

Task complexity shapes internal representations and robustness in neural networks

Robert Jankowski, Filippo Radicchi, M. Ángeles Serrano, Marián Boguñá, Santo Fortunato

Network Science Informs AI (NSIA)

NetSci 2026 Satellite

Can trained networks encode task difficulty?

Given two neural networks, can we tell which one learned the harder task?

Existing approaches compare networks using [1,2]:

- Representational measures: CCA, CKA, RSA, ...
- Functional measures: accuracy, predictions, decision boundaries, ...

Previous work shows that task structure affects learned representations [4].

Hypothesis: Task complexity should also leave signatures in the learned weight structure

[1] Klabunde, M., Schumacher, T., Strohmaier, M., & Lemmerich, F. (2025). Similarity of neural network models: A survey of functional and representational measures. *ACM Computing Surveys*, 57(9), 1-52.

[2] Kornblith, S., Norouzi, M., Lee, H., & Hinton, G. (2019, May). Similarity of neural network representations revisited. *ICML* (pp. 3519-3529).

[3] Lampinen, A. K., Chan, S. C., & Hermann, K. (2024). Learned feature representations are biased by complexity, learning order, position, and more. *arXiv:2405.05847*.

[4] Mukherjee, K., & Rogers, T. T. How does task structure shape representations in deep neural networks?. In *NeurIPS 2020 Workshop SVRHM*.

Choosing easy and hard tasks

Structural Similarity Index (SSIM) -
a method used to assess the similarity
between two images.

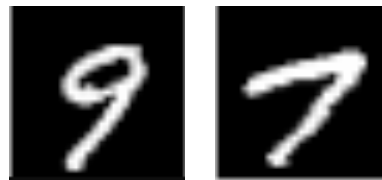
SSIM distance = $1 - \text{SSIM}$

- Lower SSIM distance \rightarrow harder task
- Higher SSIM distance \rightarrow easier task

