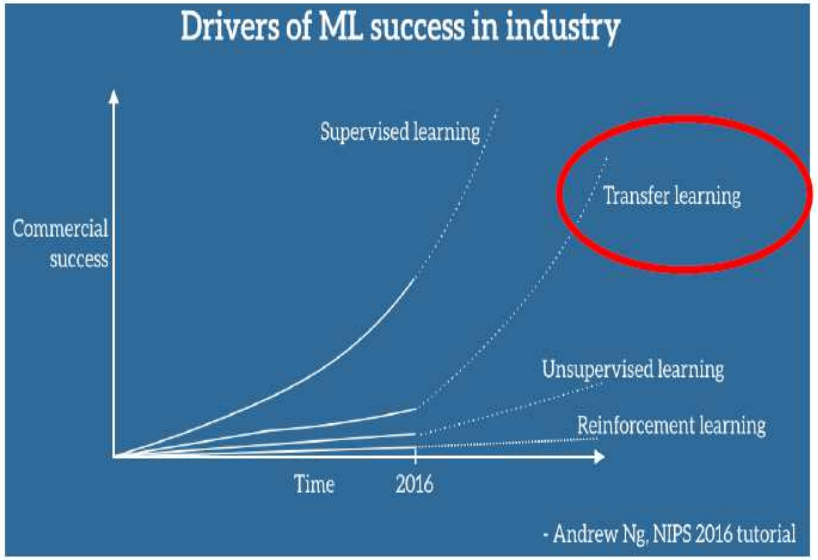
Cloud

Learning accuracy bounded by learning database quality rather than size



This is a limit to transfer learning as flaws in data is common when using data from one field in another

Sampling bias is the rule in industry



Efficient reinforcement learning will require embedded learning

NB: Andrew Ng is VP & Chief Scientist of Baidu, Co-Chairman and Co-Founder of Coursera, and an Adjunct Professor at Stanford University.

Methods:

-10-15 Y : SVM, KNN

-5-10 Y : RF, XGBoost

-Today : CNN

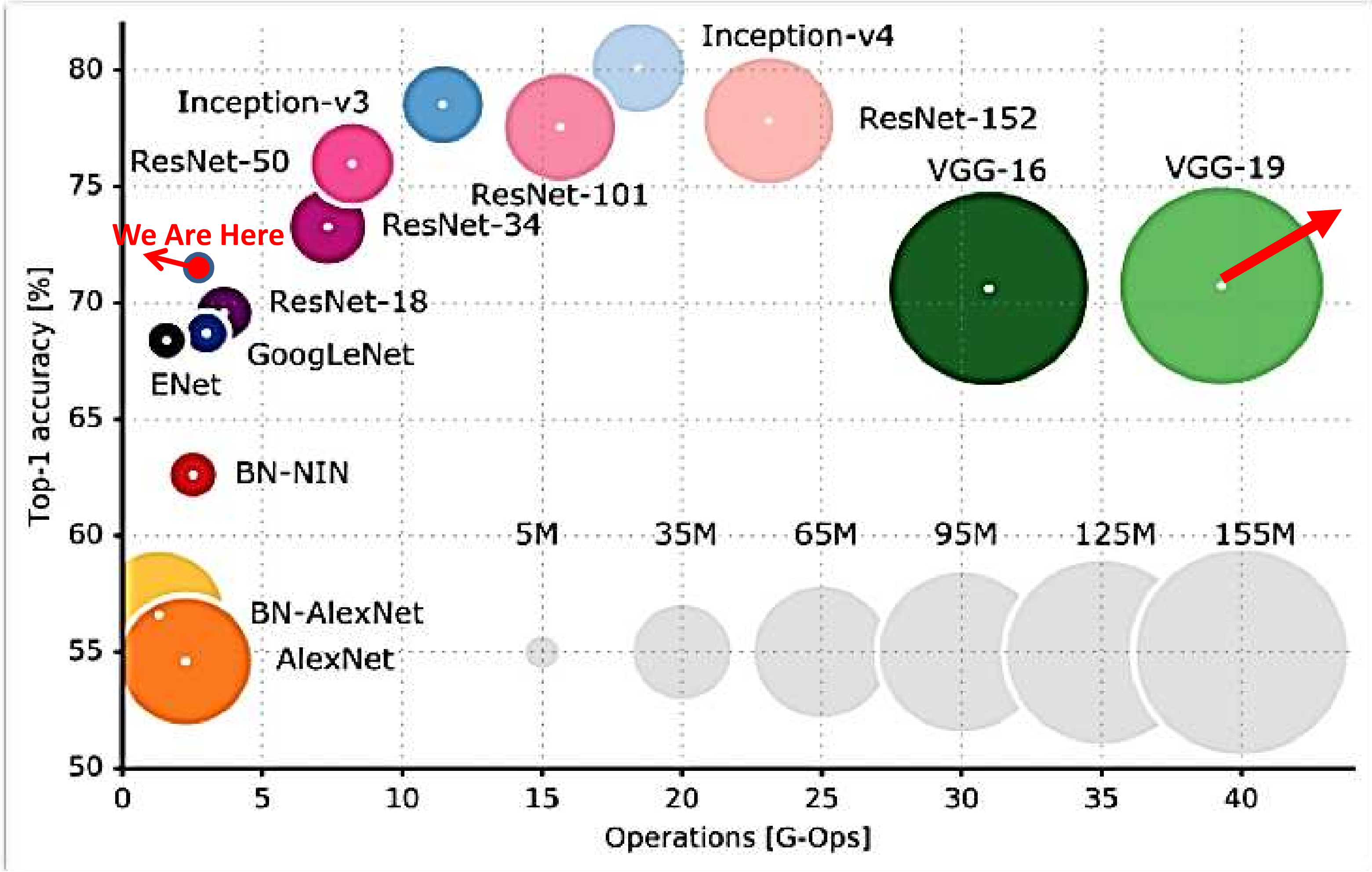
Ref: Simon Knowles, graphcore, 2018

Limitations:

-Large output dimension impossible with RF, SVM

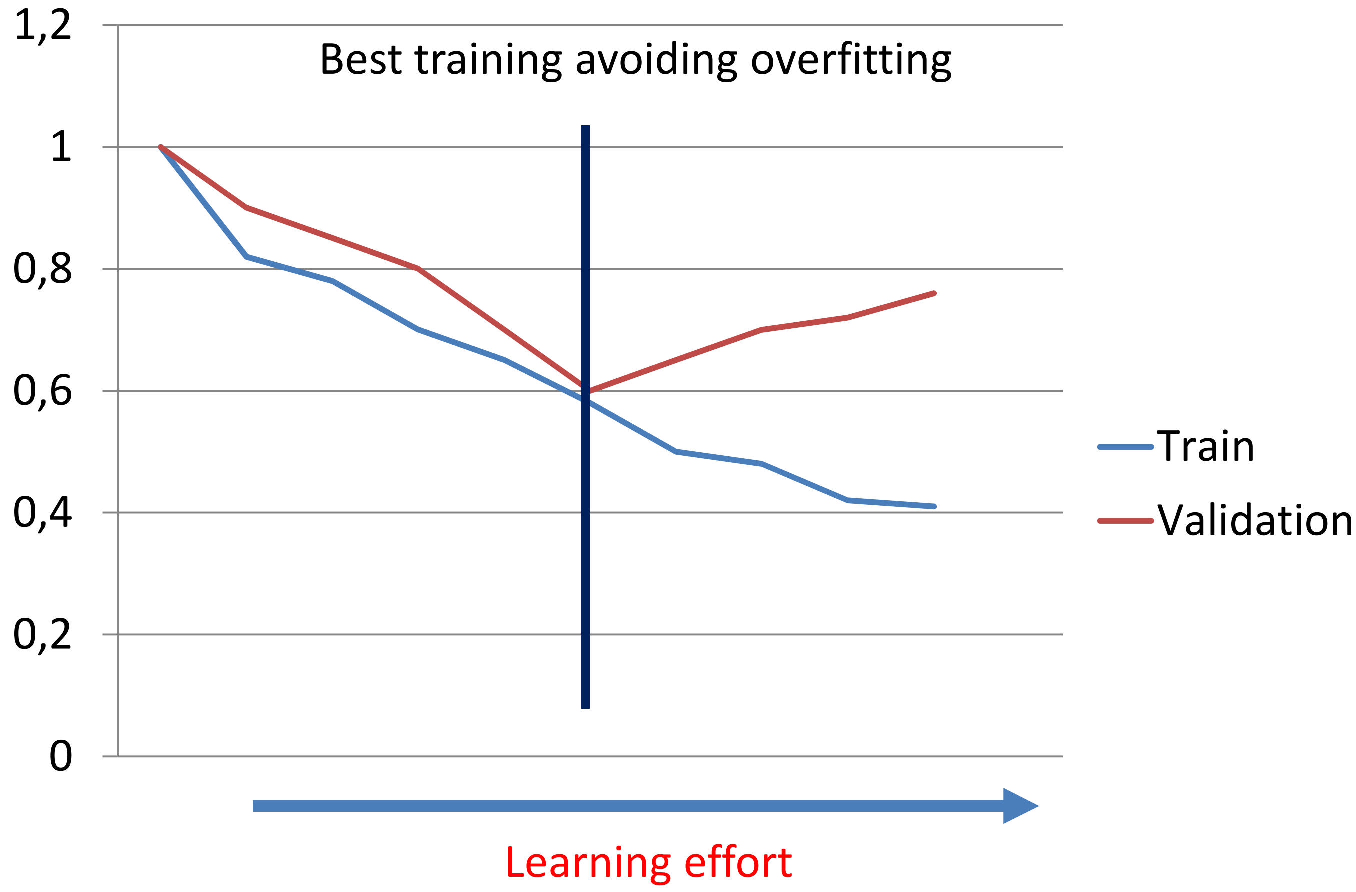
-CNN difficult to build (know-how)

-CNN Obesity issue (VGG-19 : 0.55GB)



