Reliable Machine Learning with Distributional Robustness

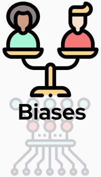
Franck IUTZELER
Univ. Toulouse III – Institut de Mathématiques & chair TRIAL

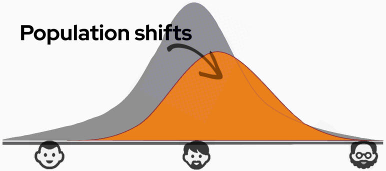
ANITI Days '24

Motivation • Trust in artificial intelligence

- **Pressing issues** from **public** + **academics** + **industry**

Failures





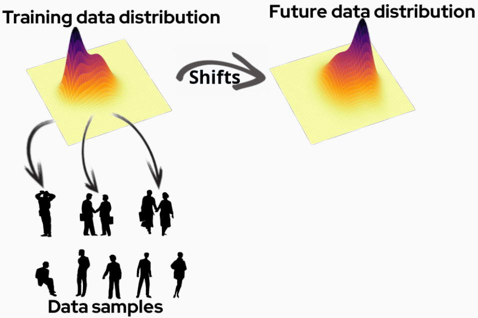
- **Mathematical modeling** of trustworthiness...
  - What **lifecycle changes** can one be **resilient** to?
  - How to **evaluate expected performance**?

Interplay between **robust stochastic optimization** and **machine learning**

- ... fitting the **operational constraints** of AI
  - **Provably robust** but **not too pessimistic**
  - Efficient **open-source implementation**

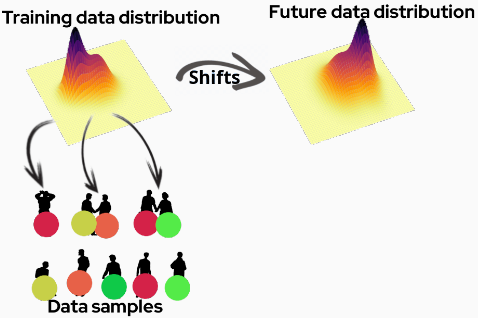
Bridge the **theoretical vision** of reliability with the public's **practical expectations**

- **Training** a model for **reliable performance**
  - Can only use collected **samples** from some unknown **training distribution**
  - Target a **low error** (eg. squared, logistic loss) on **average** for **future data**



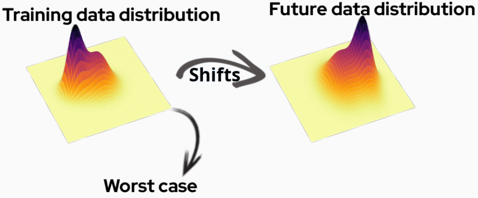
- **Training** a model for **reliable performance**
  - Can only use collected **samples** from some unknown **training distribution**
  - Target a **low error** (eg. squared, logistic loss) on **average** for **future data**

- **Classical** quantities optimized
  - **Empirical error over the data** → overly optimistic about future + replicates biases



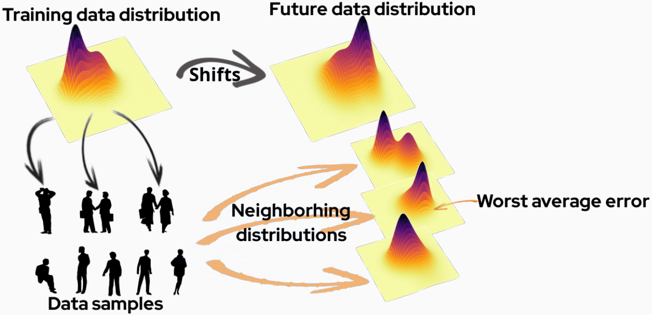
- **Training** a model for **reliable performance**
  - Can only use collected **samples** from some unknown **training distribution**
  - Target a **low error** (eg. squared, logistic loss) on **average** for **future data**

- **Classical** quantities optimized
  - **Worst error possible** → pessimistic + data agnostic



- **Training** a model for **reliable performance**
  - Can only use collected **samples** from some unknown **training distribution**
  - Target a **low error** (eg. squared, logistic loss) on **average** for **future data**

- A **sweet spot** in between: **Distributional robustness**
  - Infer a **distribution** with a **similar performance** as future → data-driven + trustworthy



Optimize
model parameters

Average error of parametrized model
on the robust distribution of data points

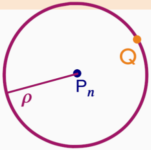
$$\min_{\theta} \quad \sup_{\mathsf{Q}} \quad \mathbb{E}_{\xi \sim \mathsf{Q}} \Big[L_{\theta}(\xi) \Big]$$

$$\text{s.t. } W(\mathsf{P}_n, \mathsf{Q}) \leq \rho$$

Eg. $L_\theta(\xi = (x, y)) = (\langle \theta, x \rangle - y)^2$

Worst distribution
within some Wasserstein distance
of the empirical distribution of the data

$$\mathsf{P}_n := \frac{1}{n} \sum_{i=1}^{n} \delta_{\xi_i}$$



- The perfect tool for **statistical reliability**
  - **Generalization** of the performance to unseen samples
  - **Resilience to shifts** between training and future data
  - **Future performance** is controlled

- **Limitations of classical WDRO**
  - **Numerically out-of-reach** in most situations **Eg.** Linear regression ✔ Neural nets ✖
  - **Modeling gaps** with reality **Eg.** High dimension ✖ Text, images ✖

- We proposed a **differentiable approximation** of Wasserstein distributional robustness
  - Built on a **entropic regularization** of the WDRO problem

    W. Azizian, F. Iutzeler, J. Malick : **Regularization for Wasserstein Distributionally Robust Optimization**, ESAIM: Control, Optimisation, and Calculus of Variations, 2023. ArXiv 2205.08826

  - Benefiting from **generalization** and **shift-resilience** guarantees

    W. Azizian, F. Iutzeler, J. Malick : **Exact Generalization Guarantees for (Regularized) Wasserstein Distributionally Robust Models**, NeurIPS 2023. ArXiv 2305.17076

  - Implemented in a **Python library** with both **sklearn** estimators and **torch** wrappers

    github.com/iutzeler/skwdro + pip / conda

    F. Vincent, W. Azizian, F. Iutzeler, J. Malick : **skwdro: a library for Wasserstein distributionally robust machine learning**, preprint, 2024. ArXiv 2410.21231