Neural networks for density learning

Eduardo García-Portugués • Department of Statistics, UC3M

Density estimation is a fundamental problem in statistics. It has applications to a wide range of topics in multivariate analysis, such as visualization, simulation, outlier hunting, classification and dimension-reduction. Flexible estimates of densities often rely on nonparametric approaches, the most popular being Kernel Density Estimation (KDE). As many other methods, KDE suffers from the curse of dimensionality, and estimating a density function f in $\mathbb{R}^p$ becomes challenging as p increases. KDE also depends crucially on the selection of a smoothing bandwidth matrix $\mathbf{H}$. Chapters 2–3 in García-Portugués (2025) give a review of KDEs and their applications, with links to R codes. See also Chacón & Duong (2018) for further mathematical details.

While typically used for regression or classification tasks, Neural Networks (NNs) can also be employed for density estimation, with some caveats. In generative modeling, convolutional NNs like U-Nets (Ronneberger et al., 2015) have been employed to model the *score function* $\mathbf{x} \mapsto s(\mathbf{x}) := \nabla_{\mathbf{x}} \log f(\mathbf{x})$ of a density f (Hyvärinen, 2005). This approach removes the need to account for the density constraints (integration to one and nonnegativity) at the price of learning s instead of f. However, although the density f is not explicitly known, *sampling* from the score function alone can be achieved through Langevin dynamics. See, e.g., Song et al. (2019) and Song (2021) for references.

The purpose of this thesis is to explore the use of NNs to learn a density through its score function, and to compare this alternative with KDEs. Unlike the KDE-estimator $\hat{f}_{\mathbf{H}}$, the NN-estimator $\hat{f}_{\mathrm{NN}}$ is not explicit and is only induced by the NN-estimated score function $\hat{s}_{\mathrm{NN}}$. However, sampling is possible from $\hat{f}_{\mathrm{NN}}$. In the thesis, the quality of the samples from $\hat{f}_{\mathbf{H}}$ and $\hat{s}_{\mathrm{NN}}$ will be benchmarked against those from f to investigate the performance of both estimation approaches across different dimensions and scenarios. The recovery of $\hat{f}_{\mathrm{NN}}$ from $\hat{s}_{\mathrm{NN}}$ will be also investigated through numerical approaches to provide a more complete comparison with $\hat{f}_{\mathbf{H}}$.

References

Chacón, J. E., & Duong, T. (2018). *Multivariate kernel smoothing and its applications*. CRC Press. https://doi.org/10.1201/9780429485572

García-Portugués, E. (2025). *Notes for nonparametric statistics*. https://bookdown.org/egarpor/NP-UC3M/

Hyvärinen, A. (2005). Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research*, *6*(Apr), 695–709. http://jmlr.org/papers/v6/hyvarinen05a.html

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In N. Navab, J. Hornegger, W. Wells, & A. Frangi (Eds.), *Medical image computing and computer-assisted intervention – MICCAI 2015* (Vol. 9351, pp. 234–241). Springer. https://doi.org/10.1007/978-3-319-24574-4_28

Song, Y. (2021). *Generative modeling by estimating gradients of the data distribution*. https://yang-song.net/blog/2021/score/.

Song, Y., Garg, S., Shi, J., & Ermon, S. (2019). Sliced score matching: A scalable approach to density and score estimation. *Conference on Uncertainty in Artificial Intelligence*. https://auai.org/uai2019/proceedings/papers/204.pdf