Pic source: Eugenio Culurciello, topbots.com

Avoid **overfitting** monitoring **validation error**

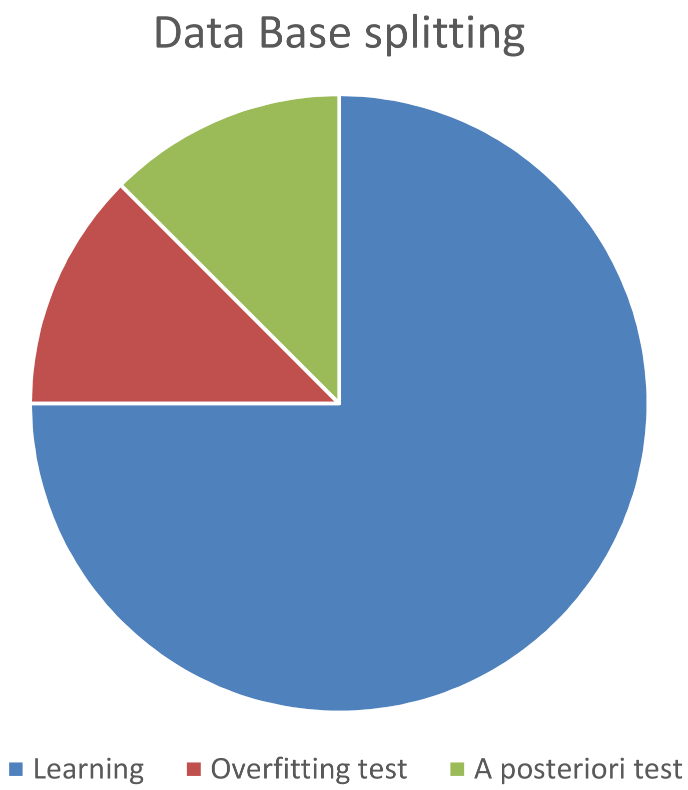


https://en.wikipedia.org/wiki/Comparison_of_deep_learning_software

roNNie.ai	Kevin Lok	MIT license
BigDL	Jason Dai	Apache 2.0
Caffe	Berkeley Vision and Learning Center	BSD
Deeplearning4j	Skymind engineering team; Deeplearning4j community; originally Adam Gibson	Apache 2.0
Chainer	Preferred Networks	MIT license
Darknet	Joseph Redmon	Public Domain
Dlib	Davis King	Boost Software License
DataMelt (DMelt)	S.Chekanov	Freemium
DyNet	Carnegie Mellon University	Apache 2.0
Intel Data Analytics Acceleration Library	Intel	Apache License 2.0
Intel Math Kernel Library	Intel	Proprietary
Keras	François Chollet	MIT license
MATLAB + Neural Network Toolbox	MathWorks	Proprietary
Microsoft Cognitive Toolkit	Microsoft Research	MIT license[25]
Apache MXNet	Apache Software Foundation	Apache 2.0
Neural Designer	Artelnics	Proprietary
OpenNN	Artelnics	GNU LGPL
PaddlePaddle	Baidu	Apache License
PlaidML	Vertex.AI	AGPL3
PyTorch	Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan	BSD
Apache SINGA	Apache Incubator	Apache 2.0
TensorFlow	Google Brain team	Apache 2.0
TensorLayer	Hao Dong	Apache 2.0
Theano	Université de Montréal	BSD
Torch	Ronan Collobert, Koray Kavukcuoglu, Clement Farabet	BSD
Wolfram Mathematics	Wolfram Research	Proprietary
VerAI	VerAI	Proprietary

Data format and destination

Train/validation/test



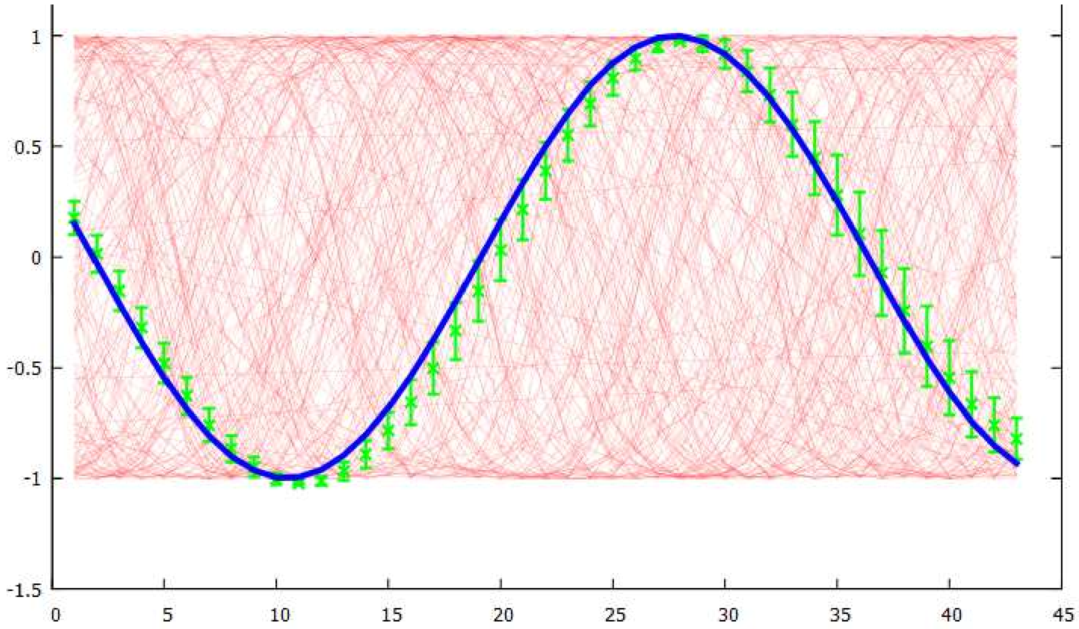
N scenarii of (#n → #m)

