Machine Learning and Computer Vision

I|S|T AUSTRIA

Institute of Science and Technology

About us...

Christoph

Nikola (since 2018)
“Robust and Trustworthy Machine Learning”

Alex (since 2019, with Dan Alistarh)
“Compression in Deep Network”

Paul (since 2019; PD)
“Deep generative models of 3D scenes”

Bernd (since 2020)
“Interpretable Computer Vision Models”

Jonny (since 2021)
“Trustworthy federated learning”

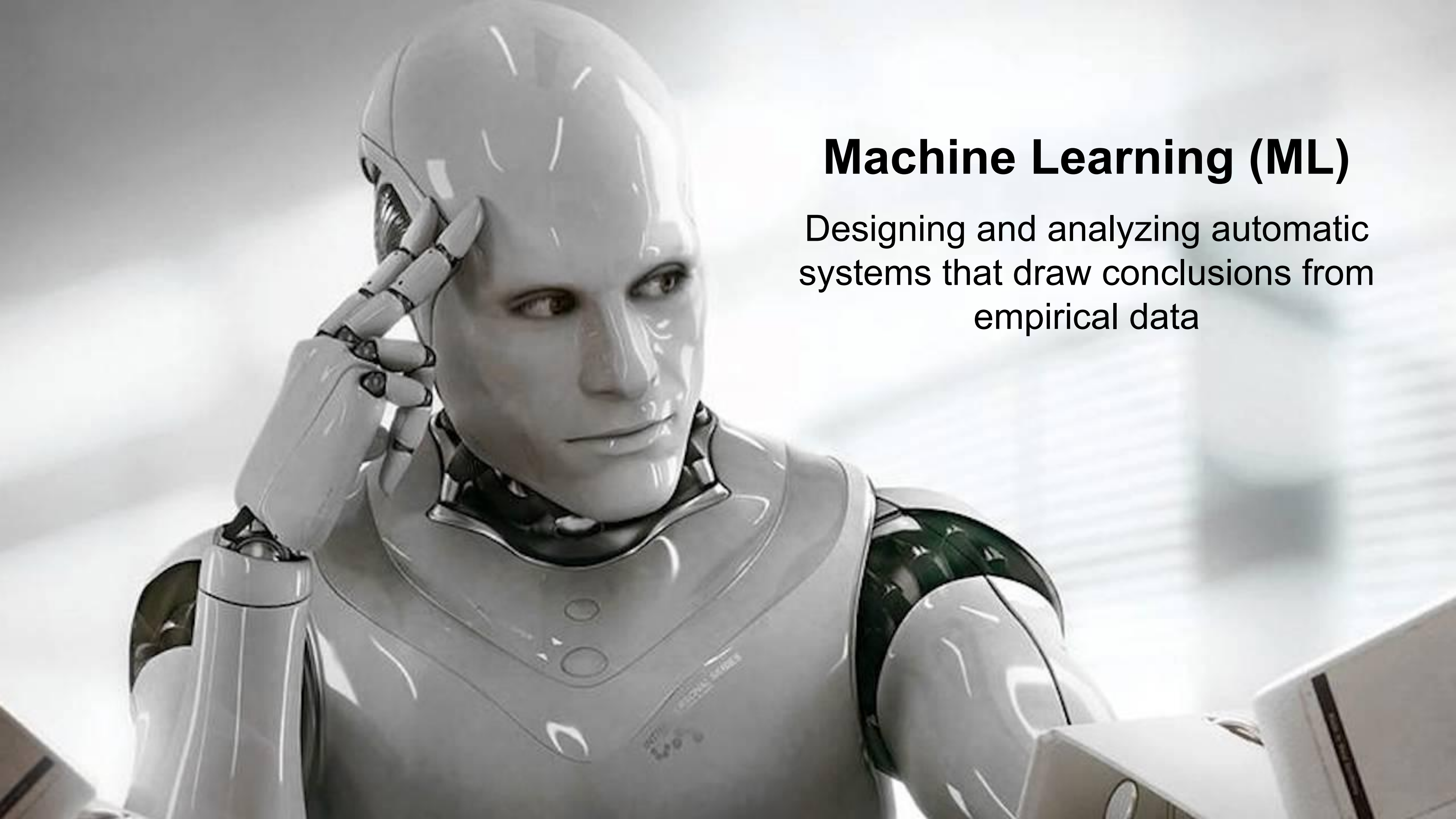
you?

