ON THE BENEFITS OF DEFINING VICINAL DISTRIBUTIONS IN LATENT SPACE

Anonymous authors
Paper under double-blind review

ABSTRACT

Various empirical and theoretical studies have shown that minimizing Empirical Risk (Vapnik, 1998) over training datasets in over-parameterized settings leads to their memorization and thus poor generalization on out-of-distribution shifts. To mitigate this problem, Vicinal Risk Minimization (VRM) was proposed which essentially chooses to train networks on similar but different examples to the training data by defining a vicinity/neighbourhood around each training example. Once defined, more examples can be sampled from their vicinity to enlarge the support of training distribution.

One of the popular choices to create the vicinal distribution is Mixup (Zhang et al., 2017) which has emerged as a popular technique to train models for better generalisation in the last couple of years. Recent works have also shown that the idea of Mixup can be leveraged during inference (Pang* et al., 2020) and in many existing techniques like data augmentation (Hendrycks* et al., 2020), adversarial training (Lamb et al., 2019), etc. to improve the robustness of models to various input perturbations and corruptions. Other efforts on Mixup (Thulasidasan et al., 2019) have shown that Mixup-trained networks are also significantly better calibrated than regular ones.

Although still in its early phase, the above efforts (Zhang et al., 2017; Verma et al., 2019; Pang* et al., 2020; Thulasidasan et al., 2019) also indicate a trend to viewing Mixup from perspectives of robustness and calibration. In this work, we take another step in this direction and propose a new vicinal distribution/sampling technique called *VarMixup (Variational Mixup)* to sample better Mixup images during training to induce out-of-distribution robustness as well as improve predictive uncertainty of models. In particular, we hypothesize that the latent unfolded manifold underlying the data (through a generative model, a Variational Autoencoder in our case) is linear by construction and hence more suitable for the defining vicinal distributions involving linear interpolations, such as Mixup. Importantly, we show that this choice of the distribution for Mixup plays an important role towards robustness (by evalauting their robustness on CIFAR corrupted Hendrycks & Dietterich (2019) datasets) and predictive uncertainty (by measuring the Expected Calibration Error).

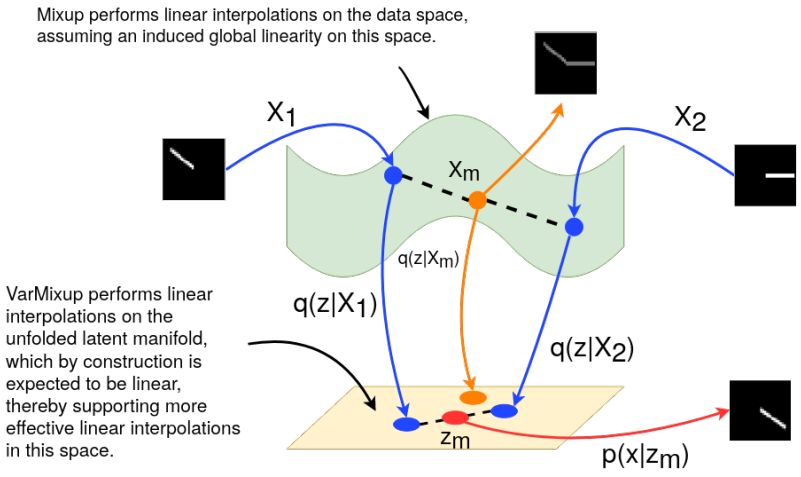

Figure 1: Illustration of conceptual idea behind VarMixup. We interpolate on the unfolded manifold, as defined by a generative model (VAE, in our case).

REFERENCES

Dan Hendrycks and Thomas Dietterich. Benchmarking neural network robustness to common corruptions and perturbations. *Proceedings of the International Conference on Learning Representations*, 2019.

Dan Hendrycks*, Norman Mu*, Ekin Dogus Cubuk, Barret Zoph, Justin Gilmer, and Balaji Lakshminarayanan. Augmix: A simple method to improve robustness and uncertainty under data shift. In *International Conference on Learning Representations*, 2020.

Alex Lamb, Vikas Verma, Juho Kannala, and Yoshua Bengio. Interpolated adversarial training: Achieving robust neural networks without sacrificing too much accuracy. AISec'19, 2019.

Tianyu Pang*, Kun Xu*, and Jun Zhu. Mixup inference: Better exploiting mixup to defend adversarial attacks. In *International Conference on Learning Representations*, 2020.

Sunil Thulasidasan, Gopinath Chennupati, Jeff A Bilmes, Tanmoy Bhattacharya, and Sarah Michalak. On mixup training: Improved calibration and predictive uncertainty for deep neural networks. In *Advances in Neural Information Processing Systems 32*, pp. 13888–13899. Curran Associates, Inc., 2019.

Vladimir N. Vapnik. *Statistical Learning Theory*. Wiley-Interscience, 1998.

Vikas Verma, Alex Lamb, Christopher Beckham, Amir Najafi, Ioannis Mitliagkas, David Lopez-Paz, and Yoshua Bengio. Manifold mixup: Better representations by interpolating hidden states. In *Proceedings of the 36th International Conference on Machine Learning*, pp. 6438–6447, 2019.

Hongyi Zhang, Moustapha Cissé, Yann N. Dauphin, and David Lopez-Paz. mixup: Beyond empirical risk minimization. *CoRR*, abs/1710.09412, 2017.