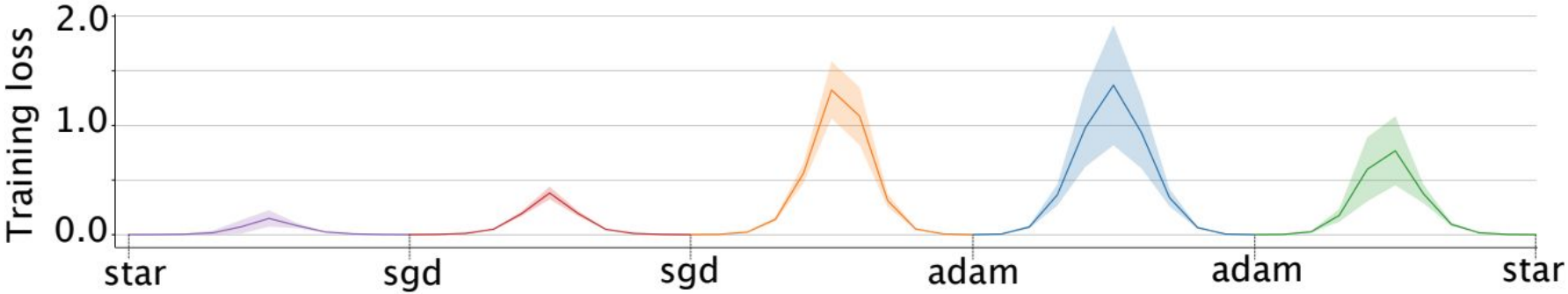
Do Deep Neural Network Solutions form a Star Domain?



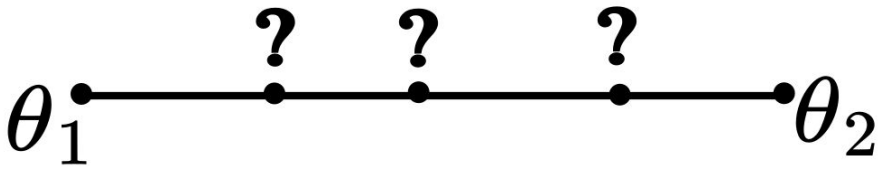
Ankit Sonthalia[1], Alexander Rubinstein[1], Ehsan Abbasnejad[2], Seong Joon Oh[1]

[1] Scalable Trustworthy AI at the Tübingen AI Center, University of Tübingen
[2] Australian Institute for Machine Learning

Different seeds lead to different minima

Dataset + architecture + hyperparameters → Model

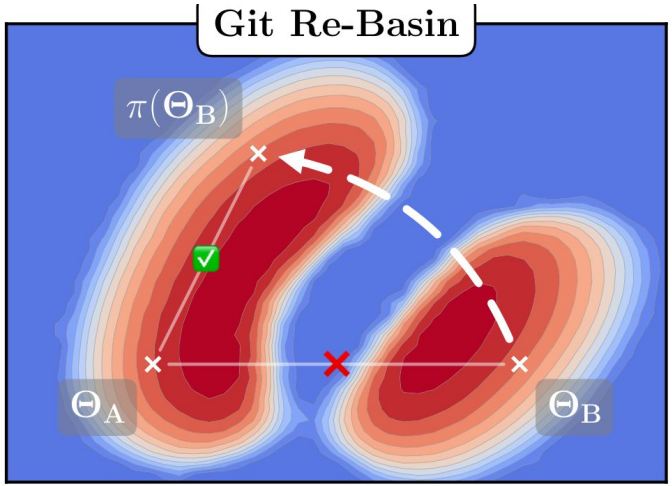
- θ_1, θ_2 are solutions.

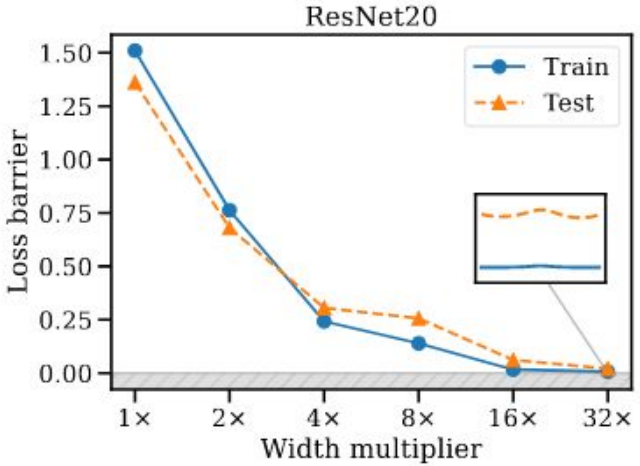


- What about 0.3 x θ_1 + 0.7 x θ_2?
- What about t x θ_1 + (1-t) x θ_2 ?

