THALES

The Landscape of Deep Learning

A Quick Overview

Teodora Petrisor
Reasoning and Analysis in Complex Systems Lab.
Thales Research and Technology, France

OPEN

www.thalesgroup.com

Deep Learning

(wiki & scholarpedia) Term coined in 1986 by R. Dechter
adopted wrt artificial neural networks by I. Aizenberg et al in 2000
then widely used wrt neural networks with many hidden layers since ~2006
(Y. Bengio, G. Hinton, Y. LeCun et al)

many, many… many

Réf. : TRT-Fr/STI/LRASC//18,0047 – 08/10/2018
Thales Research & Technology France
Template trtp version 8,0,2 / template : 87211168-GRP-EN-003

Intro

Deep Learning

Deep

What makes it work

Learning

The KEY concept

Deep Learning

Machine Learning (ML) (A. Samuel 1959)

(Mitchell 1997) **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 it**s performance at tasks in T, as measured by P, **improves** with experience E"

Réf. : TRT-Fr/STI/LRASC//18,0047 – 08/10/2018
Thales Research & Technology France
Template trtp version 8,0,2 / template : 87211168-GRP-EN-003

THALES

Intro

Machine Learning (ML)

(Mitchell 1997) **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"

Statistical

Réf. : TRT-Fr/STI/LRASC//18,0047 – 08/10/2018
Thales Research & Technology France
Template trtp version 8,0,2 / template : 87211168-GRP-EN-003

Statistical ML

i.e. Based on data

> Examples of experience

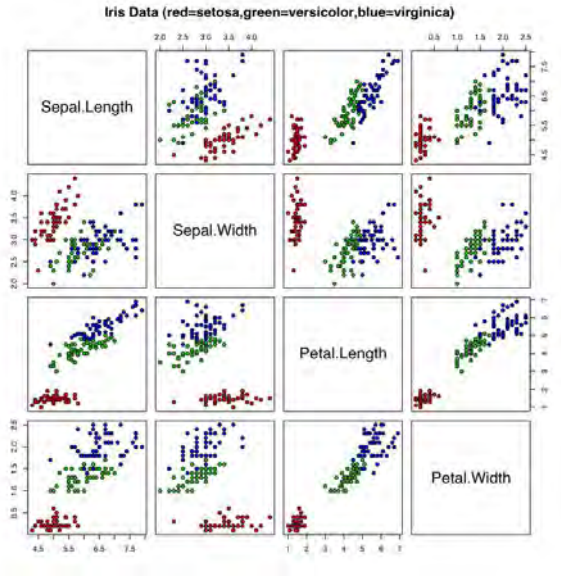
Using statistical tools

> e.g. algos to make predictions

> estimate (complicated) functions

- from (random) observations

For automated decision-making



one of the many goals of Artificial Intelligence (AI)

Réf. : TRT-Fr/STI/LRASC//18,0047 – 08/10/2018
Thales Research & Technology France
Template trtp version 8,0,2 / template : 87211168-GRP-EN-003

THALES

Artificial Intelligence – The (other) Goals

- Perception
- Reasoning
- Knowledge Representation
- Automated Planning and Scheduling
- Natural Language Processing

- Learning

Symbolic vs Data-driven

Réf. : TRT-Fr/STI/LRASC//18,0047 – 08/10/2018
Thales Research & Technology France
Template trtp version 8,0,2 / template : 87211168-GRP-EN-003

THALES

(a parenthesis) Symbolic - Artificial Intelligence (AI)

- **Top-Down Approach**
- **Try to model the world (deterministically)**
- **Using expert knowledge**
  - Rules (predicates)
  - Combination of Rules (using logic)
- **Reasoning**
- **Decision**
- **Action**
- **Planning**

What defines a cat?

IF

$$FUR \wedge (COLOR: isOrange \vee isBlack \vee isGray) \wedge (EARS \in P\{1,2\}) \wedge (EYES \in P\{1,2\}) \wedge NOSE \wedge MOUTH \wedge WHISKERS(?) \wedge (PAWS \in P\{1,2,3,4\}) \wedge [\dots]^*$$

THEN CAT

* Several thousands rules later

(Simple) Reasoning: What is this?
Decision: Recognize that this is a cat
Action: Smile
Plan: Check out other similar pictures

OPEN

Réf. : TRT-Fr/STI/LRASC//18,0047 – 08/10/2018
Thales Research & Technology France
Template trtp version 8,0,2 / template : 87211168-GRP-EN-003

THALES

(a parenthesis) Symbolic - Artificial Intelligence (AI)

- Top-Down Approach
- Try to model the world (deterministically)
- Using expert knowledge
  - Rules (predicates)
  - Combination of Rules (using logic)
- Reasoning
- Decision
- Action
- Planning

This one doesn't play by the rules!

Réf. : TRT-Fr/STI/LRASC//18,0047 – 08/10/2018
Thales Research & Technology France
Template trtp version 8,0,2 / template : 87211168-GRP-EN-003

THALES

(a parenthesis) Symbolic - Artificial Intelligence (AI)

How to automatically define FUR, EARS, …

- Another set of rules?

