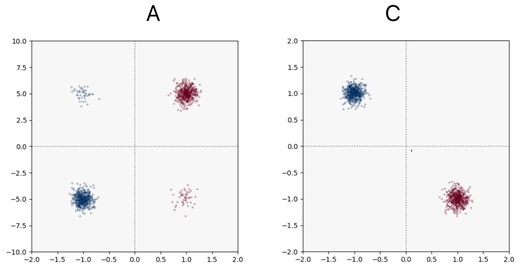
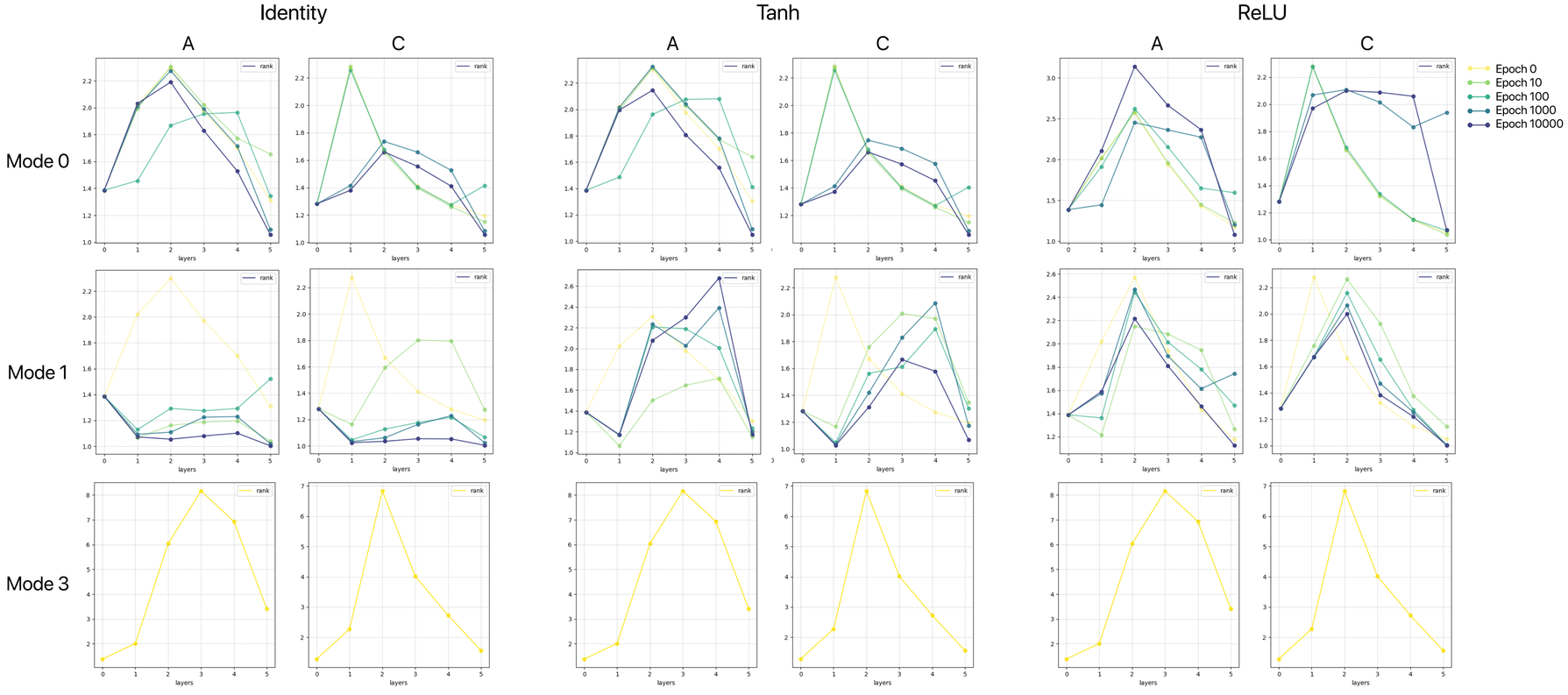
Setups:

	core corr	core gamma	spu corr	spu gamma	noise dim	noise gamma
Setting A	1	1	0.9	5	0	n/a
Setting B	1	1	0.9	5	40	1
Setting C	1	1	0	1	0	n/a
Setting D	1	1	0	1	40	1



Mode 0	Standard rich training; layer-wise representation rank computed
Mode 1	Lazy training; layer-wise representation rank computed
~~Mode 2~~	~~No training; Shred the NTRF matrix into chunks and computing the rank~~
Mode 3	No training; NTRF rank of separate l-layer networks are computed for $0 \leq l < 6$



Hidden Manifold Model (HMM):

n: the number of data points

dl: the dimension of the latent space

da: the dimension of the ambient space

p: the probability of generating non-zero entries in the latent patterns

$W \in \mathbb{R}^{dl \times da}$: a random feature matrix drawn from a standard Gaussian

$Z \in \mathbb{R}^{n \times dl}$: a binary matrix sampled from a Bernoulli with parameter p

$b \in \mathbb{R}^{dl \times 1}$: the teacher weights drawn from a standard Gaussian

$X \in \mathbb{R}^{n \times da}$ denotes the inputs

$Y \in \mathbb{R}^{n \times 1}$ denotes the labels

$$X = \sigma\left(\frac{ZW}{\sqrt{dl}}\right) \quad Y = \begin{cases} 1 \text{ if } Zb^T > 0 \\ 0 \quad o.w. \end{cases}$$

Teacher:

A HMM with

n = 2000 (1000 for train and 1000 for valid)

dl = 4, da = 512, p = 0.5

σ = identity / tanh

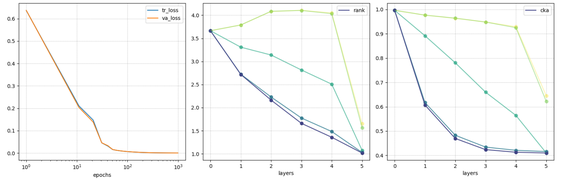
Student:

A 5-layer model: dims=[512, 100, 100, 100, 100, 2]

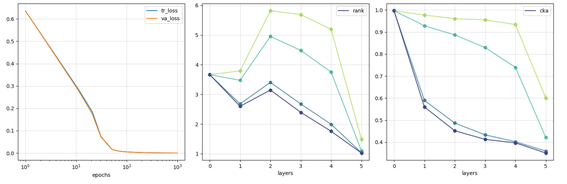
act_fn = identity / tanh

Results:

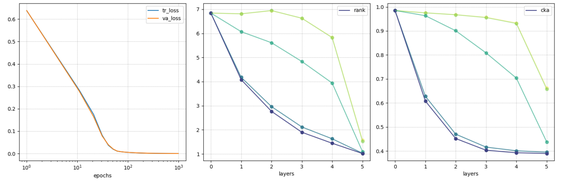
σ = identity, act_fn= identity



σ = identity, act_fn=tanh



σ = tanh, act_fn= identity



σ = tanh, act_fn=tanh