Are All Linear Regions Created Equal?

Matteo Gamba, Adrian Chmielewski-Anders, Josephine Sullivan, Hossein Azizpour, Mårten Björkman

KTH Royal Institute of Technology

Problem

ReLU networks express **continuous piece-wise affine** functions of their input $\mathbf{x}$. To **quantify their expressivity**, the number of linear "pieces", i.e. the **density of linear regions**, has been studied as a proxy for model **nonlinearity**. Intuitively, high nonlinear behaviour requires expressing many regions.

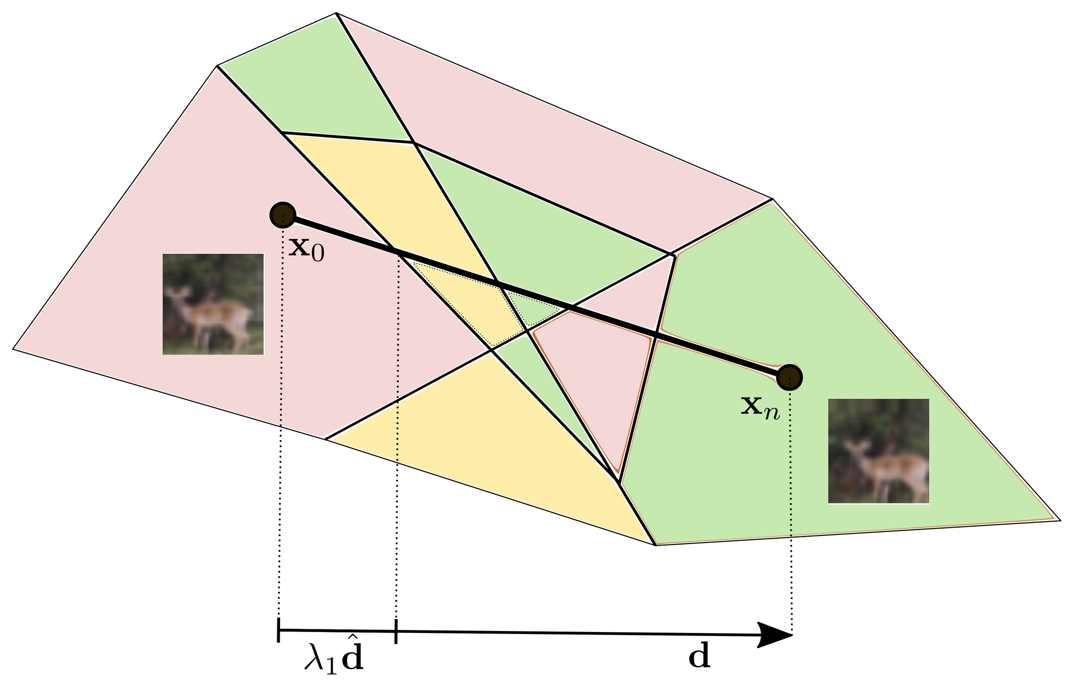
Research Question

Linear region density heavily assumes that all regions **meaningfully contribute nonlinearity** relevant to learning. Is it the case for **overparameterized** models?