A good rule is a non-ambiguous rule

A good rule is a generic yet discriminative rule

Some problems are really complex to model

- a picture is worth a 1000 … rules
- or, more formally, high dimensional functions are very hard to describe analytically
  - approximate/estimate…
  - predict
- Iterative approaches

Instead of doing this (or in addition* to doing this)

- Let the Model refine itself progressively from **examples**
  - And hope for the best (no longer controllable term by term)

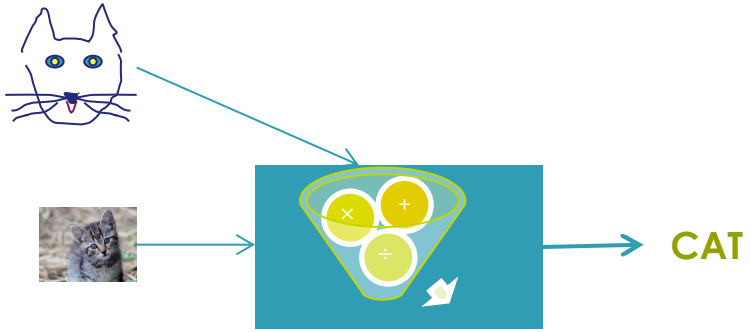
* Disclaimer: Symbolic AI works in many cases … and in general we need a mixture of both!

Réf. : TRT-Fr/STI/LRASC//18,0047 – 08/10/2018
Thales Research & Technology France
Template trtp version 8,0,2 / template : 87211168-GRP-EN-003

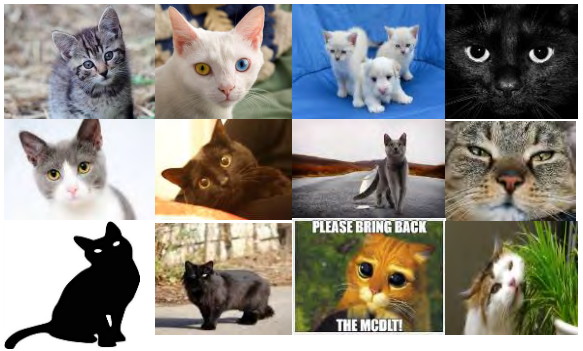
THALES

(this talk) Data Driven - AI

- Generalization
- Inference
- Model emerges from them
- Instead of explicit rules increase knowledge by accumulating examples
- Bottom-up Approach



Trial-and-error process



Réf. : TRT-Fr/STI/LRASC//18,0047 – 08/10/2018
Thales Research & Technology France
Template trtp version 8,0,2 / template : 87211168-GRP-EN-003

OPEN

THALES

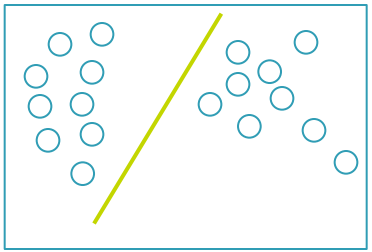
Agenda

- some more **Machine Learning Basics**
- **Artificial Neural Networks**
- **Deep Learning Models**
- **Final industrial Considerations**

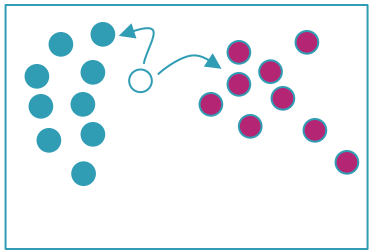
Réf. : TRT-Fr/STI/LRASC//18,0047 – 08/10/2018
Thales Research & Technology France
Template trtp version 8,0,2 / template : 87211168-GRP-EN-003

THALES

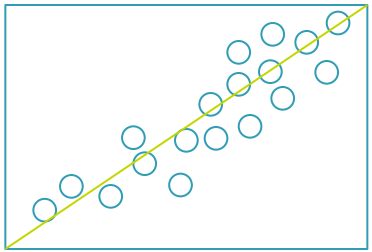
Machine Learning: Tasks, Needs & Approaches



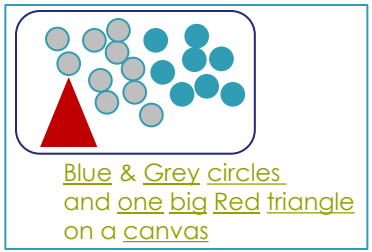
Clustering



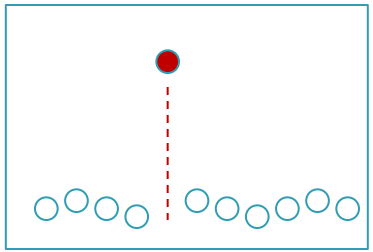
Classification



Regression



Transcription



Anomaly detection

And many more:

machine translation,
density estimation,
interpolation (missing values),
synthesis,
denoising …

Réf. : TRT-Fr/STI/LRASC//18,0047 – 08/10/2018
Thales Research & Technology France
Template trtp version 8,0,2 / template : 87211168-GRP-EN-003