What do NN solution sets look like?

Prior work: Given *sufficient width*, all solutions are linearly connected, modulo permutations. [1, 2]





(Figures from [2])

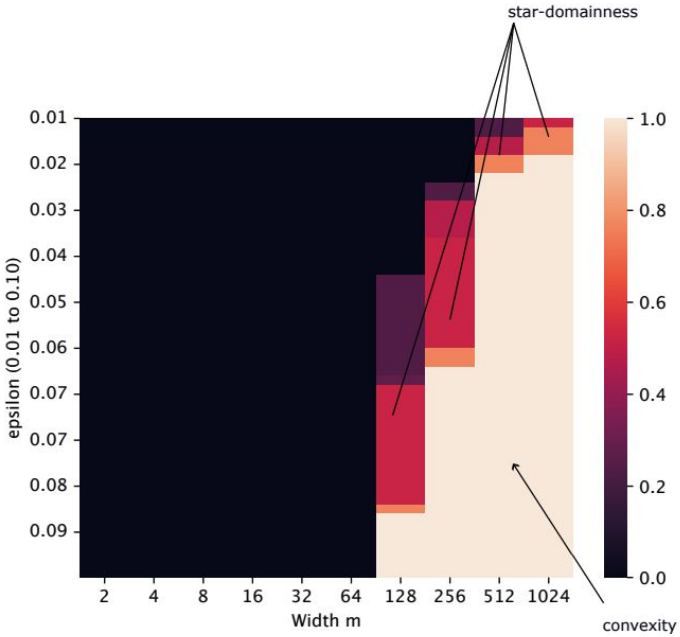
[1] Entezari, R., Sedghi, H., Saukh, O., & Neyshabur, B. The role of permutation invariance in linear mode connectivity of neural network. ICLR 2022.
[2] Ainsworth, S. K., Hayase, J., & Srinivasa, S. Git re-basin: Merging models modulo permutation symmetries. ICLR 2023.

Our claim: at *smaller* widths, star-domainness holds.

Claim: Width increases → star-domainness increases until solution set becomes convex.

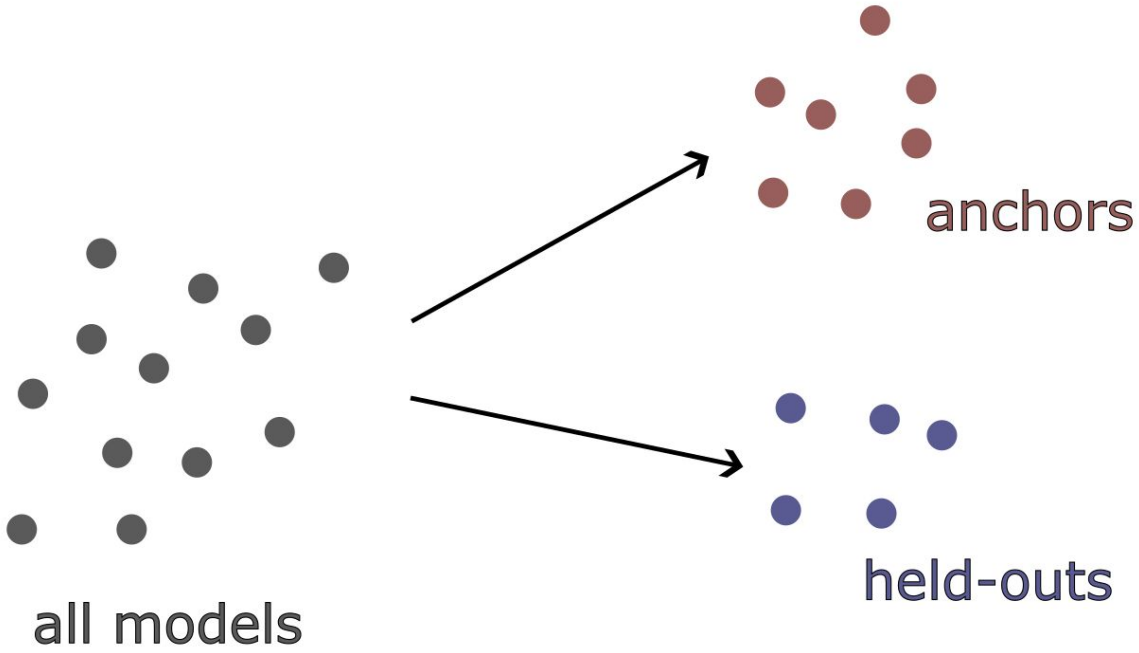
Toy experiment: 2-layer linear networks (minimal non-convex NN).