Existing Methods

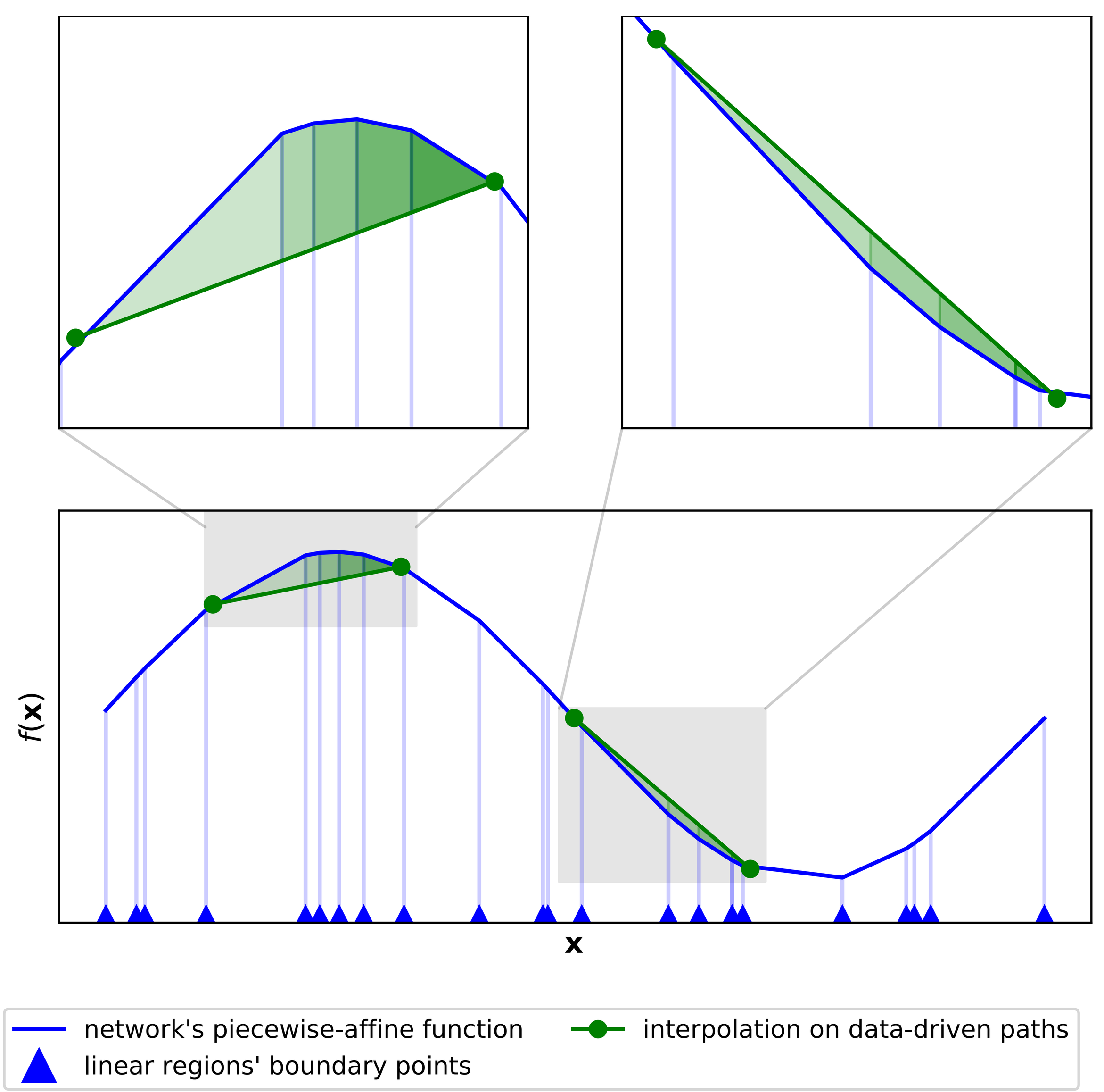
Prior empirical approaches have been limited to MLPs (Hanin & Rolnick, 2019; Novak et al., 2018) or small ConvNets (Zhang & Wu, 2020). Analytic methods don't scale to modern networks, while numerical ones assume *uniform* density in the input space.