THALES

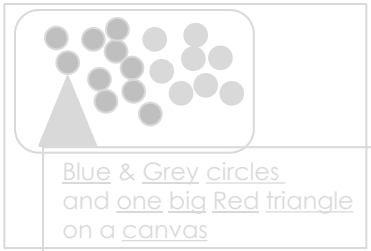
Machine Learning: Tasks, Needs* & Approaches

*algorithmic

Clustering

Classification

Regression



Transcription

Anomaly detection

Balanced datasets capturing the diversity of the subsequent space

- e.g. binary classification: comparable number of positive/negative examples
- e.g. face classification: reduced bias (gender, skin colour, accessories… all represented)
- statistical significance: adequate number of available samples

Diverse datasets

- e.g. images: different angles, views, occlusions, illumination, deformations, clutter

Strategies to avoid underfitting/overfitting

Réf. : TRT-Fr/STI/LRASC//18,0047 – 08/10/2018
Thales Research & Technology France
Template trtp version 8,0,2 / template : 87211168-GRP-EN-003

OPEN

THALES

Machine Learning: Tasks, Needs & Approaches

A Machine Learning problem/solution: a two-stage process:

TRAINING and **TEST** (sometimes called validation)

Each with its own (distinct) dataset

GOAL: minimize error on both (sometimes this fails)

Two major ways of learning

- **SUPERVISED***
- **UNSUPERVISED**

*and a continuum between the two:
weakly supervised, reinforcement, active learning

Supervised Learning: classification, regression, …

- Involves **ground truth** (class labels, reference signal …)
- minimize the error between the obtained output and the expected one

Unsupervised Learning: clustering, …

- No ground truth available: structure emerges by setting metrics on the data
  - E.g. distances between points (smallest inter-cluster distance in conjunction with largest intra-cluster distance)

(stochastic list of some) Methods

Supervised Learning

- Linear regression*
- Logistic regression
- Nearest-neighbours
- Support Vector Machines
- Boosting
- Decision trees
- ….
- Artificial Neural Networks
- ….

Unsupervised Learning

- K-Means
- DBScan
- Expectation Maximization
- Hierarchical clustering
- Principal/Independent Components
- Sparse coding
- ….
- Artificial Neural Networks
- ….

*technically, a method in Statistics

Réf. : TRT-Fr/STI/LRASC//18,0047 – 08/10/2018
Thales Research & Technology France
Template trtp version 8,0,2 / template : 87211168-GRP-EN-003

THALES

Agenda

some more Machine Learning Basics

Artificial Neural Networks

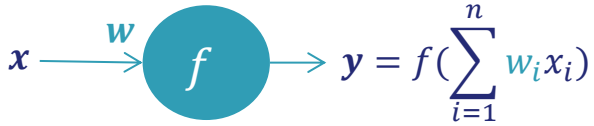
Deep Learning Models

Final Considerations

Réf. : TRT-Fr/STI/LRASC//18,0047 – 08/10/2018
Thales Research & Technology France
Template trtp version 8,0,2 / template : 87211168-GRP-EN-003

THALES

WHAT IS AN ARTIFICIAL NEURAL NETWORK

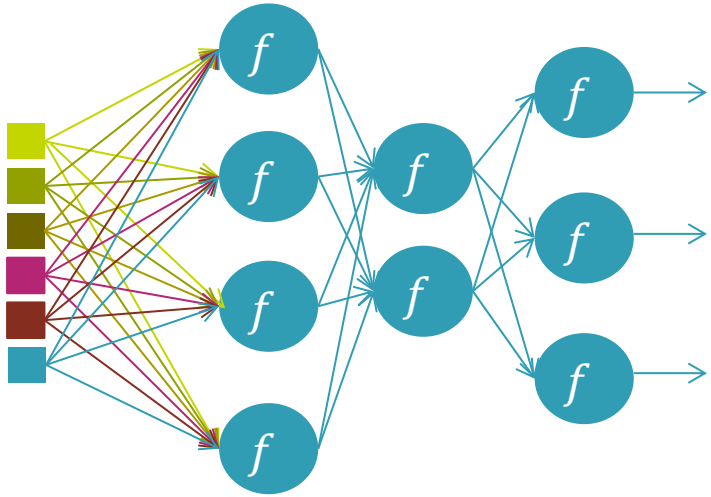


An artificial neuron

[wikipedia] Usual activation functions, f

Identity		$f(x) = x$
Binary step		$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$
Logistic (a.k.a. Sigmoid or Soft step)		$f(x) = \sigma(x) = \frac{1}{1+e^{-x}}$ [1]
TanH		$f(x) = \tanh(x) = \frac{(e^x - e^{-x})}{(e^x + e^{-x})}$
Rectified linear unit (ReLU)[11]		$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$
Leaky rectified linear unit (Leaky ReLU)[12]		$f(x) = \begin{cases} 0.01x & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$

A network of artificial neurons



Input $\boldsymbol{x}$
aka Layer$^{(0)}$

Hidden Layers

Output Layer$^{(n)}$

$$y = f(\boldsymbol{x}) = f^{(n)}(f^{(n-1)}(f^{(n-2)}(\dots(f^{(1)}(\boldsymbol{x})))))$$