$\{X_1^1, \ldots, X_n^1\}$
...
$\{X_1^N, \ldots, X_n^N\}$

$\{Y_1^1, \ldots, Y_m^1\}$
...
$\{Y_1^N, \ldots, Y_m^N\}$

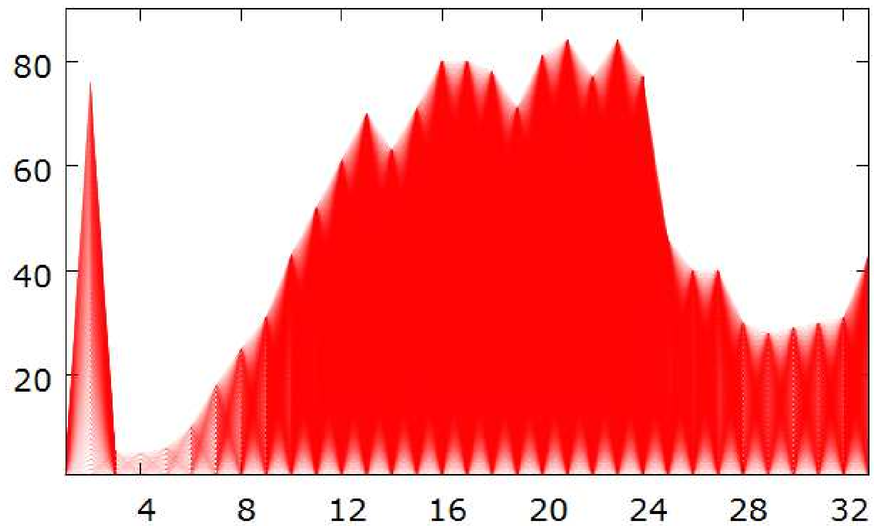
Learning ODE solutions

ndim_in=2, ndim_out=43, nb_obs=200



$$y' = x_1 \cos(x_1 t + x_2)$$
$$y(0) = \sin(x_2)$$

33 layers with up to 80 variables / layer

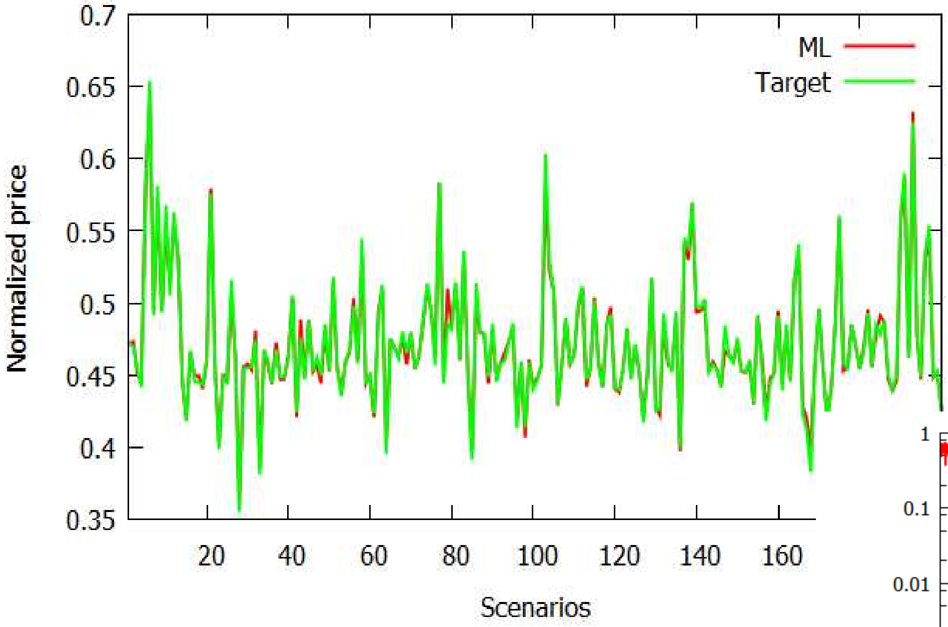


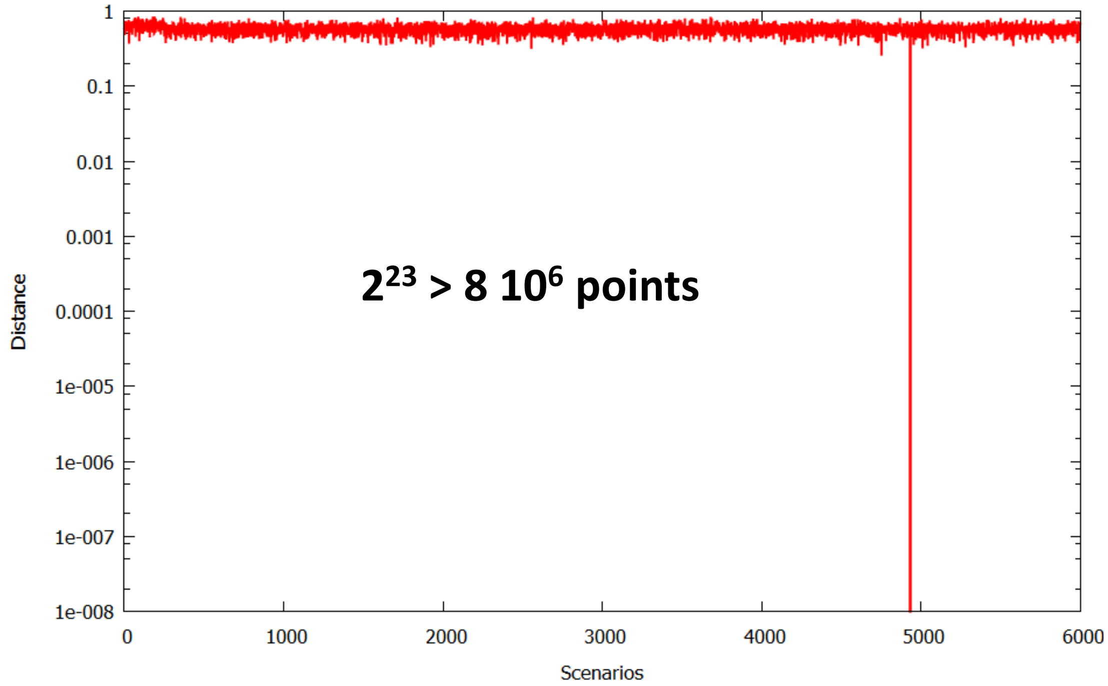
Challenges

https://www.kaggle.com/ ---- https://challengedata.ens.fr/

Exotic pricing with multidimensional non-linear interpolation by Natixis

Training set of 1 million prices to learn how to price a specific type of instruments described by 23 parameters by nonlinear interpolation on these prices. Testing set of 400K scenarios.





Variables

ML works on real variables

Real
Integer
Characters → integer
Categorical → integer or One-hot-encoding

One-Hot-Encoding for both input & output variables

Classification in 10 classes (MNIST) → dimension 10
0 → (1,0,…,0)
2 → (0,0,1,0…,0)
9 → (0,…,0,1)
Also permits to have a probability distribution of belonging to given class

Missing data → Imputation
Median…

Gradients

-Finite differences
-Complex variables
-Adjoint variables
-Automatic differentiation