PhD Alumni...

 **Viktoriia (PhD 2015)** now at Imperial College London	 **Asya (PhD 2016)** now at Bayer Research, Berlin	 **Alex (PhD 2018)** now at Google Brain, Zurich
 **Alex (PhD 2018)** now at Amazon Research, Berlin	 **Amelie (PhD 2020)** now at Qualcomm Research, Amsterdam	 **Mary (PhD 2021)** currently intern at Deepmind, London

About us...

central office building, 3rd floor



Machine Learning (ML)

Designing and analyzing automatic systems that draw conclusions from empirical data

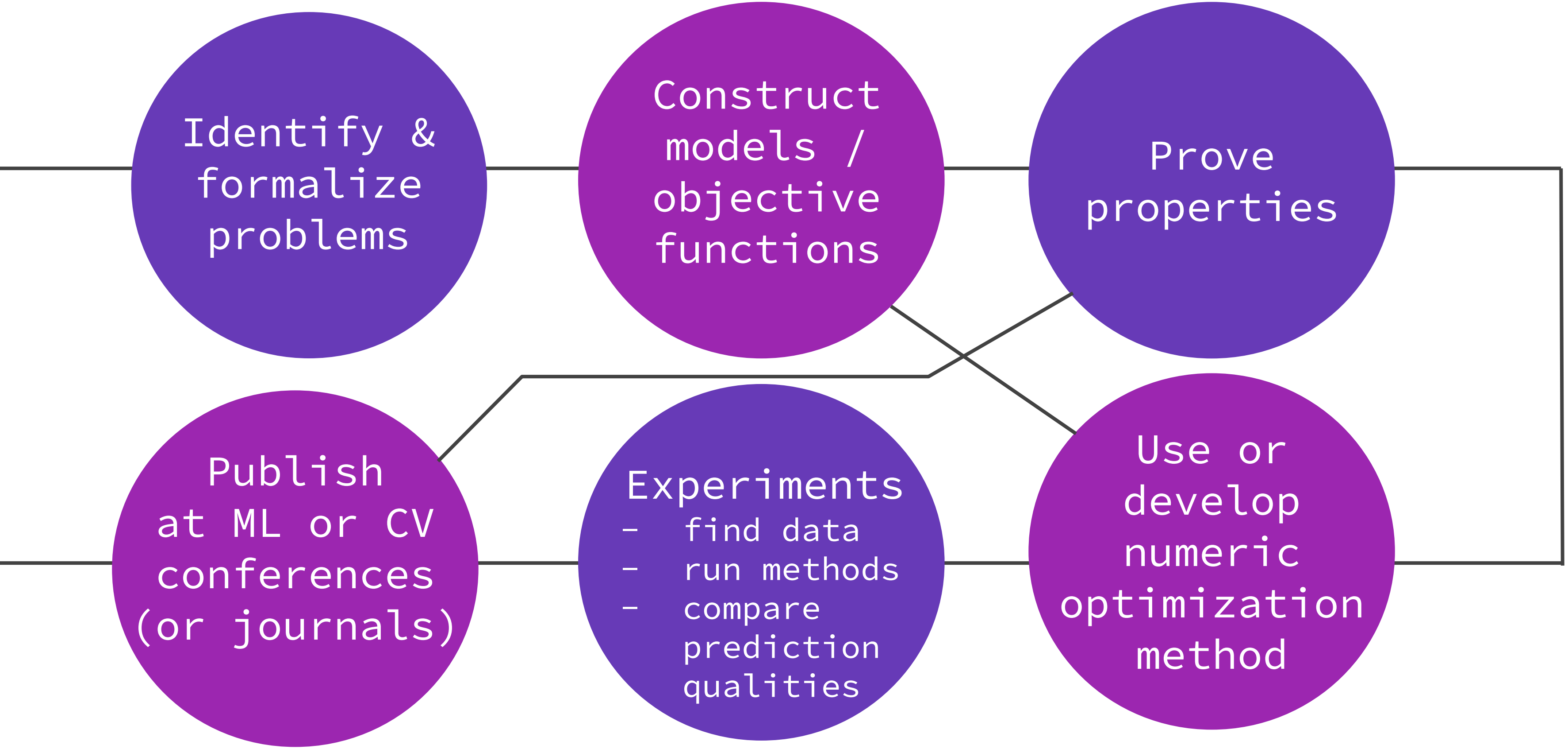
Computer Vision (CV)