How to Make the Network Learn: Error Gradient Backpropagation

Goal: approximate some function f^* i.e. $f(\boldsymbol{x}) \rightarrow f^*(\boldsymbol{x})$

- Classification example:
  - Input example $\boldsymbol{x}$ mapped into an output category: $y = f^*(\boldsymbol{x})$
  - At each point $\boldsymbol{x}$ the network outputs a value $\hat{y}$ close to y
  - $\hat{\boldsymbol{y}} = f(\boldsymbol{x}; \boldsymbol{w})$: find parameters $\boldsymbol{w}$ bringing f the closest to f^*

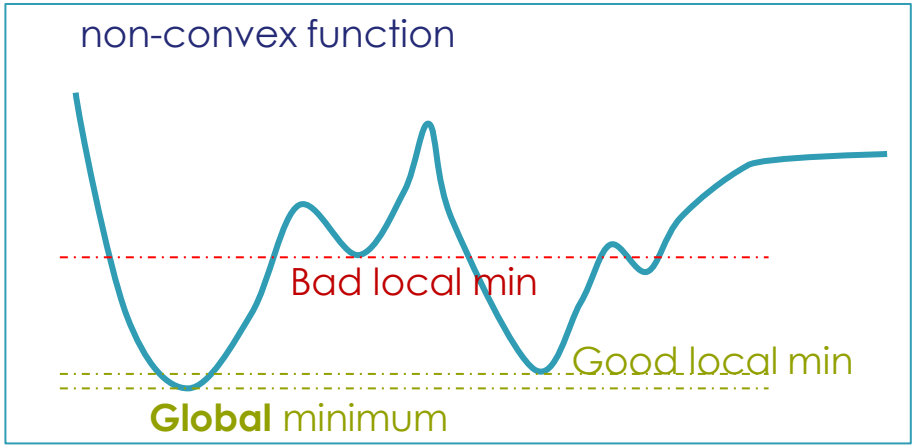
Principle: minimizing a non-convex loss (cost) function

- Forward propagation: compute the loss function
- Back propagation* ⇔ chain rule computation: Jacobian-gradient product from layer to layer
- Learning: (stochastic) gradient descent of the cost function
  - cost function to minimize: the "distance" (error) between the obtained output and the expected one (e.g. the cross-entropy)

* a.k.a backprop

Réf. : TRT-Fr/STI/LRASC//18,0047 – 08/10/2018
Thales Research & Technology France
Template trtp version 8,0,2 / template : 87211168-GRP-EN-003

Tips or tricks



- **Learning rate: too low – slow convergence; too high – (possibly) no convergence**
- **"Plateaux" – can lead to vanishing gradient; "cliffs" – exploding gradient**
  - Most methods introduce some kind of regularisation
    - e.g. Parameter norm

Réf. : TRT-Fr/STI/LRASC//18,0047 – 08/10/2018
Thales Research & Technology France
Template trtp version 8,0,2 / template : 87211168-GRP-EN-003

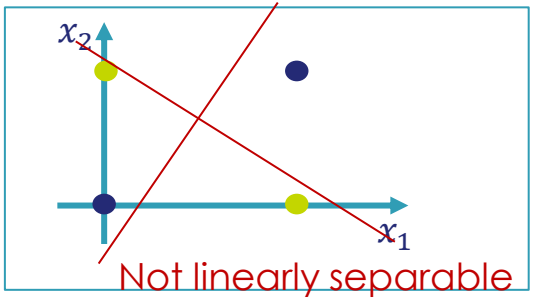
Artificial Neural Networks – The Begining, the Fall-out

brief and laconic History

- 1943 McCulloch&Pits: artificial neuron
- 1958 Rossenblat: perceptron linear model $y = w^T x$
- 1960 Widrow and Hoff: ADALINE
- 1980 Fukushima: Neocognitron
- 1988 Rumelhart et al, 1989 LeCun et al: backpropation
- 1991 Hinton: Multi Layer Perceptron for speech
- 1998 LeCun: LeNet-5

Fall-out

- The XOR problem
- The universal approximation theorem
- Support Vector Machines doing better and simpler
- Lacking computational power for learning



Réf. : TRT-Fr/STI/LRASC//18,0047 – 08/10/2018
Thales Research & Technology France
Template trtp version 8,0,2 / template : 87211168-GRP-EN-003

OPEN

THALES

Agenda

some more Machine Learning Basics

Artificial Neural Networks

Deep Learning Models

Final industrial Considerations

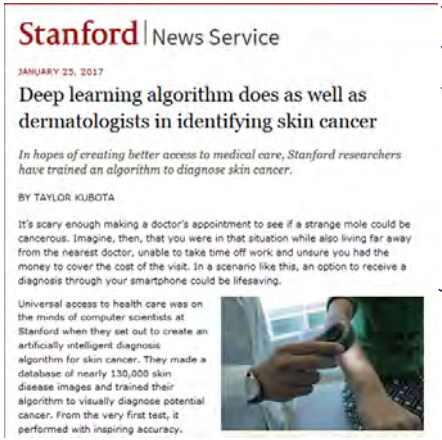
Artificial Neural Networks – The Begining, the Fall-out and the Revival

from [Krizhevsky et al 2012] ImageNet Classification

2006: unsupervised learning of representations

- Pre-train each layer
- Train each layer at a time on top of the previously trained
- Supervised training of the global architecture (fine tuning)

Hinton et al, 2006; Bengio et al. 2007; Ranzato et al, 2007



Stanford | News Service

JANUARY 25, 2017

Deep learning algorithm does as well as dermatologists in identifying skin cancer

In hopes of creating better access to medical care, Stanford researchers have trained an algorithm to diagnose skin cancer.

