Ensembles Learn Equivariance Through Data Augmentation

UMEÅ UNIVERSITET

Oskar Nordenfors, Umeå University

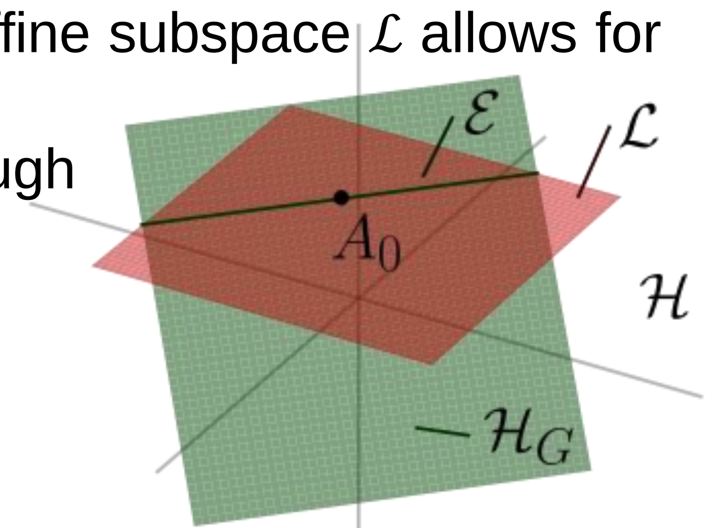
Department of Mathematics and Mathematical Statistics

Abstract

For many machine learning tasks, known symmetries of the task are a strong inductive bias which can be used to improve models. In the language of the representation theory of groups, models which respect symmetries are called equivariant. Recently, it was shown that equivariance emerges in ensembles of neural networks trained under data augmentation in the limit of infinite width.[1] In our paper, we show that this happens in the finite width setting as well. For general architectures, we provide a sufficient condition on the relationship between the architecture and the group action for our results to hold.

Overview

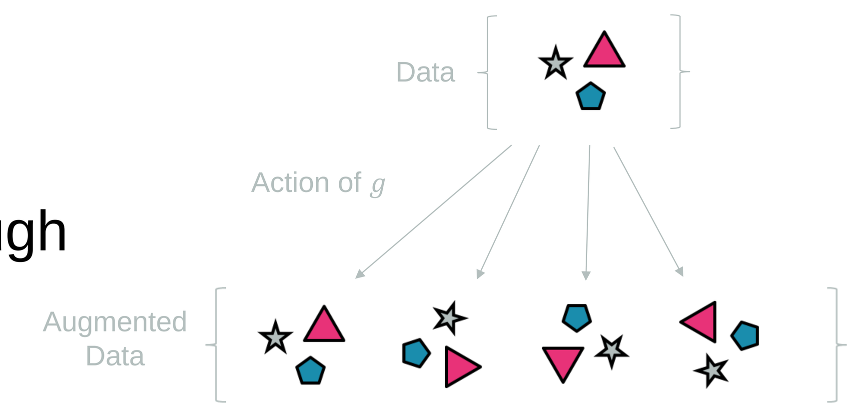
- Neural network, $\varphi_A : X \to Y$, defined recursively by
  $x_0 := x, \quad x_{i+1} := \sigma_i(A_i x_i), \quad \varphi_A(x) := x_L := y, \qquad i \in [\mathrm{L}],$
  where $x_i \in X_i$, $A_i \in \mathrm{Hom}(X_i, X_{i+1})$, and σ_i non-linearities.
- Restricting the linear layers to an affine subspace $\mathcal{L}$ allows for more general architectures.
- A group G is acting on X and Y through a representation ρ.
- ρ can be pulled through φ_A onto the parameters $A \in \mathcal{L}$.[3]
- Training with data augmentation means optimizing
  $R^{\mathrm{aug}}(A) := E_g E_{\mathcal{D}}[l(\varphi_A(\rho_X(g)x), \rho_Y(g)y)] = E_g[\mathrm{R}(\rho_{\mathcal{L}}(g)A)]$



