LIONS EDIC Intro

Volkan Cevher, Associate Professor EPFL

https://lions.epfl.ch
Preface

My research:
  Machine Learning (ML)
  Optimization
  Signal Processing
  Information Theory
  Statistics
  Automatic Control

My current courses:
  Mathematics of Data
  Reinforcement Learning
  Advanced Topics in ML
  Online Learning in Games
Machine Learning (ML) works
Machine Learning (ML) works... but many new challenges A-RISE!

1. **Robustness**
2. **Interpretability**
3. **Bias & Fairness**
4. **Reproducibility**

---

GPT-3 medical chatbot tells suicidal test patient to kill themselves

Rob Beschizza  2 days ago

Researchers experimenting with GPT-3, the AI text-generation model, found that it is not ready to replace human respondents in the chatbox.
Robustness and interpretability are entangled

- Robust ML problems
  a. Adversarial training
  b. Robust reinforcement learning
  c. GAN training

- Interpretability
  a. Integrated gradients
  b. Sparse features
  c. Disentangled representations

Challenges to robustness & interpretability

- Critical challenges to robustness
  a. High-dim. geometry
  b. Spurious attractors
  c. Reproducibility & Verification

Trust but verify

- **Trustworthiness** requires effort
  - a. Geometric boundary
  - b. Algebraic formulations
  - c. Uncertainty estimation

1. **On certifying non-uniform bounds against adversarial attacks.** Liu, Tomioka and Cevher. ICML’19.
2. **Lipschitz Constant Estimation of Neural Network via Sparse Polynomial Optimization.** Latorre, Rolland and Cevher. ICLR’20.
Challenges to Fairness & Bias

• Models
  b. Generative models
  c. Manifold models…

• Regularizers (and optimizers!)
  a. Robustness: Path norm.
  b. Structured (graph) sparsity
  c. Learning-based regularizers

**New projects: Mathematical foundations for RISE of AI**

\[
\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} \Phi(x, y)
\]

### Game Theory
1. Distributionally robust
2. Multi-player setting
3. (Mixed) Nash equilibrium
4. Universal & adaptive

### Geometry
1. Trust but verify
2. NAS & Algorithm
3. Explicit bias
4. Limit cycles

### Adaptation
1. Enhance interpretability
2. Model-based adaptation
3. Reinforcement learning
4. Representation theory

---

LIONS EDIC | Volkan Cevher | [https://lions.epfl.ch](https://lions.epfl.ch)
Scalable, rigorous, and reproducible optimization

Game Theory

1. Distributionally robust
2. Multi-player games
3. (Mixed) Nash equilibria
4. Universal & adaptive
Geometric and algebraic foundations for NNs

Geometry

1. Trust but verify
2. NAS & Algorithm
3. Explicit bias
4. Limit cycles

Geometry is determined by:
- raw input
- fair or disentangled representations
- causal representations.

radius $\propto \frac{1}{\text{Lips}}$

NN verification with Lipschitz constants

$\ell_0$-norm perturbation

Model-based perturbation

$\Rightarrow$ interpretable robustness = fairness

SCALABILITY CHALLENGE!

POLYNOMIAL

Positivity

Certification

= $O(\eta^d)$

depth

# of neurons

MUST EXPLOIT STRUCTURE:

- Sparse connections
- Symmetry (invariance)
- Algebraic constraints

I'm working on Theory of Deep Learning.

For more than two layers, right?
Interpretability: New technical developments

Adaptation

1. Enhance interpretability
2. Model-based adaptation
3. Robust RL
4. Representation theory

\[ \Delta x \Delta \rho \geq \frac{\hbar}{2} \]
New applications: Robust domain adaptation (DA), Causality, GNNs

- Mathematical formulations for Robust DA
  - formulation challenges
  - algorithmic challenges

- Causality
  - Joint with Amazon Research
  - Nonlinear models

- Knowledge graphs & graph generation
  - Scalable graph inference
  - Scientific ML & circuit applications

Ideal profile

- Pareto frontier of theory & crack coding
  - super collaborative group (within EPFL and outside)
  - strong connections to industry (Amazon, Google,...)

- Academic track & research labs
  - Faculty@Rice, NUS, Umea (2), Zhejiang, UNC, Linkoping, AIMS, UoCB, VNU, NTU, Technion, UCL, Alborg
  - Postdoc@ETHZ (3), McGill, Turing, UWisc, University of Geneva
  - Others@Kandou bus, SwissRE, TUM, Meta
Ideal profile

- Basic, future-proof research
  - NeurIPS, ICML, ICLR, SIAM Opt, Math Prog,
    - [https://lions.epfl.ch](https://lions.epfl.ch)
    - [http://csrankings.org/#/index?mlmining&world](http://csrankings.org/#/index?mlmining&world)
    - [https://ml.epfl.ch](https://ml.epfl.ch)
Ideal profile

- Basic, future-proof research
  - NeurIPS, ICML, ICLR, SIAM Opt, Math Prog,...
    - [https://lions.epfl.ch](https://lions.epfl.ch)
    - [http://csrankings.org/#/index?mlmining&world](http://csrankings.org/#/index?mlmining&world)
    - [https://ml.epfl.ch](https://ml.epfl.ch)

- Interest in integrated teaching & research
  - Mathematics of data
  - Reinforcement learning
  - Advanced topics in ML
  - Online learning in games
Thank you!

Visit us at ELD 243 and have coffee with my group members!

Also, open house next Friday in our lab!