BY TAYLOR KUBOTA

It's scary enough making a doctor's appointment to see if a strange mole could be cancerous. Imagine, then, that you were in that situation while also living far away from the nearest doctor, unable to take time off work and unsure you had the money to cover the cost of the visit. In a scenario like this, an option to receive a diagnosis through your smartphone could be lifesaving.

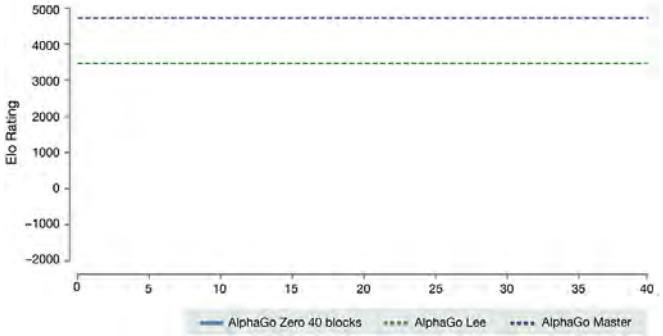
Universal access to health care was on the minds of computer scientists at Stanford when they set out to create an artificially intelligent diagnosis algorithm for skin cancer. They made a database of nearly 130,000 skin disease images and trained their algorithm to visually diagnose potential cancer. From the very first test, it performed with inspiring accuracy.

Source: https://news.stanford.edu/press-releases/2017/01/25/artificial-intelleitfy-skin-cancer/

Input sentence:	Translation (PBMT):	Translation (GNMT):	Translation (human):
李克強此行將啟動中加總理年度對話機制，與加拿大總理杜魯多舉行兩國總理首次年度對話。	Li Keqiang premier added this line to start the annual dialogue mechanism with the Canadian Prime Minister Trudeau two prime ministers held its first annual session.	Li Keqiang will start the annual dialogue mechanism with Prime Minister Trudeau of Canada and hold the first annual dialogue between the two premiers.	Li Keqiang will initiate the annual dialogue mechanism between premiers of China and Canada during this visit, and hold the first annual dialogue with Premier Trudeau of Canada.

Source: https://ai.googleblog.com/2016/09/a-neural-network-for-machine.html
Google Neural Machine Translation system, 2016

https://deepmind.com/blog/alphago-zero-learning-scratch/



OPEN

THALES

Deep Learning: What's New

DEPTH: Deep Learning starts at 3 hidden layers

- Complex concepts are a hierarchical composition of simple concepts (presumably similar to humans)

… and now goes to hundreds of layers

Optimisation algorithms:

- Stochastic gradient descent (SGD); SGD with momentum; approximate second order methods

Meta algorithms

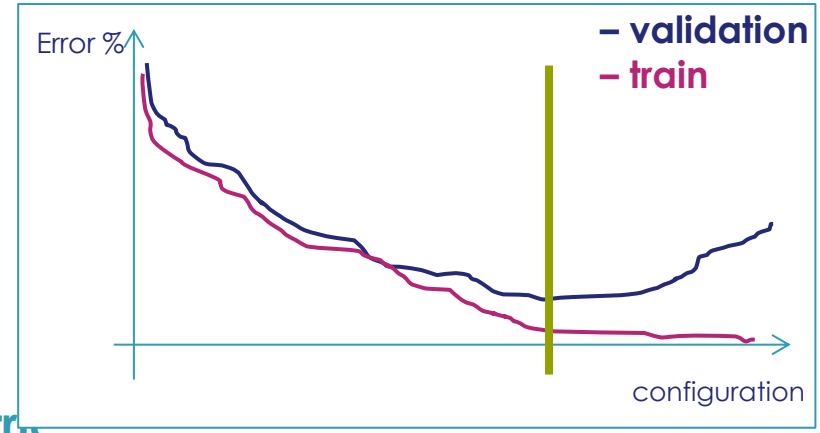
- Adaptive moments – Adam – (2014): adaptive learning rate optimization
  - Other: AdaGrad (2011), RMSProp (2012)
- Batch normalization (2015): adaptive reparametrization, z-norm of a minibatch of activations
- Dropout (2014): randomly drop units during training

Réf. : TRT-Fr/STI/LRASC//18,0047 – 08/10/2018
Thales Research & Technology France
Template trtp version 8,0,2 / template : 87211168-GRP-EN-003

How to make Deep Learning work

Hyperparameter tuning (manual/automatic)