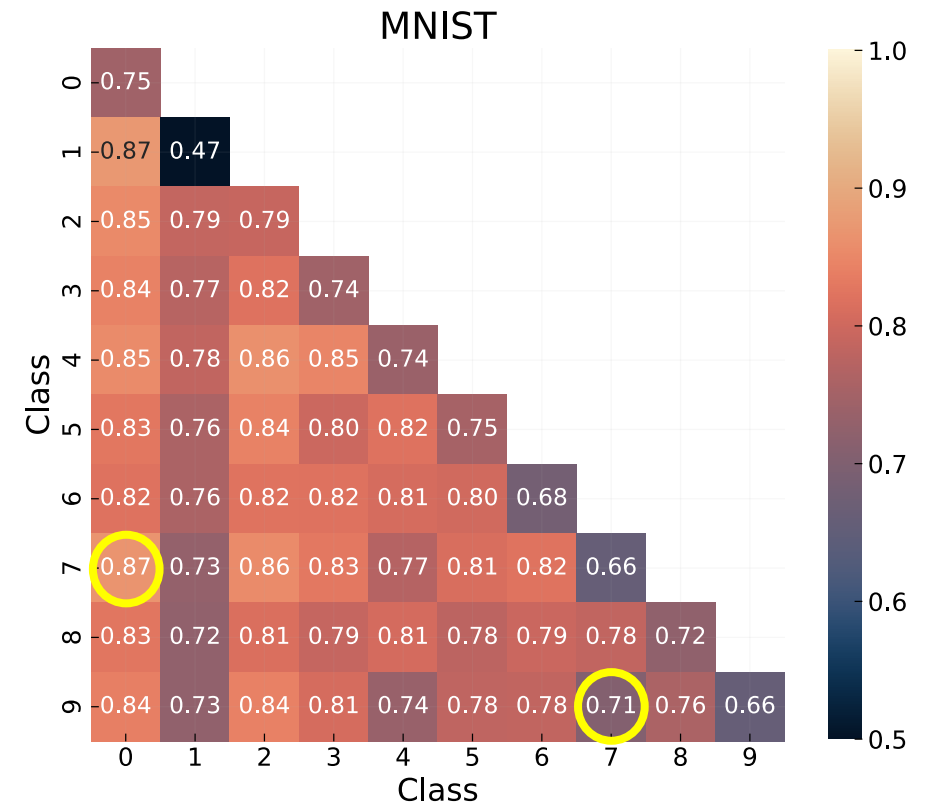
Easy task



Hard task

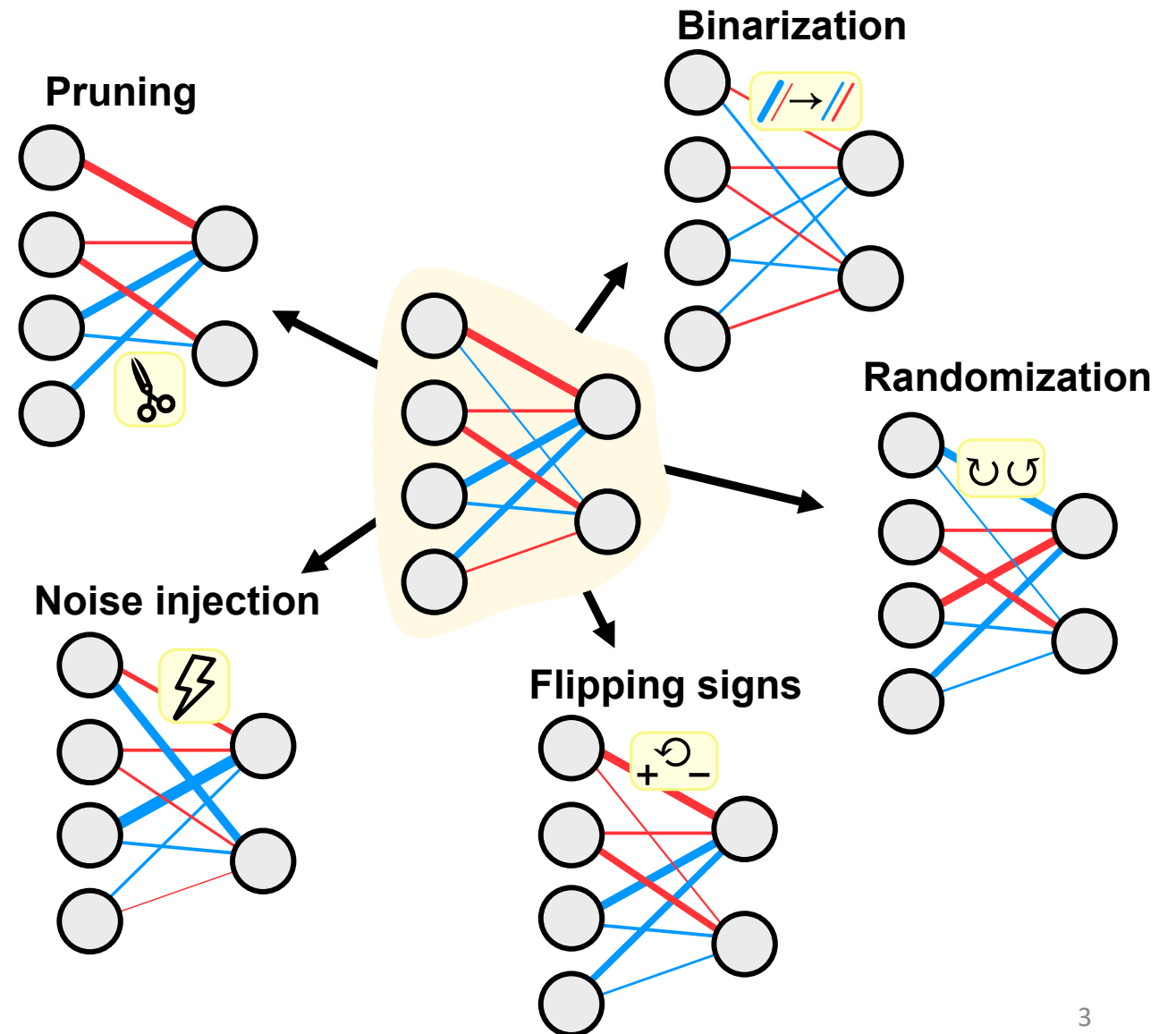


Heatmap of SSIM distances



Data-agnostic probes for MLPs

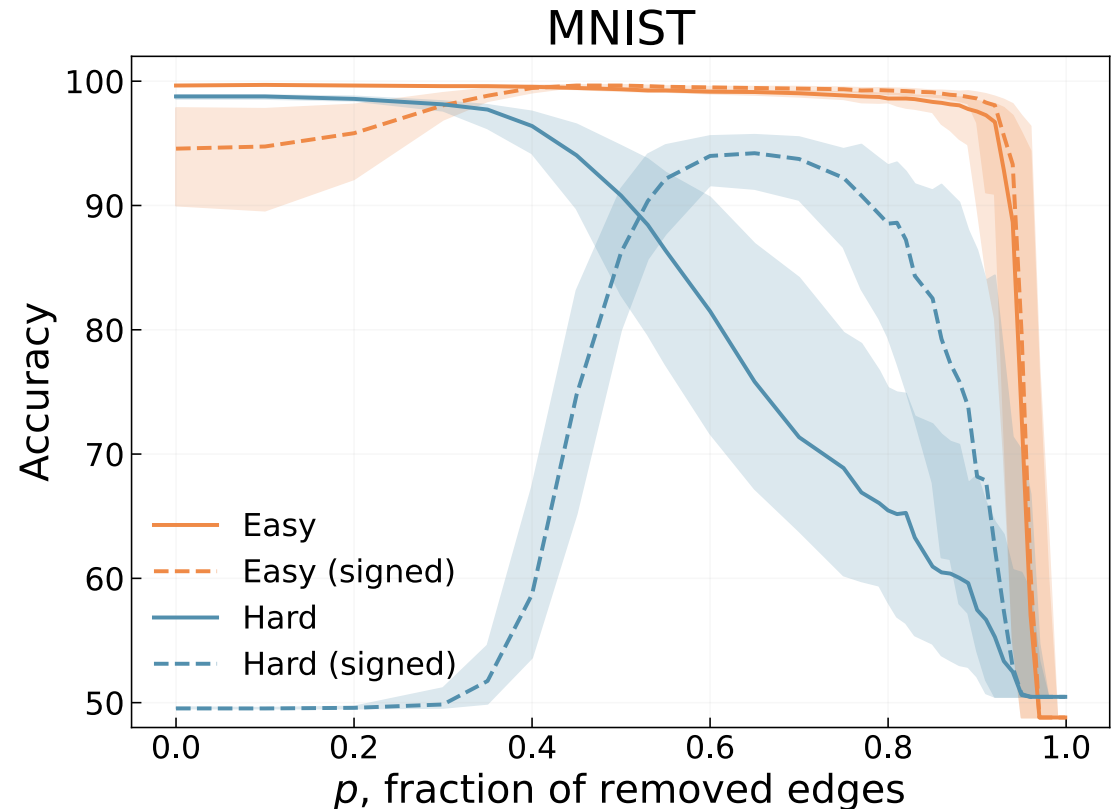
- Pruning and binarization
- Noise injection
- Flipping signs
- Bipartite network randomization



Pruning and binarization

Setup:

- Remove smallest weights first
 - Binarize and measure accuracy
- Signed hard-task model exhibit a performance transition
 - The easy-task model remains quite robust



Easy task



Hard task



Noise injection

Setup:

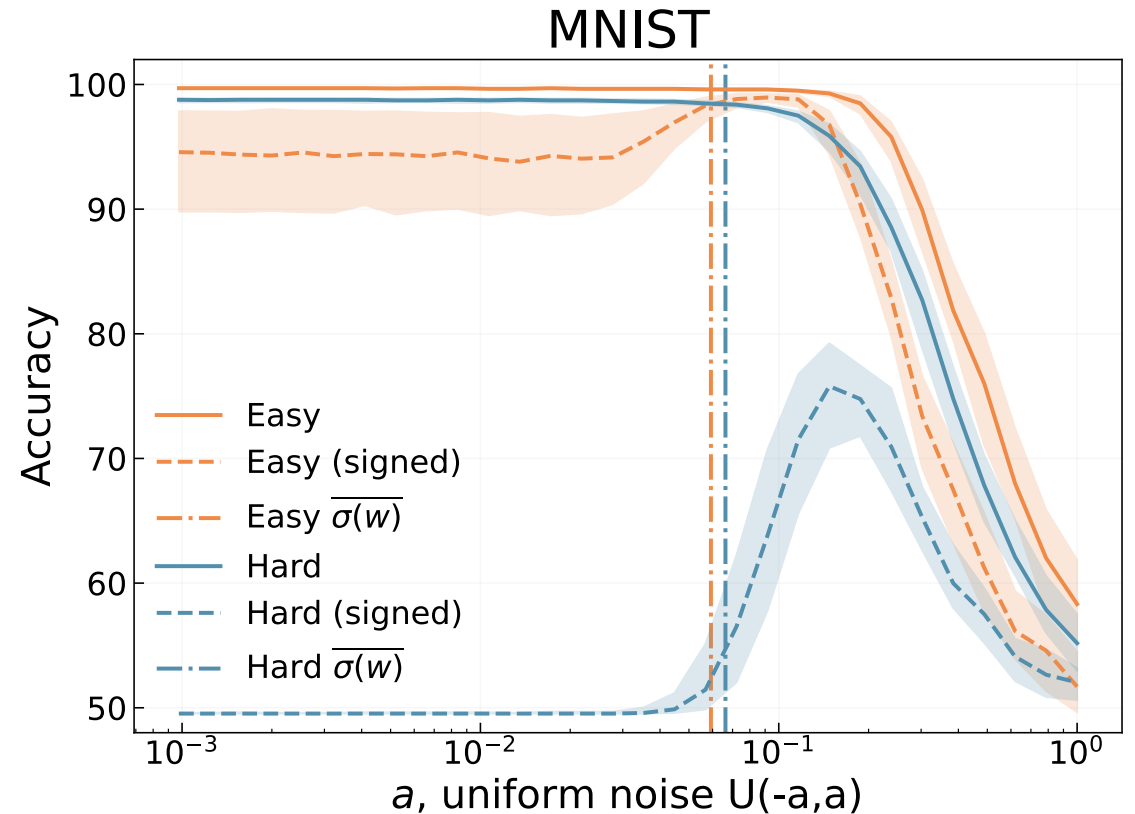
- Perturb each edge weight
- Binarize and measure accuracy