Adaptive Linear Region Discovery

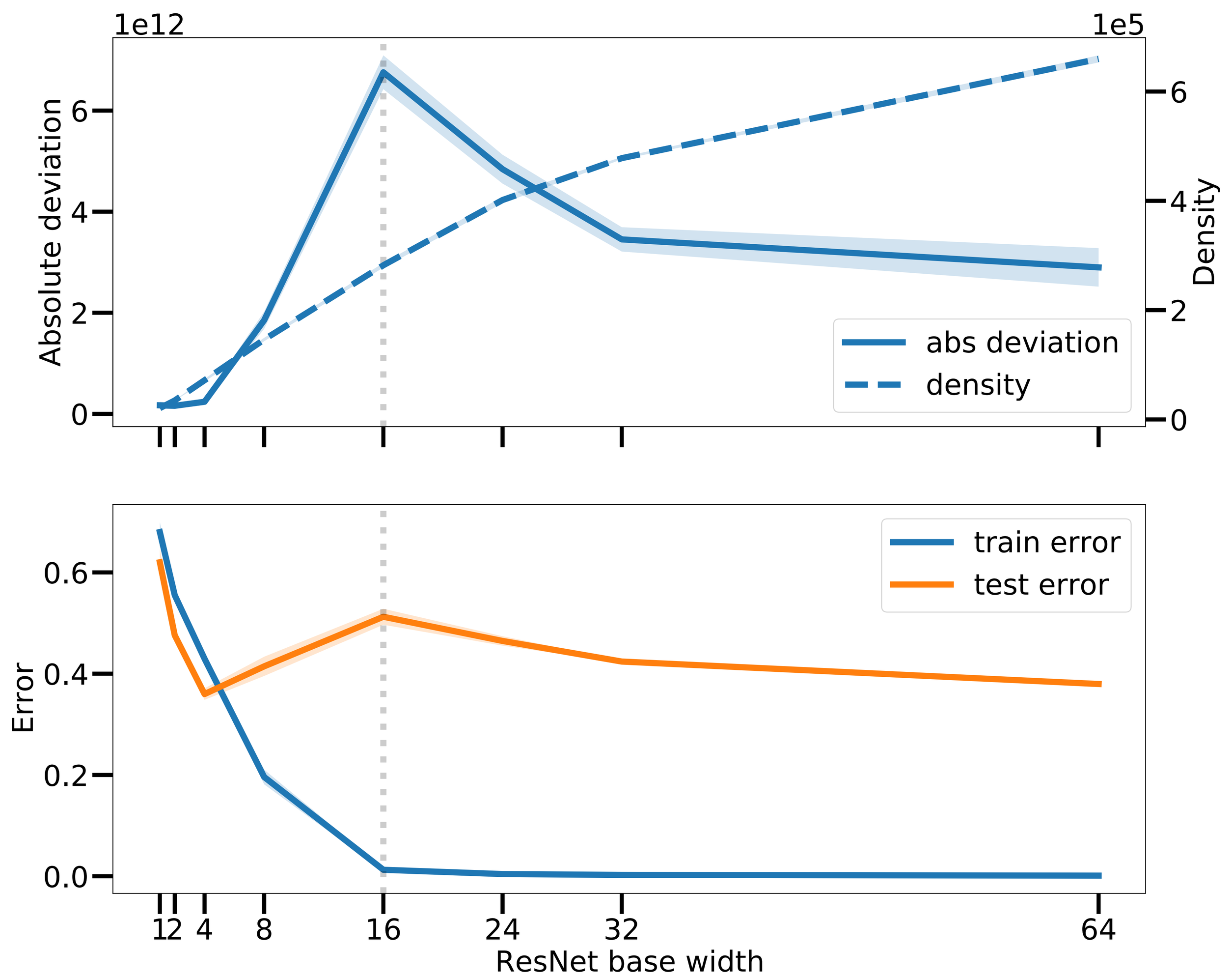


Our fast numerical method finds the smallest $\lambda > 0$ to cross the nearest region boundary, in directions tangent to the "data manifold", **with a single forward pass**. We use it to compute our novel measure of nonlinearity, *absolute deviation*.