- Network topology (width = number of units per layer; depth = number of layers)
- Activation function(s)
- Learning rate



Additional dataset – for validation

Choose the appropriate performance metric

- e.g. Accuracy vs. receiver operating characteristic (binary classification)

Regularization

Réf. : TRT-Fr/STI/LRASC//18,0047 – 08/10/2018
Thales Research & Technology France
Template trtp version 8,0,2 / template : 87211168-GRP-EN-003

THALES

Agenda

some more Machine Learning Basics

Artificial Neural Networks

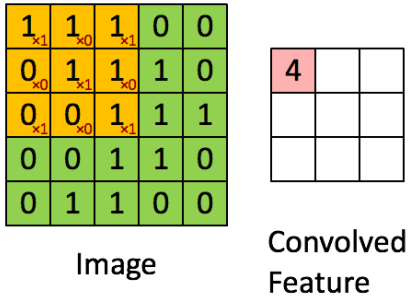
Deep Learning Models

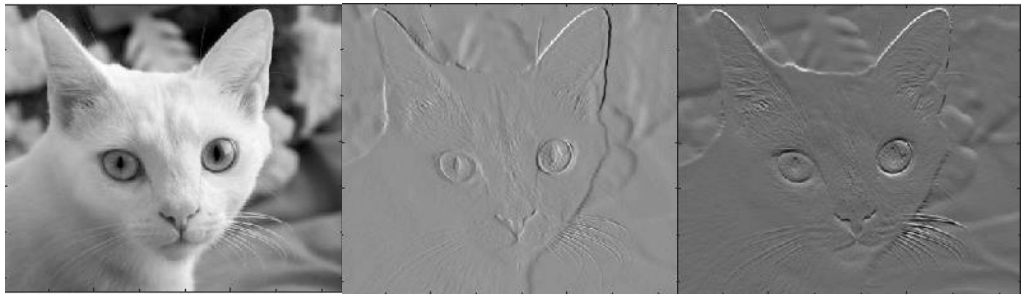
- Only a very few

Final industrial Considerations

Convolutional Neural Networks

What is a convolution





Input: $x(i,j)$ $g_v(i,j) = x(i,j) - x(i+1,j)$ $g_h(i,j) = x(i,j) - x(i,j+1)$

source:
http://deeplearning.stanford.edu/wiki/index.php/Feature_extraction_using_convolution

- **Particular architecture with much sparser connections**
- **Exploit signal particularities (e.g. spatial locality in images)**
  - Learn filters that represent an input as a progressively richer abstraction
  - Inspiration from the visual system

Convolutional Neural Networks

The basic architecture

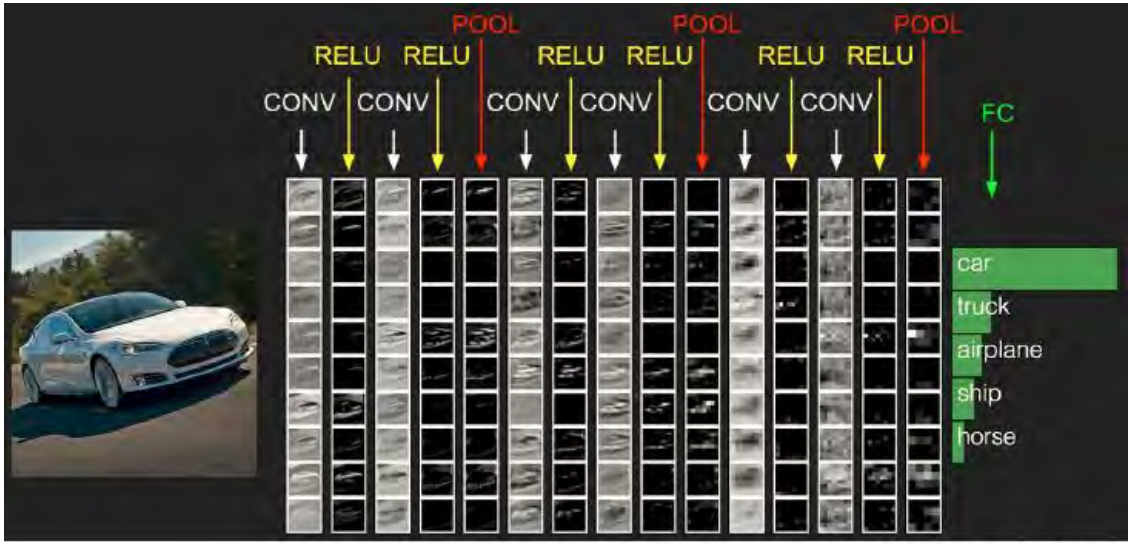


Figure source: A. Karpathy

Three types of "processing"

- Convolution layers: very few connections (kernel size)
- Pooling layers: further signal subsampling with a non-linear function – summary statistics of neighbouring outputs
- Dense layer(s)