- Stochastic resonance-like effect for both models

Easy task



Hard task

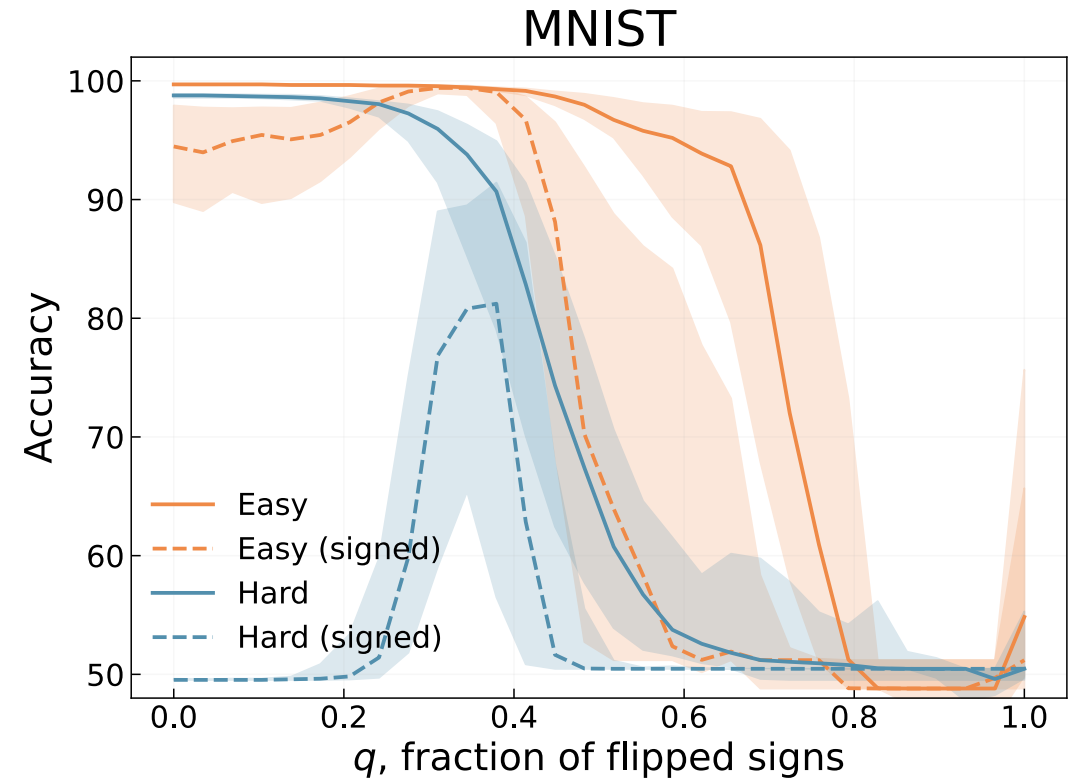


Flipping edge sign

Setup:

- Flip signs of the smallest weights
- Binarize and measure accuracy

- Again, performance peaks when a certain fraction of the weights' signs are flipped.