Designing and analyzing automatic systems that autonomously process visual data

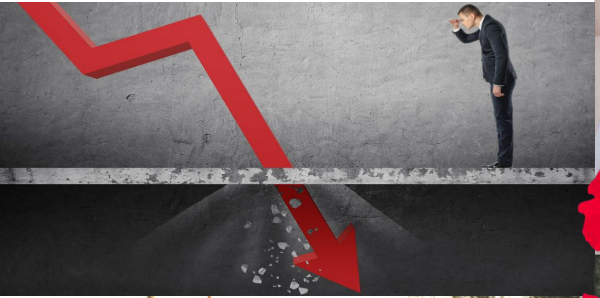
"Three men sit at a table in a pub, drinking beer. One of them talks while the other two listen."

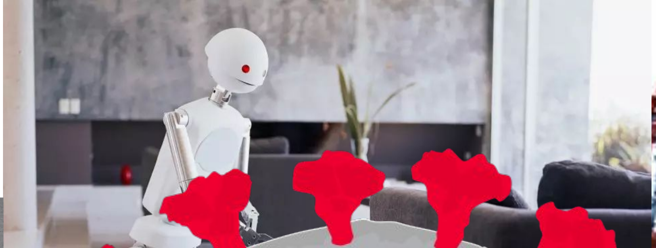
Image: British Broadcasting Corporation (BBC)