Weight sharing

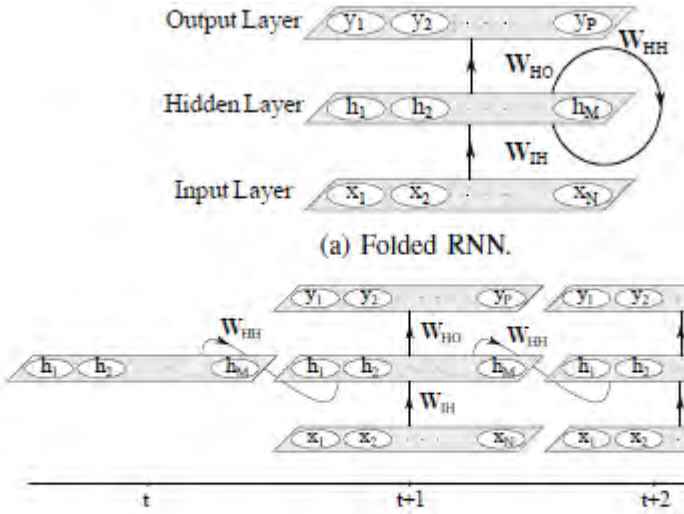
Recurrent Networks

Another class of data: sequential data (time series, text, speech)

- Learn temporal dependencies
- Related to convolution across a temporal (1D) sequence: but this is shallow

Parameter sharing through recurrence

- No longer feed-forward: current output depends on previous output
- Gradient backpropagation through time



(a) Folded RNN.

(b) Unfolded RNN through time.

Fig. 1: A simple recurrent neural network (RNN) and its unfolded structure through time t. Each arrow shows a full connection of units between the layers. To keep the figure simple, biases are not shown.

from [Salehinejad et al, 2017] - Recent Advances in Recurrent Neural Networks

Autoencoders (AE), Variational AE

Basic principles

- A two-part neural network: encoder and decoder

Initially used for dimension reduction of feature learning

Now: generative models

- Used for dataset augmentation
- Used for denoising

Training:

- Minibatch gradient descent; backpropagation
- Recirculation

Réf. : TRT-Fr/STI/LRASC//18,0047 – 08/10/2018
Thales Research & Technology France
Template trtp version 8,0,2 / template : 87211168-GRP-EN-003

Generative/Graphical Models

Key idea: Stochastic

Use graph theory to infer relations between random variables

- Hidden units are probabilistic
- Undirected models (causality is not obvious)
- Propagate distributions in a structured model

Some are energy-based models

- To enforce the assumption that the state of the input have non-zero probability
  - Boltzmann distribution

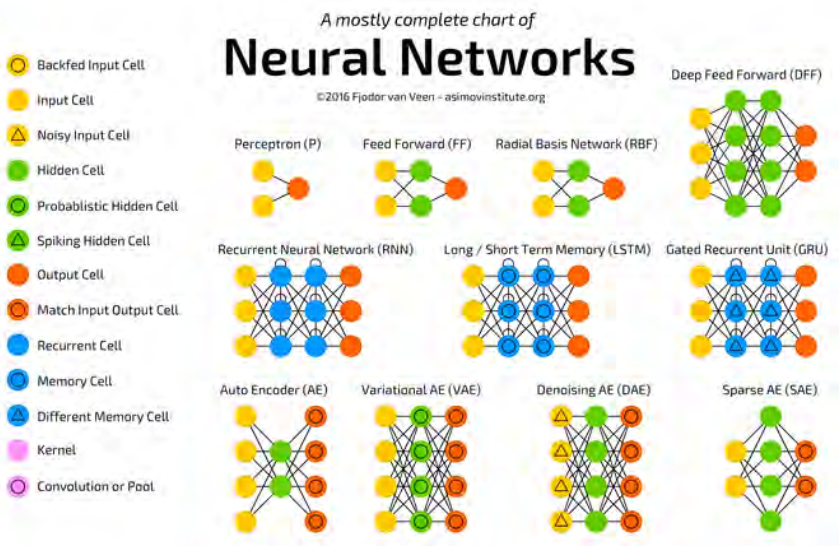
Training with contrastive divergence (Hebbian like)

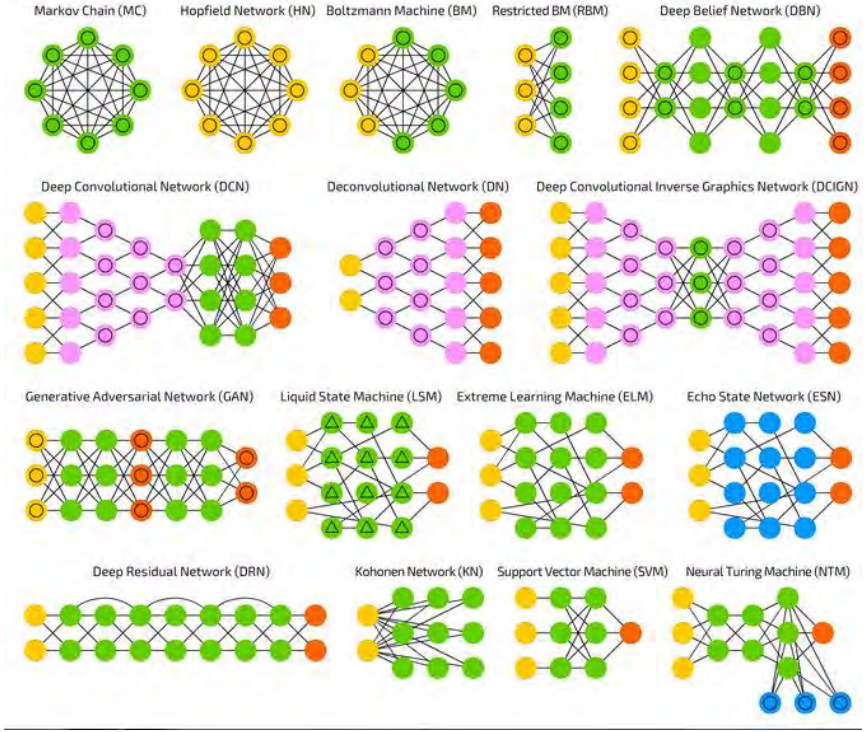
Boltzmann Machine (1985), Deep Belief Nets (Restricted Boltzmann Machine)

