Tanguy Lefort

Montpellier, France
tanguy.lefort@umontpellier.fr
GitHub: tanglef
Website: https://tanglef.github.io

Ph.D student in machine learning: expertise in deep-learning and crowdsourcing

Experience

Ph.D Student

IMAG, University of Montpellier and INRIA Montpellier, France *10/2021 – Present*

- **Supervisors:** Benjamin Charlier (CNRS), Alexis Joly (INRIA) and, Joseph Salmon (CNRS)
- **Research Focus:** Noisy labels in crowdsourced classification datasets with expert feedback
- Identify data ambiguity in cooperative annoation datasets using neural networks margins
- Create a standardized framework in a Python library: `peerannot`
- Compare crowdsourcing aggregation strategies in a high class and large number of workers real dataset with Pl@ntNet

Intern

IMAG, Montpellier, France *03/2021 – 08/2021*

- Master's thesis internship on High dimensional optimization for penalized linear models with interactions
- **Supervisors:** Benjamin Charlier and Joseph Salmon
- Benchmarked descent methods for linear models with L1 and L2 penalties with first-order interactions
- Applied strategies for GPU acceleration and applied them to genomics datasets

IMAG, Montpellier, France *07/2020 – 08/2020*

- Contributed to the PyKeOps library under the supervision of Benjamin Charlier
- Rewrote Scipy's Fortran conjugate gradient routine for symbolic matrices in PyKeOps
- Benchmarked Ridge-Tikhonov regularization and worked on optimal transport problems

Education

Master Biostatistics, Montpellier, France *2019–2021*

- Theoretical and applied statistics and probabilities
- Classification algorithms, survival analysis, and modelization of populations

Bachelor in mathematics, Dijon, France *2016–2019*

- Bachelor with honors, specialized in applied mathematics
- Final project on skeletonization algorithm for gamma-ray surgery

Publications

A full list of research publications with abstracts is available at `https://tanglef.github.io/research/`

Journal

- Under review: *Peerannot: classification for crowdsourced image datasets with Python* in Computo by T. Lefort, B. Charlier, A. Joly and J. Salmon *2023*
- Under review: *Identify ambiguous tasks combining crowdsourced labels by weighting Areas Under the Margin* in **TMLR** by T. Lefort, B. Charlier, A. Joly and J. Salmon *2023*

In Proceedings of Conferences

- *Weighting areas under the margin in crowdsourced datasets* in Journées des statistiques de France by T. Lefort, B. Charlier, A. Joly and J. Salmon *2023*
- *Crowdsourcing label noise simulation on image classification tasks* in Journées des statistiques de France by T. Lefort, B. Charlier, A. Joly and J. Salmon *2022*
- *Benchopt: Reproducible, efficient and collaborative optimization benchmarks* in **NeurIPS 2022** by T. Moreau, M. Massias et. al *2022*

Other involvements

Open source library contributions

- Lead developer of the Peerannot library for handling crowdsourced datasets in image classification
- Developer on the BenchOpt library for reproducible benchmarks in optimization problems

Community services

- Reviewer for Computo journal *2023 – Current*
- Co-organizer of the Ph.D seminar at IMAG, University of Montpellier *2022–2023*

Teaching

- Advisor of 4 master students in the creation of a data visualization workshop on genomics data at the Montpellier Omics Days conference *2023*
- TA for Convex Optimization to undergraduate mathematics students *2021–2023*
- TA for a first-year biology course covering mathematical concepts *2021-2023*
- TA for mathematical undergraduates covering logic and proof techniques *2023*
- TA for second year undergraduates in chemistry: linear algebra, derivation and integration *2023*

Talks

Here is a list of some talks given. More details and sources are available on my personal website.

- *ChatGPT & co, Myths and Reality. Everything you wanted to ask about Deep Learning but did not dare to ask*, General knowledge Seminar with Francois David Collin, IMAG. *10/2023*
- *Data collection from a crowd: where is the noise coming from?* at Ph.D students seminar, IMAG. *09/2023*
- *Crowdsourcing label noise simulation on image classification tasks* at Journées des Statistiques de France (JDS) Univ. Lyon. *06/2022*
- *High dimensional optimization for penalized linear models with interactions using graphics card computational power* at Probability and Statistics (EPS) team seminar *11/2021*
- *Introduction to neural networks* with Joseph Salmon at ML-MTP seminar *10/2021*
- *Paper club Ridge Regularization: an Essential Concept in Data Science* by Trevor Hastie with Florent Bascou at ML-MTP seminar *04/2021*

Skills

- **Tools and Languages:** Python, R,Git, LaTeX, JavaScript, HTML, CSS
- **Machine learning algorithms:** logistic regression, trees, SVM, dimension reduction methods, KNN, neural networks (from CNN to ViT and more)
- **Communication:** Create a library, Interactive applications using Shiny/Dash/Flask/quarto
- **Languages:** French, English, Spanish, Italian (in decreasing order)