Deep Learning in Practice

Guillaume Charpiat
Victor Berger & Wenzhuo Liu

TAU team, LRI, Paris-Sud / INRIA Saclay

... and guests!

Overview

- Course summary and organization
- Chapters overview

Context

- Deep learning: impressive results in the machine learning literature
- yet difficult to train, and still poorly understood;
  results = black-boxes missing explanations.
- Huge societal impact of ML today (assistance in medicine, hiring process, bank loans...)
  $\Longrightarrow$ explain their decisions, offer guarantees?
- Real world problems: usually do not fit standard academic assumptions (data quantity and quality, expert knowledge availability...).
- This course: aims at providing insights and tools to address these practical aspects, based on mathematical concepts and practical exercises.

Organisation and evaluation

- Most courses: a lesson + practical exercises (evaluated)
- Extras: guest talks, Jean Zay visit (1000 GPUs cluster), ...

Schedule

8 classes of 3 hours, most often on Monday mornings (9h – 12h15 with a break) at CentraleSupelec (not every week, check the webpage for details).

Webpage & mailing-list: `https://www.lri.fr/~gcharpia/deeppractice/`

Prerequisite

- The introduction to Deep Learning course by Vincent Lepetit (1st semester)
- Notions in information theory, Bayesian statistics, analysis, differential calculus

Links with other Deep Learning courses

- Introduction to Deep Learning (V. Lepetit) : prerequisite

- L'apprentissage par réseaux de neurones profonds (S. Mallat)
- Fondements Théoriques du deep learning (F. Malgouyres & al)
- Apprentissage Profond pour la Restauration et la Synthese d'Images (A. Almansa & al)

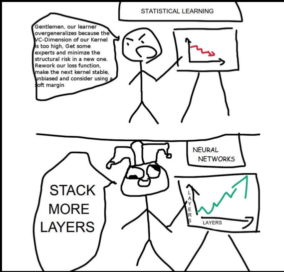
- Modélisation en neurosciences et ailleurs (J-P Nadal)
- Object recognition and computer vision (Willow team & al)

- Our course: understanding and tools to make NN work in practice with a focus on architecture design, explainability, societal impact, real datasets and tasks (e.g. small data, limited computational power vs. scaling up, RL...).
  $\implies$ negligible overlap

Outline

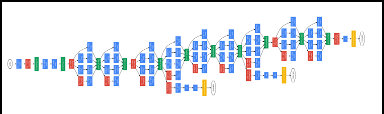
Deep learning vs. classical ML and optimization – January 13th

- Going Deep or not?
  - Examples of successes and failures of deep learning vs. classical techniques (random forests)
  - Approximation theorems vs. generalization [3, 4]
  - Why deep: ex. of depth vs. layer size compromises (explicit bounds)
- Gap between classical Machine Learning and Deep Learning
  - Forgotten Machine Learning basics (Minimum Description Length principle, regularizers, objective function different from evaluation criterion) and incidental palliatives (drop-out, early stopping, noise)
- Hyper-parameters and training basics
  - List of practical tricks
  - **Practical session** (exercise to give before early February) **(bring your laptop!)**



Architectures – February 3rd

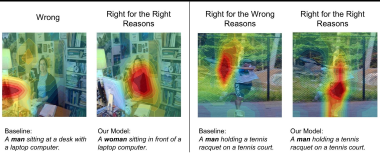
- Architectures as priors on function space
  - Change of design paradigm
  - Random initialization
- Architecture zoo
  - Reminder (CNN, auto-encoder, LSTM, adversarial...)
  - Dealing with scale & resolution (fully-convolutional, U-nets, pyramidal approaches...)
  - Dealing with depth (ResNet, Highway networks) and mixing blocks (Inception)
  - Attention mechanisms
  - GraphCNN



- Problem modeling
- Guest talk by Yaroslav Nikulin (start-up Therapixel): Deep learning for breast cancer detection

Interpretability – February 10th

- At stake: the example of medical diagnosis, and societal issues with black-box algorithms [5]