- We optimize by the projected gradient flow
  $$A'(t) = -\pi_{\mathcal{L}} \nabla R^{\mathrm{aug}}(A(t)).$$
- An ensemble model for time t is defined by $\overline{\varphi_t}\,(x) = E_A\big[\varphi_{A(t)}(x)\big]$.
- We want equivariant models through data augmentation. I.e.,

$$\begin{tikzcd} X \arrow[r, "\varphi"] \arrow[d, "\rho_X(g)"'] & Y \arrow[d, "\rho_Y(g)"] \\ X \arrow[r, "\varphi"'] & Y \end{tikzcd}$$

through

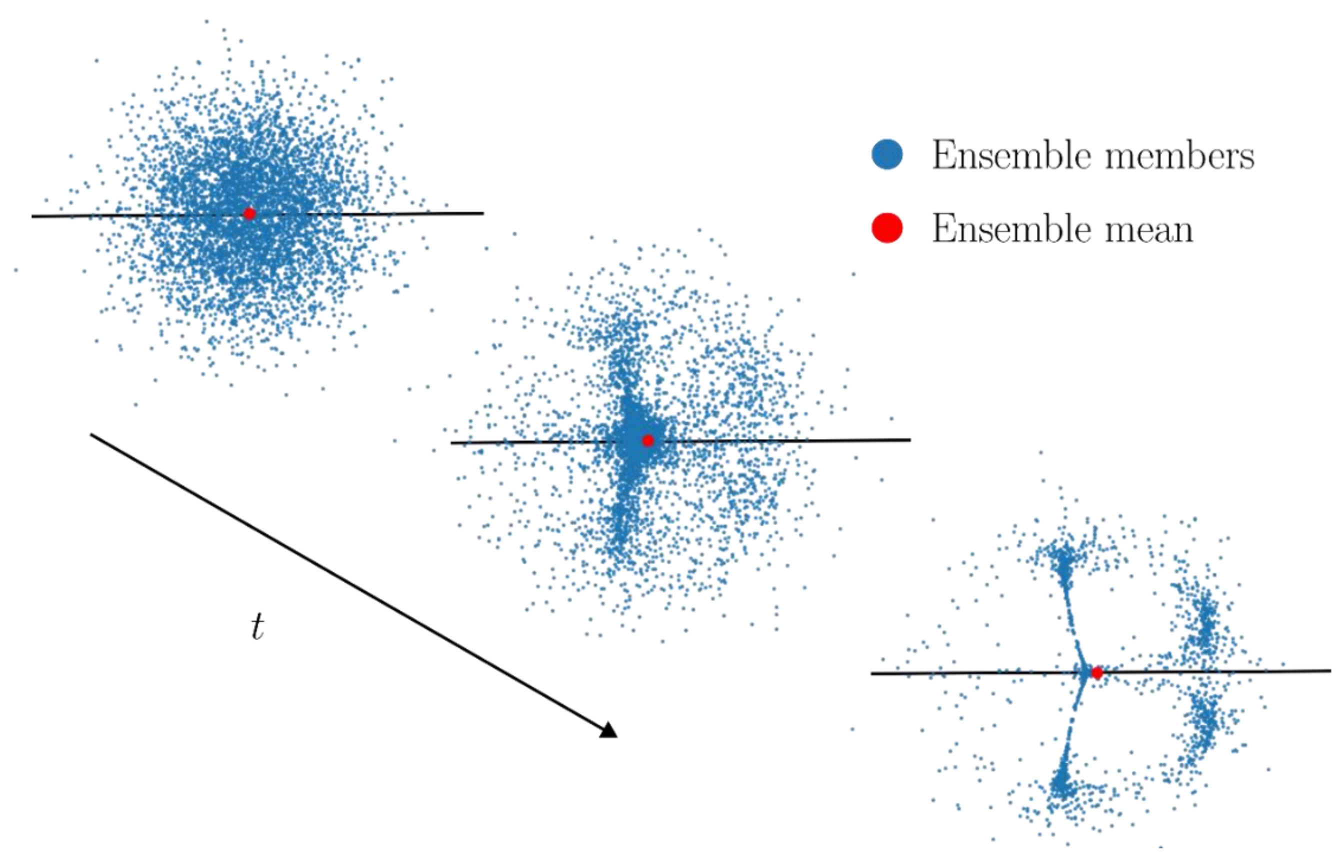


Data augmentation turns the risk into an invariant function, which yields an equivariant gradient flow. This flow will then push an invariant initial distribution of parameters forward to another invariant distribution at any later time.[2] It turns out that the expected value of our network with respect to the parameters is then an equivariant function. This is summarized on the right side of the poster, both visually and in a theorem.