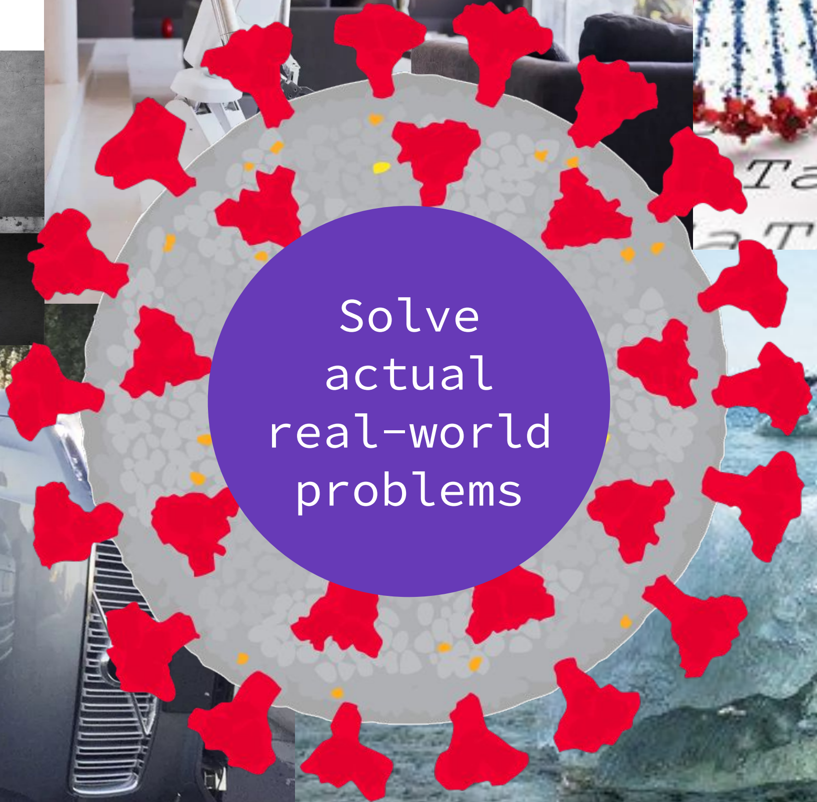
What we do


Identify & formalize problems
Construct models / objective functions
Prove properties
Publish at ML or CV conferences (or journals)
Experiments
- find data
- run methods
- compare prediction qualities
Use or develop numeric optimization method

What we don’t do





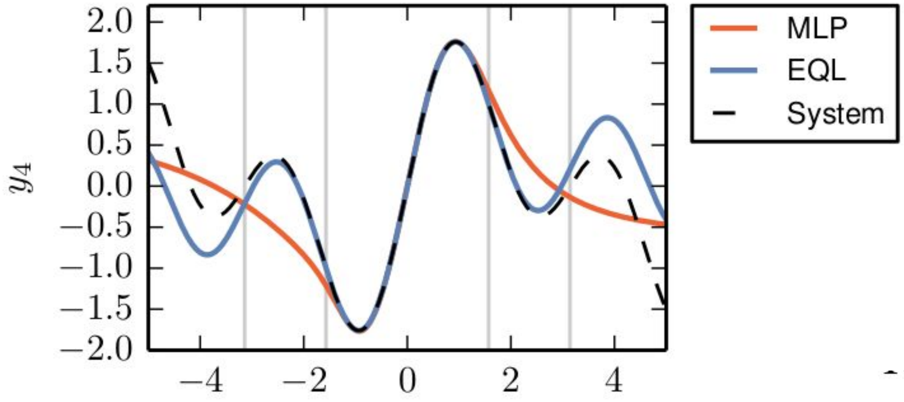


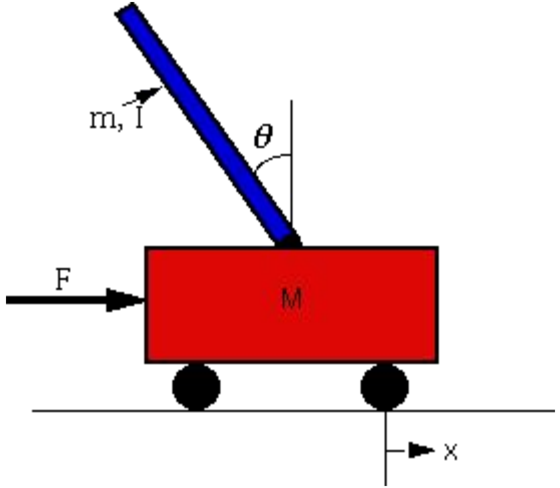
Examples

identify a problem

Extrapolation and learning equations

(Georg Martius, CHL, *ICML 2018*)



