Degree linear-probe acc (penultimate features): train=100.00%, test=100.00%  
Mask ablation accuracies:

* baseline\_acc: 95.43%
* zero\_mask\_acc: 49.97%
* shuffled\_mask\_acc: 85.83%
* random\_degree\_mask\_acc: 85.78%

explain what thus means?

Here’s what those lines indicate:

* Degree linear-probe acc: 100% train/test  
  The penultimate features encode the polynomial’s degree perfectly (linearly decodable). The network has learned/propagated degree info, very likely via the 9-D mask.
* baseline\_acc: 95.43%  
  Normal test accuracy with the true mask. Good performance.
* zero\_mask\_acc: 49.97%  
  Dropping to chance when the mask is zeroed shows the model relies heavily on the mask. A zero mask is also out-of-distribution relative to training.
* shuffled\_mask\_acc: 85.83% and random\_degree\_mask\_acc: 85.78%  
  Giving the wrong mask hurts but does not destroy performance. The model still uses coefficient patterns to classify many cases, but correct degree info from the mask contributes ~10 percentage points vs baseline.

Takeaways:

* The mask is a strong shortcut feature (degree is critical because the threshold is d/2).
* The representation preserves degree information exactly.
* If you want robustness or degree-invariance, train with mask corruption/ dropout, or remove the mask and/or add augmentation so the model can infer from coefficients alone (or pass degree explicitly as a separate, trusted input).