Density estimation via normalizing flows

Eduardo García-Portugués • Department of Statistics, UC3M

Normalizing flows are a class of probabilistic models that transform a simple probability distribution (usually a known one, like a Gaussian) into a more complex and potentially high-dimensional distribution. They achieve this transformation through a composition of invertible and differentiable mappings that result in a density that closely matches the density of the data. These transformations, known as "flows", are automatically learned using different approaches, often involving neural networks. The final object represents a nonparametric density estimator of the high-dimensional density that is constructed with a very different spirit to the standard kernel density estimator and the Gaussian mixture model. This density estimate is useful for sampling new data objects, outlier hunting, clustering, goodness-of-fit and hypothesis testing, etc. The recent survey by Papamakarios et al. (2021) collects many references to the origin of the technique (see Huang et al. (2018)) and the recent advances on the field, such as the extension of normalizing flows that allows estimating densities within manifolds (Horvat & Pfister, 2023).

This thesis proposal is focused on understanding normalizing flows for density estimation in detail, both theoretically and empirically. This includes: (1) understanding and explaining precisely the mathematics of normalizing flows and of their main variations; (2) implement pedagogical constructions of normalizing flows to explore their performance in original toy examples; (3) compare the performance of available implementations of normalizing flows in a workbench of density estimation problems, determining the weaknesses and strengths of normalizing flows with respect to kernel smoothing and Gaussian mixture models; (4) apply normalizing flows for nonparametric outlier hunting, clustering or hypothesis testing, either in simulated or real datasets.

References

Horvat, C., & Pfister, J.-P. (2023). Density estimation on low-dimensional manifolds: An inflation-deflation approach. *Journal of Machine Learning Research*, *24*(61), 1–37. http://jmlr.org/papers/v24/21-0235.html

Huang, C.-W., Krueger, D., Lacoste, A., & Courville, A. (2018). Neural autoregressive flows. *Proceedings of the 35th International Conference on Machine Learning*, 2078–2087. https://proceedings.mlr.press/v80/huang18d.html

Papamakarios, G., Nalisnick, E., Rezende, D. J., Mohamed, S., & Lakshminarayanan, B. (2021). Normalizing flows for probabilistic modeling and inference. *Journal of Machine Learning Research*, *22*(57), 1–64. http://jmlr.org/papers/v22/19-1028.html