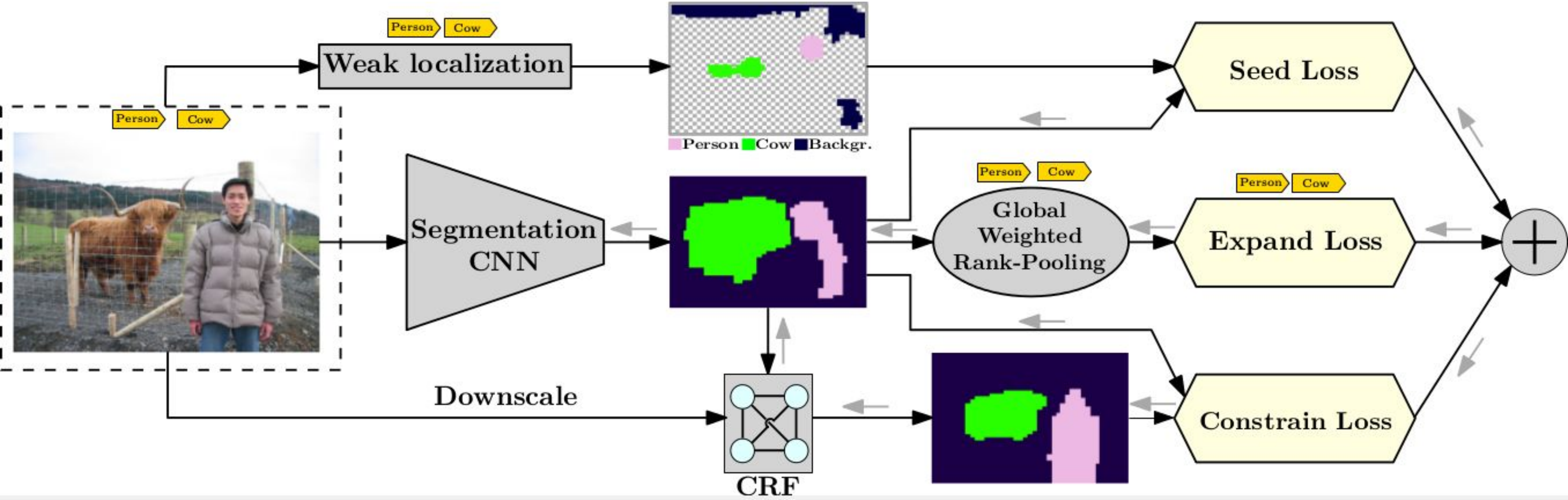


Regression methods typically find functions that **interpolate** well between observed values.
Can we learn systems that **extrapolate** well, e.g. by identifying underlying physical equations?

construct a model

Seed, Expand and Constrain: Three Principles for Weakly-Supervised Image Segmentation

(Alex Kolesnikov, CHL, *ECCV 2016*)



construct an objective function

Active Task Selection for Multi-Task Learning

(Asya Pentina, CHL, *ICML 2017*)

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{i\in I}\alpha_i^t \operatorname{disc}(S_t, S_i) + \frac{A}{T}\|\alpha\|_{2,1} + \frac{B}{T}\|\alpha\|_{1,2}$$

prove properties

Learning from Untrusted Sources

(Niko Konstantinov, CHL, *ICML 2020*)

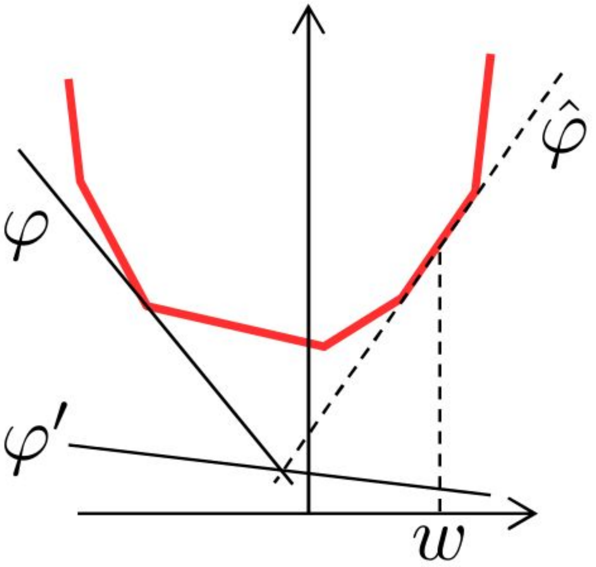
Corollary 2. *In the setup of Theorem 1, against any fixed-set adversary, it holds that*

$$\mathcal{R}(\mathcal{L}(\mathcal{A}(S^{'}))) - \min_{h\in\mathcal{H}} \mathcal{R}(h) \leq 4\mathfrak{R}_G + 6\sqrt{\frac{\log(\frac{4}{\delta})}{2km}} \tag{8}$$

$$+\alpha\left(18\sqrt{\frac{\log\left(\frac{4N}{\delta}\right)}{2m}} + 12\max_{i\in[N]} \mathfrak{R}_i\right).$$

find or develop (continuous) optimization method

Multi-Plane Block-Coordinate Frank-Wolfe Algorithm for Training Structural SVMs with a Costly max-Oracle (Neel Shah, Vladimir, CHL, *CVPR 2015*)



