Deep Density Destructors

Hunar Preet Singh (160301)[1], Ritesh Kumar (160575)[2], Yogesh Kumar (160831)[3]

1 Motivation

The idea of density destructors is an unconventional way to approach the problem of estimating complex distributions via shallow/deep networks given samples from the true distribution. The main motivation behind choosing this project is that we wish to explore this approach and determine how does this approach fare as compared to the conventional constructive approaches. Since the current framework that we plan to work on is a very generic framework it allows all the previous methods to be systematically combined, evaluated and improved upon. So, there is a huge scope of experimentation and exactly figure out the areas where this approach is extremely useful or what can be potential problems in this area.

2 Problem Statement

A density destructor is an invertible function that transforms a given density to the uniform density - destroying any structure in the original density [6]. To view the estimation of a complex distribution via deep networks given samples from the true distribution we have two perspectives - Constructive and Destructive. Examples for constructive transformation models are VAEs [7] and GANs [5]. Formally this amounts to the following approximation:

$$\begin{aligned} \mathbf{z} &\sim \text{BaseDistribution} \\ \mathbf{x} \equiv G_\phi(\mathbf{z}) &\sim \text{GenerativeDistribution} \end{aligned}$$

As a mirror of the constructive process, estimating an invertible destructive transformation, or destructive flow, from the the input distribution to the base distribution:

$$\begin{aligned} \mathbf{x} &\sim \text{DataDistribution} \\ \mathbf{z} \equiv D_\theta(\mathbf{x}) &\sim \text{ApproxmiateBaseDistribution} \end{aligned}$$

Our task is to estimate the destructive transformation $D_\theta(\mathbf{x})$. Moreover, we wish to understand the paper implementation, compare it with other state-of-the-art models and also improve the model presented.

3 Prior Work

The idea of density destructors is not a very recent breakthrough. It has been slowly developed and with increasing complexity and efficiency. The core idea can be rooted back to seminal exploratory projection pursuit paper [3] which used various multivariate transforms to make the resulting distribution more and more Gaussian. Then by sampling from a Gaussian and inverting back to get a sample from original density. [1] was based on similar lines of Gaussianization. [10] proposed many specific types of invertible tranformations to move the density towards Gaussianity comprising of linear and radial flows. Recently, there has been work on parameterizing a valid autoregressive distribution that uses neural network models to achieve the task [4, 2, 8]. These models are a hybrid of neural network and linear tranformations. They compute a linear tranformation of inputs but scale and shift parameters are a function of deep nenural networks. One drawback of these recent models is that they are restricted to a specific type of invertible transformations. [6] generalizes the idea of the previous works on invertible transformations and establish a fundamental connection to shallow density models. The mentioned framework can leverage both deep and shallow density models and thus making the framework highly modular. This paper also formally defines density destructors along with its key properties, introduces an invertible transformation for tree densities [9] and proposes a image-specific destructive transformation based on pixel locality to showcase the flexibility of the mentioned framework.

[1]hunar@iitk.ac.in
[2]riteshk@iitk.ac.in
[3]kumaryog@iitk.ac.in

4 Plan of work

Time Period	Proposed Work
Feb - Before Mid-Sem	We will go through the main paper by thoroughly understanding the concepts presented in it. We will finalize our problem for the project during this period. We will also present a **report** based on our problem statement and main topics of the paper few days before Mid-sem exams.
Post Mid-Sem - March	We will work on our research problem and also study the implementation of main paper. As we are aiming for find a novel approach towards the problem we would also implement our models and keep record of how it performs against the existing state-of-the-art models. We plan to meet the mentor minimum two times during this period to present a report on our work and this will also help us to get feed-back on our approach.
April - Before End-Sem	We will conclude our project by writing a report on our work and findings. And if we have found some novel approach towards the problem then we will also work on our paper.

References

[1] Scott Saobing Chen and Ramesh A Gopinath. "Gaussianization". In: *Advances in neural information processing systems*. 2001, pp. 423–429.

[2] Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. "Density estimation using Real NVP". In: *arXiv preprint arXiv:1605.08803* (2016).

[3] Jerome H Friedman. "Exploratory projection pursuit". In: *Journal of the American statistical association* 82.397 (1987), pp. 249–266.

[4] Mathieu Germain et al. "Made: Masked autoencoder for distribution estimation". In: *International Conference on Machine Learning*. 2015, pp. 881–889.

[5] Ian Goodfellow et al. "Generative adversarial nets". In: *Advances in neural information processing systems*. 2014, pp. 2672–2680.

[6] David Inouye and Pradeep Ravikumar. "Deep Density Destructors". In: *International Conference on Machine Learning*. 2018, pp. 2172–2180.

[7] Diederik P Kingma and Max Welling. "Auto-encoding variational bayes". In: *arXiv preprint arXiv:1312.6114* (2013).

[8] George Papamakarios, Theo Pavlakou, and Iain Murray. "Masked autoregressive flow for density estimation". In: *Advances in Neural Information Processing Systems*. 2017, pp. 2338–2347.

[9] Parikshit Ram and Alexander G Gray. "Density estimation trees". In: *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM. 2011, pp. 627–635.

[10] Esteban G Tabak, Eric Vanden-Eijnden, et al. "Density estimation by dual ascent of the log-likelihood". In: *Communications in Mathematical Sciences* 8.1 (2010), pp. 217–233.