Main paper: real-world datasets.



The Starlight Algorithm: explicitly finding θ*

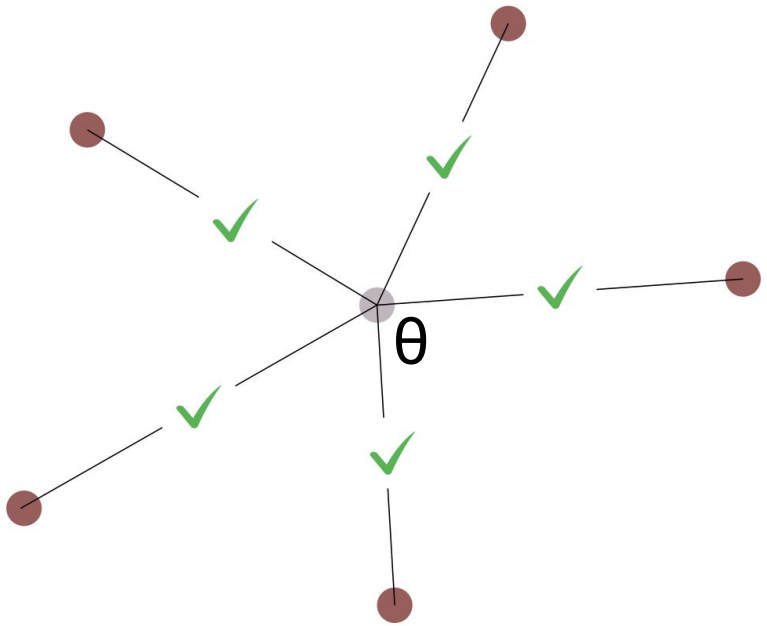
Train several solutions identically, except for the random seed.



How to find θ* ?

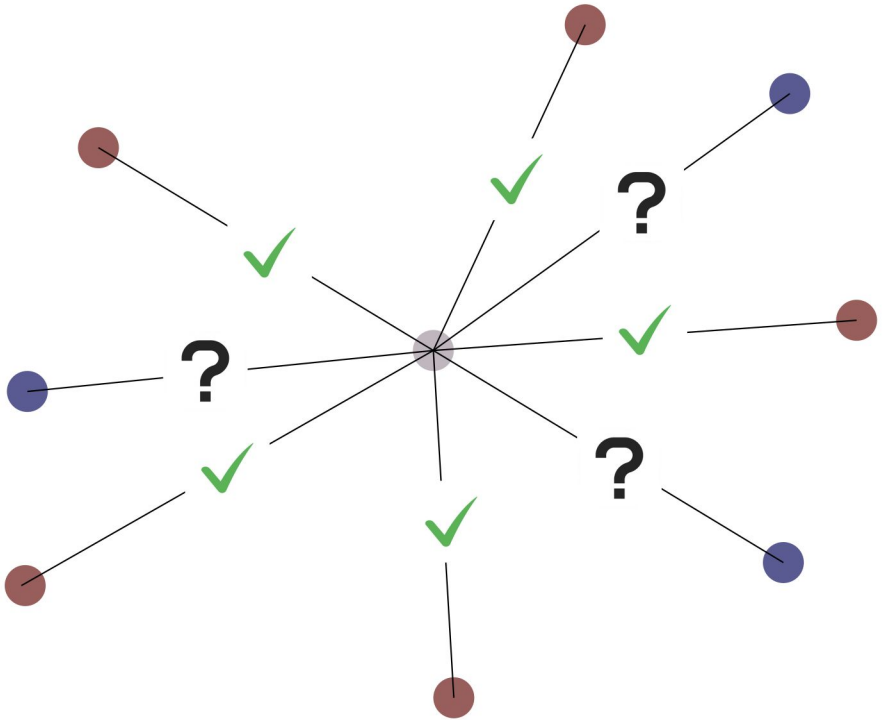
Use the anchors to train a “star model”, θ.

- Force θ to be connected to all anchors.
- Use Monte-Carlo optimization.

