informatics mathematics

INRIA Saclay, https://project.inria.fr/dirtydata

LIP6

An optimal transport view on neural implementations of GAMs

Research theme: Machine learning, data science
Keywords: Neural networks, optimal transport, deep learning, generalized additive models, dirty data
Duration & salary: 3 to 6 months, between 500 €and 800 € monthly
Research teams: Parietal (INRIA Saclay) and LIP6 (Sorbonne Université)
Adviser: Gaël Varoquaux & Edouard Oyallon
Contact: gael.varoquaux@inria.fr, edouard.oyallon@lip6.fr
Application: Interested candidate should send CV and motivation letter

Context: Many data science problems, for instance in health or business, apply machine learning not on signals, but rather on tabular data x with features of different nature: age, sex, income... In these settings, a challenge is that the feature distribution μ has very different marginal distributions μ_i (the income, for instance, has a very long tail). In practice, the most popular models on these data build upon trees for which the learning procedure only depends on the order of the values of each feature, which makes the learning algorithm $f = \mathcal{A}(\mu)$ robust to the marginal distributions. Here, this property implies that a monotonic transformation φ_i applied to the i-th marginal distribution of μ should leave invariant $\mathcal{A}$, meaning that $\mathcal{A}(\mu \circ \varphi_i^{-1}) \circ \varphi_i = \mathcal{A}(\mu)$.
Recent developments in deep learning [2], however give hope to build neural architectures with the same properties: invariance to non-informative variabilities are believed to be a core component of modern deep architectures. However, the invariance that we seek here depart from traditional invariances of deep architectures which are typically obtained with regards to groups [4].
Generalized Additive Models (GAMs) are another traditional approach to model combinations of features with different marginal distributions [3]. They model the response by the sum of univariate functions with respect to each feature, *ie*, $f(x) = \sum_{j=1...p} f_j(\mathbf{x_j})$. It has recently been shown that using empirical risk minimization with, as a regularizer, a total-variation constraint on the functions f_j, the GAM learning procedure is invariant to monotonic transforms of each of the feature [5]. Note that this type of property is also naturally exhibited by the quantiles of a given distribution.
Parallel developments have parametrized the f_j of GAMs using small neural networks, with stochastic regularizations not related to total variation [1]. The total GAM then composes a specific kind of neural architecture that can be learned by stochastic gradient descent.

Proposed work: Our hypothesis is that a neural network architecture can be formulated following the ideas above to *transport* the marginal distribution of each feature to a Gaussian, using the resulting intermediate representations to build a multivariate predictor with fully-connected layer. Neural networks are indeed an ideal candidate to progressively Gaussianize the data because they are well known to increase the linear separability. Such formulation should be beneficial for learning on data with diverse marginal distributions, that will be more robust to usual preprocessing used in the data science community. The internship will entail *1)* a theoretical study of the desirable properties and how they can appear as a dedicated neural architecture with adequate regularization *2*) an empirical study with supervised learning on various real-life databases.

Required skills:

- Knowledge of machine learning or applied maths background (mathematical optimization and statistics)
- Some familiarity with fitting deep neural networks (typically pytorch)

[1] R Agarwal, N Frosst, X Zhang, R Caruana, and G E Hinton. Neural additive models: Interpretable machine learning with neural nets. *arXiv:2004.13912*, 2020.

[2] A Bietti and J Mairal. Group invariance, stability to deformations, and complexity of deep convolutional representations. *The Journal of Machine Learning Research*, 20(1):876–924, 2019.

[3] T Hastie and R Tibshirani. Generalized additive models: some applications. *Journal of the American Statistical Association*, 82(398):371, 1987.

[4] S Mallat. Understanding deep convolutional networks. *Philosophical Transactions of the Royal Society A*, 374(2065):20150203, 2016.

[5] S Matsushima. Statistical learnability of generalized additive models based on total variation regularization. *arXiv:1802.03001*, 2018.