Generative Adversarial Nets

Réf. : TRT-Fr/STI/LRASC//18,0047 – 08/10/2018
Thales Research & Technology France
Template trtp version 8,0,2 / template : 87211168-GRP-EN-003

And many more …





From: http://www.asimovinstitute.org/neural-network-zoo/

Agenda

some more Machine Learning Basics

Artificial Neural Networks

Deep Learning Models

Final industrial Considerations

Réf. : TRT-Fr/STI/LRASC//18,0047 – 08/10/2018
Thales Research & Technology France
Template trtp version 8,0,2 / template : 87211168-GRP-EN-003

THALES

Challenges

Implementation

- Training vs Inference
- And performance: how to get it in/with non main-stream tasks/datasets

Energy

- Again: Training vs Inference
- But also: Money, Resources (natural)

Explicability

- Let the human know how the decision has been made

Trust

- Let the human know that the decision is always the same

Réf. : TRT-Fr/STI/LRASC//18,0047 – 08/10/2018
Thales Research & Technology France
Template trtp version 8,0,2 / template : 87211168-GRP-EN-003

THALES

Transfer Learning

Image classification example: not plug and play

- Train a neural network on a main-stream dataset/problem
- Use the result to improve generalization on another

Extreme cases:

- One-shot learning
- Zero-shot learning

Multimodal learning

Assumption Hidden units learn to represent causal factors explaining the data

Réf. : TRT-Fr/STI/LRASC//18,0047 – 08/10/2018
Thales Research & Technology France
Template trtp version 8,0,2 / template : 87211168-GRP-EN-003

THALES

Explainable AI

Trustable AI

Source: Ribeiro et al – *"Why should I trust you?" Explaining the predictions of any classifier*, KDD 2016

(a) Original Image (b) Explaining *Electric guitar* (c) Explaining *Acoustic guitar* (d) Explaining *Labrador*

Interpretable

Source: Goodfellow et al – *Explaining and harvesting adversarial examples*, ICLR 2015

Robust

x

"panda"
57.7% confidence

$\text{sign}(\nabla_x J(\theta, x, y))$

"nematode"
8.2% confidence

$x + \epsilon\text{sign}(\nabla_x J(\theta, x, y))$

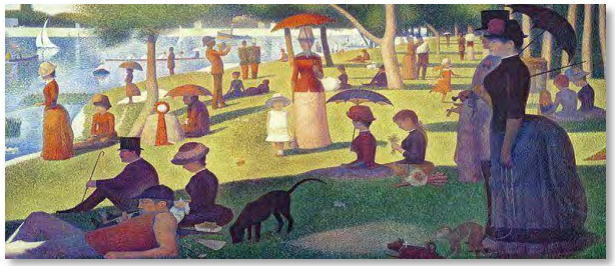
"gibbon"
99.3 % confidence

Réf. : TRT-Fr/STI/LRASC//18,0047 – 08/10/2018
Thales Research & Technology France
Template trtp version 8,0,2 / template : 87211168-GRP-EN-003

OPEN

THALES

Thales Research and Technology meets Deep Learning



Georges Seurat
Un dimanche après-midi à l'île de la Grande Jatte
Institut d'art, Chicago

Source: Thales internal communication based on Gatsy et al. 2015 – A neural algorithm of artistic style

Réf. : TRT-Fr/STI/LRASC//18,0047 – 08/10/2018
Thales Research & Technology France
Template trtp version 8,0,2 / template : 87211168-GRP-EN-003

THALES

Sources & Further Reading

- http://deeplearning.net/
- Juergen Schmidhuber 2015, http://www.scholarpedia.org/article/Deep_Learning
- I. Goodfellow, Y. Bengio, A. Courville – *Deep Learning*, MIT Press 2016, ISBN: 9780262035613
- Y. Bengio – *Learning Deep Architectures for AI*, Foundations and Trends in Machine Learning, Vol 2., no. 1, 2009
- https://adeshpande3.github.io/The-9-Deep-Learning-Papers-You-Need-To-Know-About.html