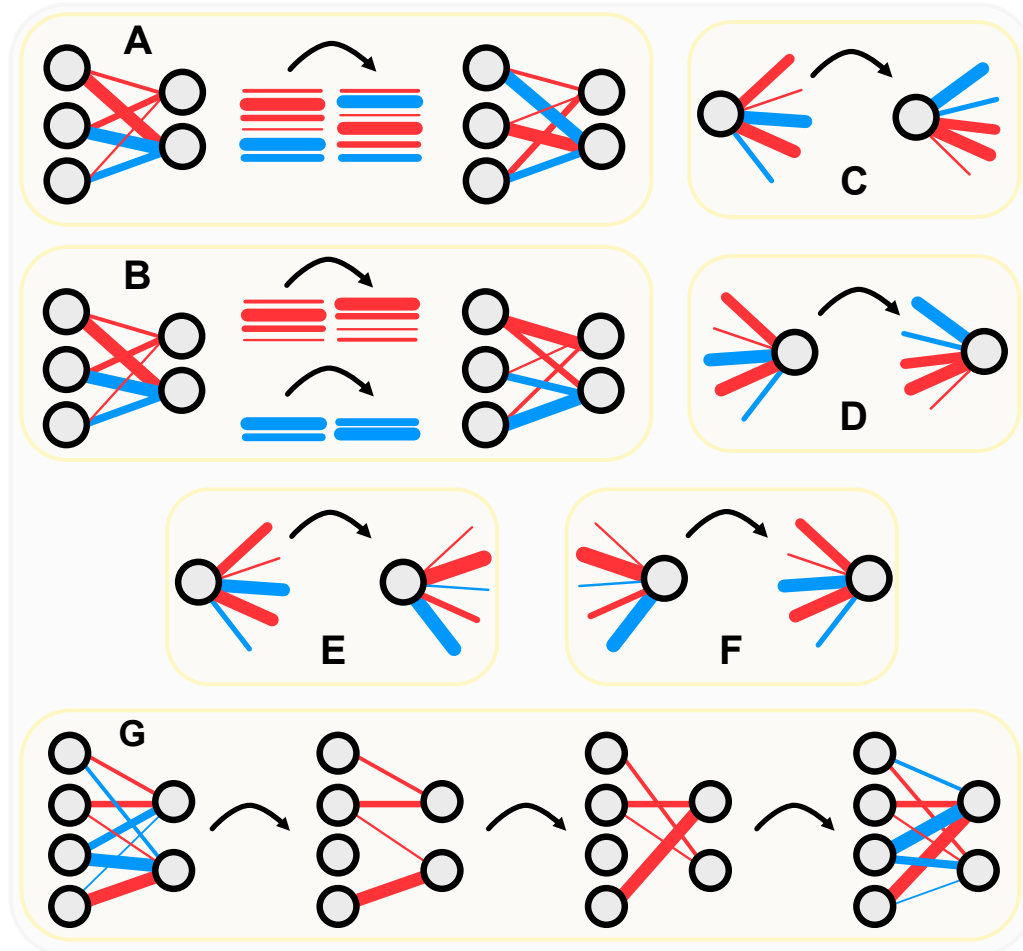
Easy task



Hard task



Bipartite network randomization



| Type | α | k_L^- | k_L^+ | s_L^- | s_L^+ | k_R^- | k_R^+ | s_R^- | s_R^+ | Keeps original sign |
|------|----------|---------|---------|---------|---------|---------|---------|---------|---------|---------------------|
| A | ✓ | - | - | - | - | - | - | - | - | ✓ |
| B | ✓ | ✓ | ✓ | - | - | ✓ | ✓ | - | - | ✓ |
| C | ✓ | ✓ | ✓ | ✓ | ✓ | - | - | - | - | ✓ |
| D | ✓ | - | - | - | - | ✓ | ✓ | ✓ | ✓ | ✓ |
| E | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | - | - | ✓ |