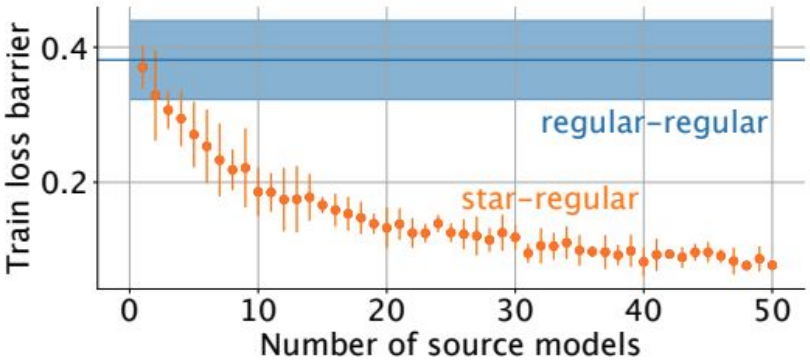


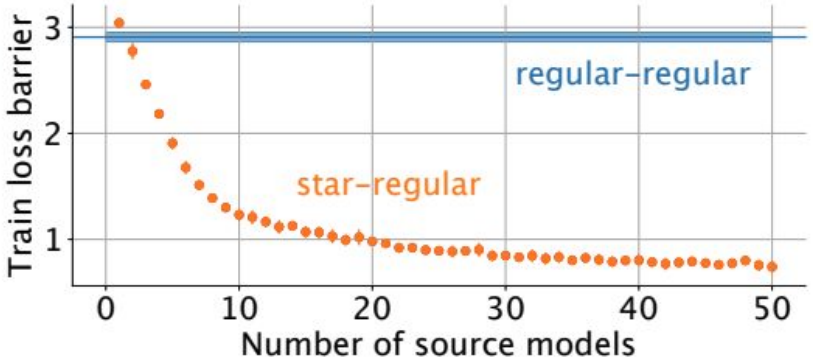
Is θ the θ* that we were seeking?

Use the held-outs to test this.

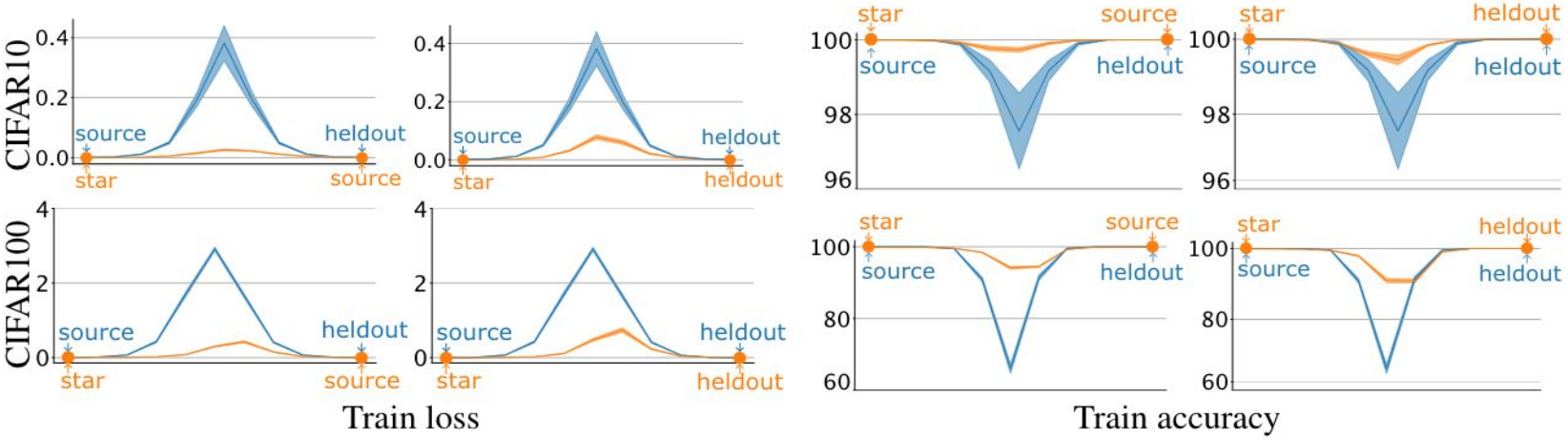


Result: Barriers with held-out models decrease #anchors increases.





Result: Barriers with held-out models decrease #anchors increases.



Code is available at https://github.com/aktsonthalia/starlight.