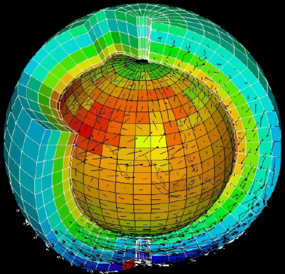
- Interpretability of neural networks
  - Analyzing the black-box
    - at the neuron level: filter visualisation, impact analysis
    - at the layer level: layer statistics...
    - at the net level: low-dimensional representation (t-SNE) + IB
    - by sub-task design: "explainable AI"
  - Adversarial examples & remedies
- Issues with datasets
  - Biases in datasets : 4 definitions of fairness
    - Getting invariant to undesirable dataset biases (e.g. gender in CVs / job offers matching)
    - Ensuring errors are uniform over the dataset
  - Differential privacy (database client protection)
- Visualization tools

Small data, weak supervision and robustness

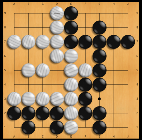
- Modes of supervision
  - Learning from synthetic data [12]
  - Learning from scratch vs. Transfer learning
  - Semi-supervised learning [11]
  - Self-supervised learning (ex: video prediction)
  - Multi-tasking
- Exploiting known invariances or priors
  - Example of permutation invariance: "deep sets" [8], applied to people genetics
  - Spatial/Temporal coherence
  - Choosing physically meaningful metrics, e.g. optimal transport (Sinkhorn approximation)[9]
  - Dealing with noisy supervision (noisy labels)
- Transfer learning

Incorporating physical knowledge / Learning physics

- Course featuring Michele Alessandro Bucci (LRI, TAU team) and Lionel Mathelin (LIMSI, Paris-Sud)
  - Data assimilation
  - Learning a PDE (equation not known)
  - Incorporating invariances/symmetries of the problem
  - Knowing an equation that the solution has to satisfy: solving PDEs!
  - Form of the solution known
  - Deep for physic dynamics : learning and controlling the dynamics



Deep Reinforcement Learning



- Guest course by Olivier Teytaud (Facebook FAIR Paris)
- Crash-course about deep RL. . .
- . . . until alpha-0!

GPU clusters

- Presentation of Jean Zay, the GPU super-cluster for French academia, by IDRIS
- Optional visit to the cluster and practical session (job scheduler, etc.)

Generative models & Variation inference

- GAN and VAE (Variational Auto-Encoder)
- GAN vs. VAE

Auto-DeepLearning – March 16th

- Guest talk by Zhengying Liu (LRI, TAU team, Isabelle Guyon's group)
- Overview of recent approaches for automatic hyper-parameter tuning (architecture, learning rate, etc.): classical blackbox optimisation, Reinforcement Learning approaches, constrained computational time budget, self-adaptive architectures...
- Additional real-world difficulties: missing data, unstructured data
- Presentation of the Auto-ML & Auto-DL challenges

To attend the course

- go see the website and subscribe to the mailing-list
- bring your laptop, and **install PyTorch, Jupyter and matplotlib beforehand!**
- See you on Monday at 9h, amphi Janet (CentraleSupelec)

Biographies

- Guillaume Charpiat is an INRIA researcher in the TAU team (INRIA Saclay/LRI/Paris-Sud). He has worked mainly in computer vision, optimization and machine learning, and now focuses on deep learning. He conducts studies on neural networks both in theory (self-adaptive architectures, formal proofs) and in applications (remote sensing, people genetics, molecular dynamics simulation, brain imagery, weather forecast, skin lesion medical diagnosis, ...).
- Victor Berger and Wenzhuo Liu are PhD students in the TAU team, working on deep generative models and on graph neural networks.

Bibliographies

- *Why does deep and cheap learning work so well?*, Henry W. Lin, Max Tegmark, David Rolnick. https://arxiv.org/abs/1608.08225
- *Representation Benefits of Deep Feedforward Networks*, Matus Telgarsky. https://arxiv.org/abs/1509.08101
- *On the structure of continuous functions of several variables*, David A. Sprecher.
- *Representation properties of networks: Kolmogorov's theorem is irrelevant*, Federico Girosi and Tomaso Poggio.
- *Weapons of Math Destruction*, Cathy O'Neil.
- *Practical Variational Inference for Neural Networks*, Alex Graves. https://papers.nips.cc/paper/4329-practical-variational-inference-for-neural-network