| F | ✓ | ✓ | ✓ | - | - | ✓ | ✓ | ✓ | ✓ | ✓ |
| G | ✓ | ✓ | ✓ | - | - | ✓ | ✓ | - | - | - |

α - fraction of positive links

k_L^- k_L^+ k_R^- k_R^+ (positive/negative) degree for (left/right) nodes

s_L^+ s_L^- s_R^- s_R^+ (positive/negative) strength for (left/right) nodes

Keeps original sign - edge retains its original positive or negative sign

Bipartite network randomization

Setup:

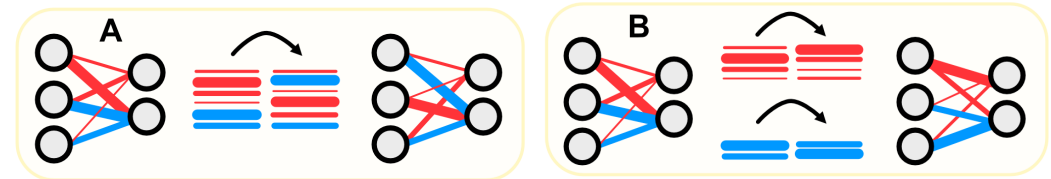
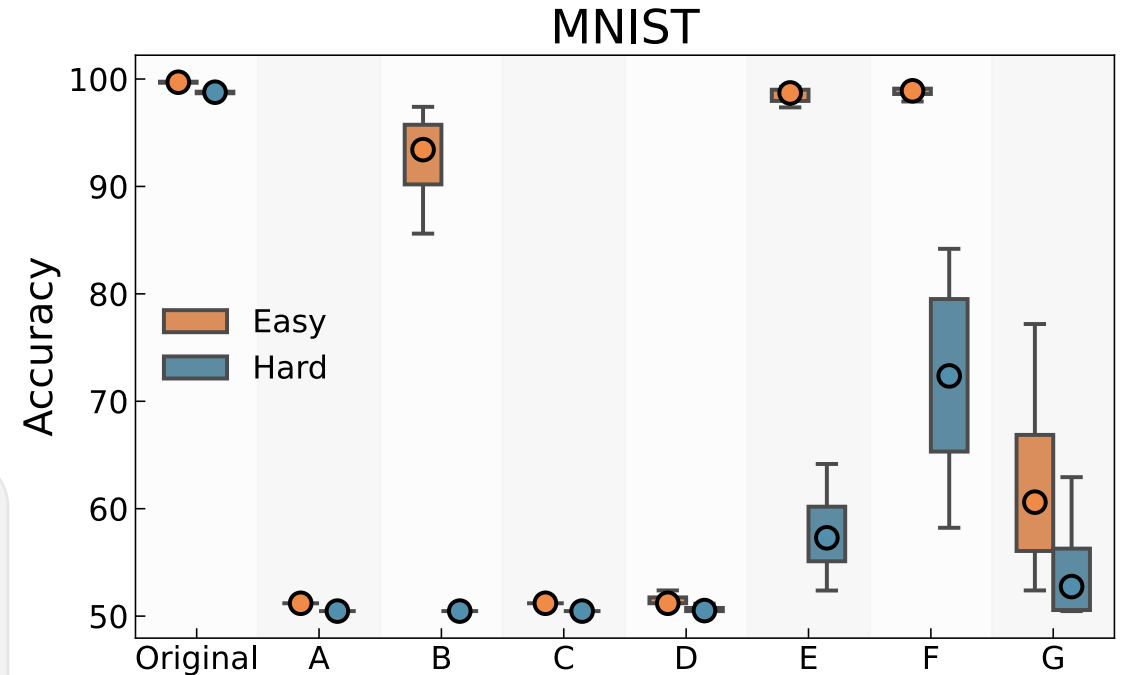
- Apply a bipartite randomization
- Measure accuracy