Measuring Nonlinearity



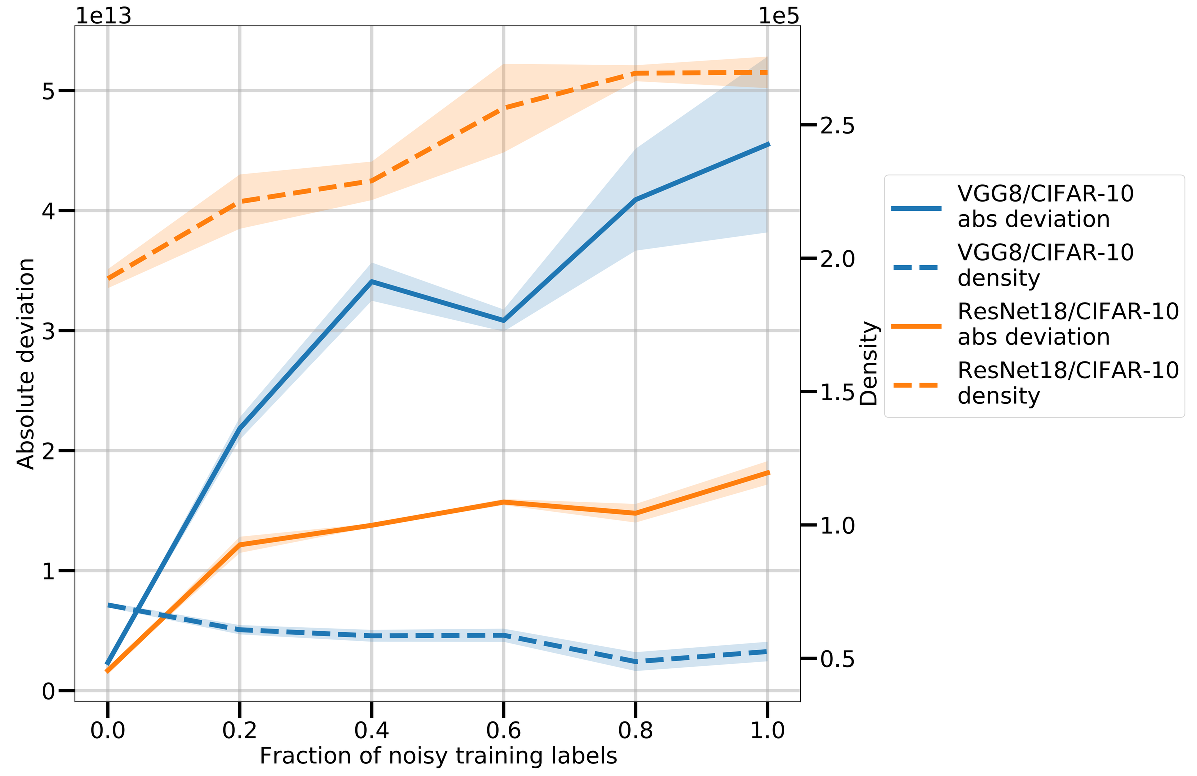
Overparameterization & Nonlinearity



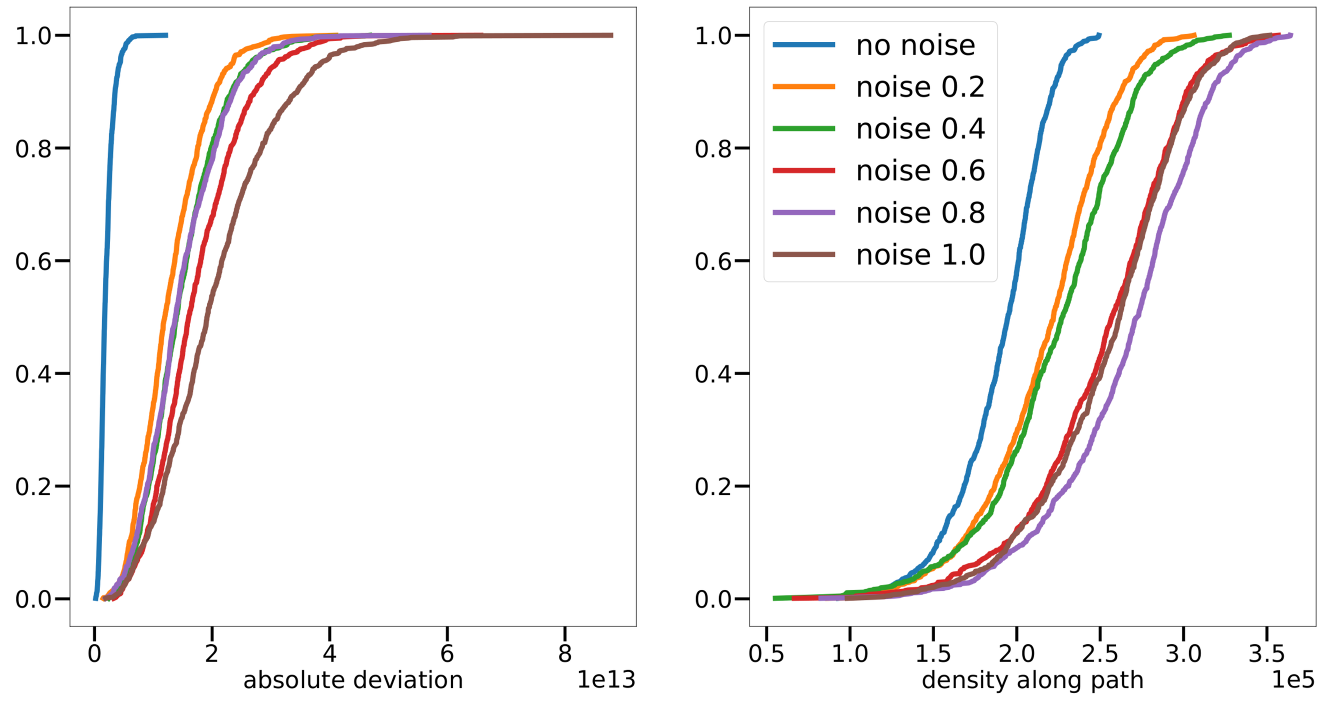
Implications

(1) **Density is** a **noisy** estimator of nonlinearity of large networks. (2) In **model-wise double descent** (Nakkiran et al., 2019), past the noise interpolation threshold, local linearity increases, connecting **reduced test error to reduced model nonlinearity**.

Ranking Nonlinearity



Capturing Effective Nonlinearity



Density & Variation

Table: Spearman rank correlation between density and absolute deviation.

	CIFAR-10		CIFAR-100
	VGG8	ResNet18	VGG8
vanilla	-0.15 ± 0.00	0.05 ± 0.03	0.00 ± 0.02
augment.	-0.08 ± 0.02	0.17 ± 0.06	0.11 ± 0.08
noise 0.2	0.04 ± 0.02	0.21 ± 0.02	
noise 0.4	0.11 ± 0.02	0.21 ± 0.04	
noise 0.6	0.18 ± 0.07	0.25 ± 0.03	
noise 0.8	0.26 ± 0.04	0.22 ± 0.02	
noise 1.0	0.14 ± 0.04	0.27 ± 0.04	

References

B. Hanin and D. Rolnick. Complexity of linear regions in deep networks. In *ICML*, pp. 2596–2604, 2019.

G. Montufar, R. Pascanu, K. Cho, and Y. Bengio. On the number of linear regions of deep neural networks. In *Advances in NeurIPS*, 2014.

P. Nakkiran, G. Kaplun, Y. Bansal, T. Yang, B. Barak, and I. Sutskever. Deep double descent: Where bigger models and more data hurt. In *ICLR*, 2019.

R. Novak, Y. Bahri, D.A. Abolafia, J. Pennington, and J. Sohl-Dickstein. Sensitivity and generalization in neural networks: an empirical study. In *ICLR*, 2018.

X. Zhang and D. Wu. Empirical studies on the properties of linear regions in deep neural networks. In *ICLR*, 2020.

Source code

https://bit.ly/linear-regions