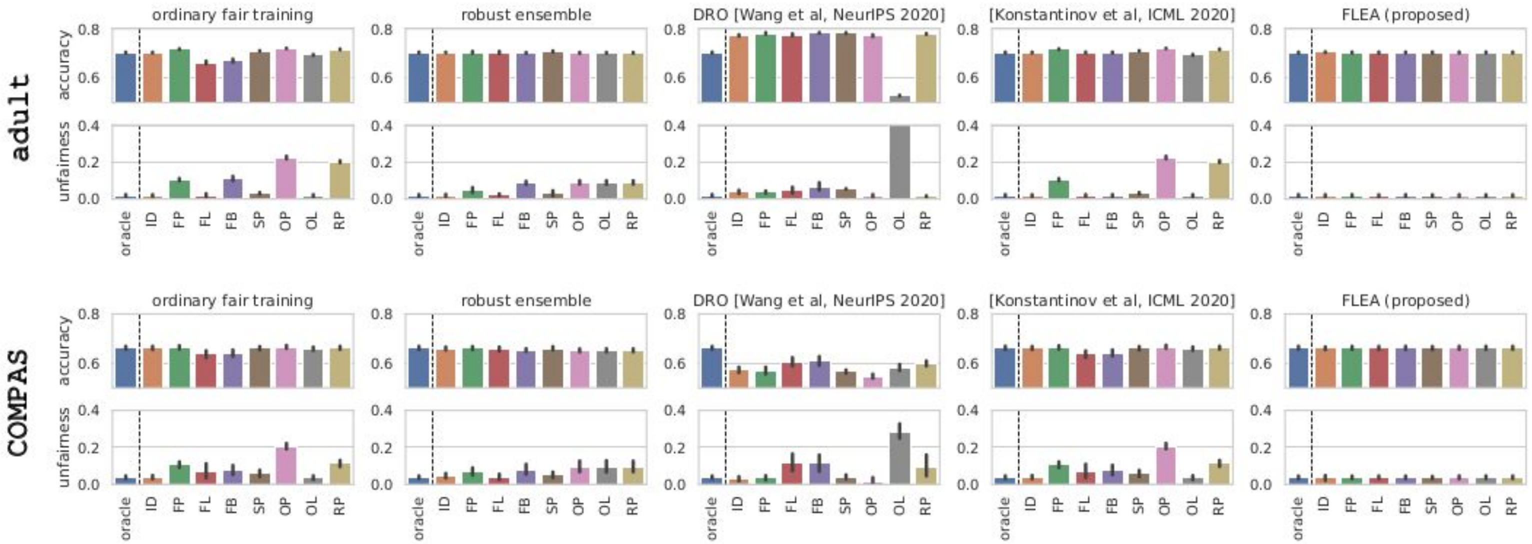
Algorithm 1 Frank-Wolfe algorithm for the dual of (4)

1: set $\varphi \leftarrow \varphi^{\bar y}$ for some $\bar y \in \overline{\mathcal{Y}}$
2: **repeat**
3: compute $w \leftarrow \arg\min_w \frac{\lambda}{2}||w||^2 + \langle \varphi, [w\ 1]\rangle$;
the solution is given by $w = -\frac{1}{\lambda}\varphi_\star$
4: call oracle for vector w: compute $\hat\varphi \leftarrow \arg\max\limits_{\varphi^{\bar y}:\bar y\in\overline{\mathcal{Y}}} \langle \varphi^{\bar y}, [w\ 1]\rangle$
5: compute $\gamma \leftarrow \arg\max_{\gamma\in[0,1]} \mathcal{F}((1-\gamma)\varphi + \gamma\hat\varphi)$ as follows:
set $\gamma \leftarrow \frac{\langle \varphi_\star - \hat\varphi_\star, \varphi_\star\rangle - \lambda(\varphi_\circ - \hat\varphi_\circ)}{||\varphi_\star - \hat\varphi_\star||^2}$ and clip γ to $[0,1]$
set $\varphi \leftarrow (1-\gamma)\varphi + \gamma\hat\varphi$
6: **until** some stopping criterion