- Preserving the edge sign maintains high accuracy (randomization B, E, F)
- Hard-task model is less robust to randomization

Easy task

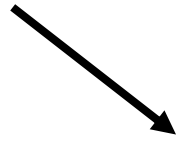
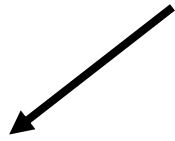


Hard task



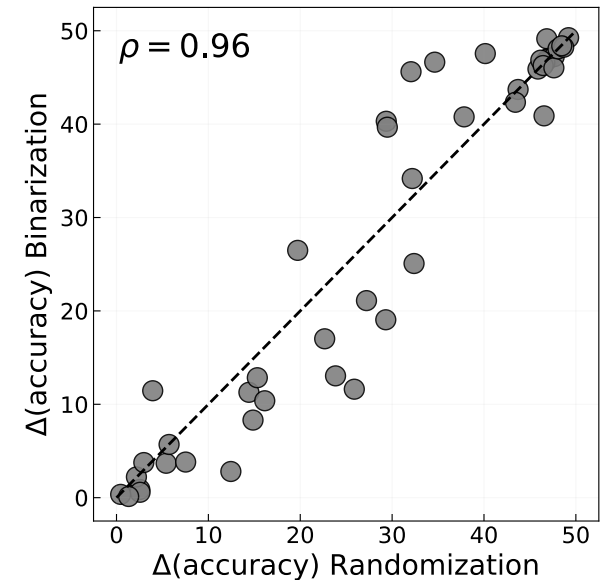
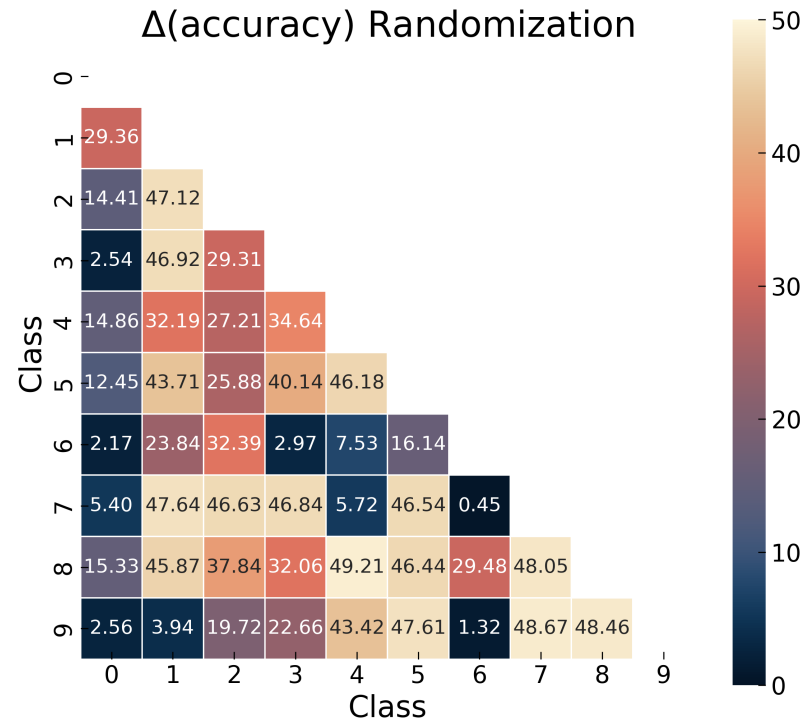
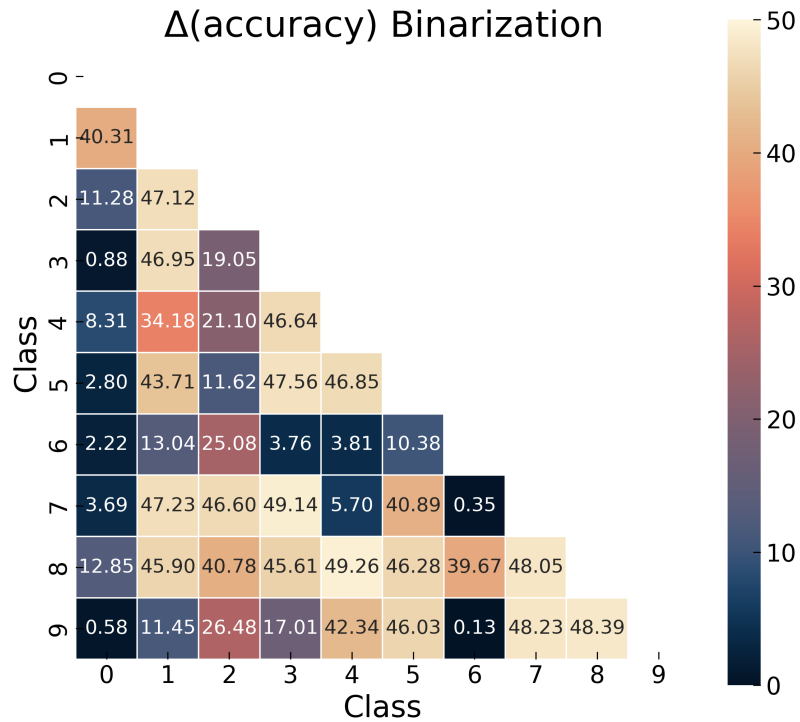
Defining task complexity

Performance gap between



full-precision and binarized models

full-precision and randomized models



Conclusions

- Data-agnostic probes for measuring task complexity.
- The robustness of learned weight structures depends on task difficulty.
- Sign structure matters more than weight magnitude.

Future directions:

- Compare with representational similarity measures: CCA, CKA, RSA.
- Can task complexity guide the design of better architectures, initialization schemes, or compression methods?

Questions?

MACHINE LEARNING
Science and Technology

PAPER • OPEN ACCESS

Task complexity shapes internal representations and robustness in neural networks

Robert Jankowski*, Filippo Radicchi, M Ángeles Serrano, Marián Boguñá and Santo Fortunato

Published 8 April 2026 • © 2026 The Author(s). Published by IOP Publishing Ltd

[Machine Learning: Science and Technology](#), Volume 7, Number 2

Citation Robert Jankowski et al 2026 *Mach. Learn.: Sci. Technol.* 7 025045

DOI 10.1088/2632-2153/ae5503



R.Jankowski@tudelft.nl



<https://robertjankowski.github.io/>