References

1. Jan E. Gerken and Pan Kessel. Emergent Equivariance in Deep Ensembles. In *Proceedings of the 41st International Conference on Machine Learning* (ICML'24), volume 235, pp. 15438–15465, 2024.
2. Jonas Köhler, Leon Klein, and Frank Noé. Equivariant Flows: Exact Likelihood Generative Learning for Symmetric Densities. In *Proceedings of the 37th International Conference on Machine Learning* (ICML'20), volume 119, pp. 5361–5370, 2020.
3. Oskar Nordenfors, Fredrik Ohlsson and Axel Flinth. Optimization Dynamics of Equivariant and Augmented Neural Networks. Preprint at https://arxiv.org/abs/2303.13458.

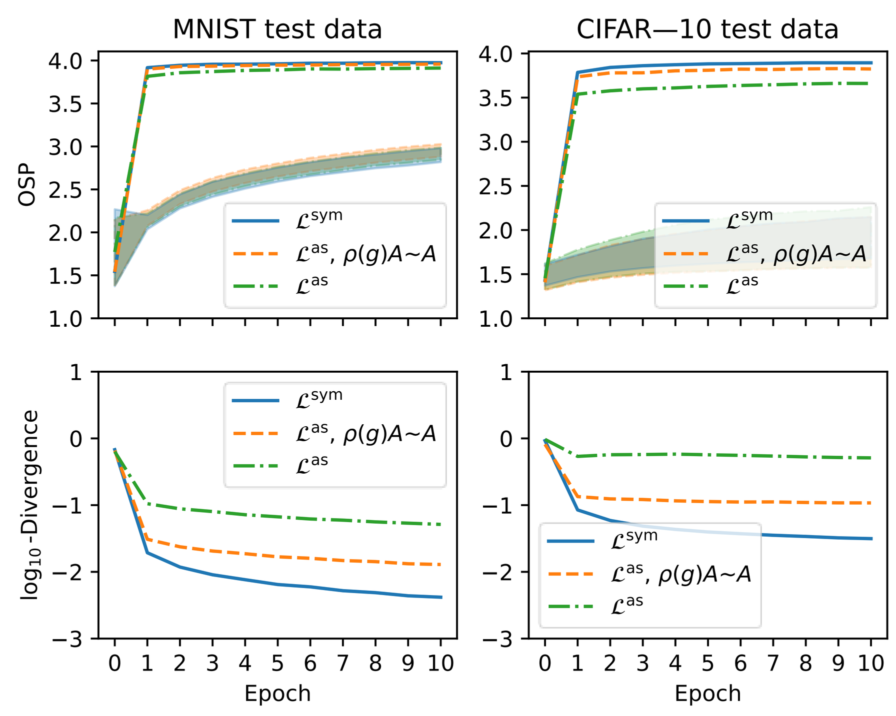
Findings



Main Result

If (i) the network parameters are invariantly distributed at initialization and (ii) the space $\mathcal{L}$ is invariant, then the ensemble model trained under data augmentation satisfies
$$\overline{\varphi_t}\,(\rho_X(g)x) = \rho_Y(g)\,\overline{\varphi_t}\,(x),$$
for every g, x and t.

We test our theory emprically by training both models that fulfill our theorem's assumptions and ones that violate them on an invariant MNIST classification task. The results can be viewed in the figure below.



- The ensembles which satisfy the conditions display more invariance.
- The difference is not so big however. All ensembles show much higher invariance than members.
- Could invariant distributions be attractors of the augmented strategy?

For more details, scan the QR code or click the link to read the preprint! Oskar Nordenfors and Axel Flinth. Ensembles Provably Learn Equivariance Through Data Augmentation. Preprint at https://arxiv.org/abs/2410.01452.