experiments
- find data and baselines
- run method
- evaluate quality

FLEA: Provably Fair Multisource Learning from Unreliable Training Data

(Jen Iofinova, Niko Konstantinov, CHL, *under review*)



Publish at CV or ML conferences (or journals)

Conferences (double blind, peer-reviewed, prestigious):

- Neural Information Processing Systems (NeurIPS)
- International Conference on Machine Learning (ICML)
- International Conference on Learning Representations (ICLR)
- Computer Vision and Pattern Recognition (CVPR)
- International Conference on Computer Vision (ICCV)
- European Conference on Computer Vision (ECCV)

Journals:

- Journal of Machine Learning Research (JMLR)
- IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)
- International Journal of Computer Vision (IJCV)

Amelie, CHL, “Classifier Adaptation at prediction time”, CVPR 2015]

Concepts that we use frequently

probability
random variables, expected values, Bayes' rule, concentration inequalities

linear algebra / calculus
function spaces, inner products, gradients, convexity

numerics/ continuous optimization
gradient-based, stochastic

public data sources
images or text, downloaded from the web

Concepts that we use rarely

classical statistics
parametric data distributions, p-values

physical intuition
differential equations, dynamical system

biological intuition
"how does the brain do it?"

complex algorithms
recursion, computational complexity classes

Machine Learning concepts that we frequently use

- supervised learning
- transfer learning
- empirical risk minimization
- convolutional networks

Machine Learning concepts that we rarely use

- generative adversarial networks
- reinforcement learning
- recurrent neural networks
- artificial general intelligence

Potential Rotation Topics

<u>If you consider affiliating with my group</u>	**<u>If you do not consider affiliating with my group</u>**
A topic that • shows what PhD research in our group is like, • builds on your prior knowledge, • ideally is useful for your actual PhD topic. Examples: • *“Metric learning for face recognition”* • *“Compression bounds for deep networks”* • “*Online guarantees for lifelong learning*”	A topic that • provides insight into CV/ML research • builds on your prior knowledge, • ideally is useful for your actual PhD topic. Examples: • Biology: *“Image processing for ant tracking”* • Cryptography: *“Learning with encrypted data”*, • Computer Graphics: *“Segmenting Meshes”*

Prerequisites

- **Mathematics**: Probability, Linear Algebra, Calculus
- **Computer Science**: Programming, preferably in Python (except for “theory” rotations)

Useful Courses

- **Track core courses:**
  - *"Data Science and Scientific Computing"*

  or
  - *"Computer Science"*

- **Fall 1:** *"Statistical Machine Learning"* (myself; inverted classroom format)
- **Fall 1:** *"Information Theory (for Data Science)"* (Marco Mondelli)
- **Fall 2:** *"Methods of Data Analysis"* (Gasper Tkacik)
- **Fall 2:** *"Probability in High Dimension"* (Jan Maas)
- **Spring?:** *"Probabilistic Graphical Models"* (myself, Paul Henderson; inverted)

Public Events

- **"Tea talks"** (15 min. talk series)
- **PhD status talks**
- **Reading/Writing/Coding group(s)**

Join us at the "ELLIS Presents: ML@IST" event on October 6th, 15:00

Contact

- open door (when on campus)
- or send me email: `chl@ist.ac.at`

If you would like to do a rotation in our group and haven't contacted me yet, please do so **today**!