Additive Decoders for Latent Variables Identification and Cartesian-Product Extrapolation

Sébastien Lachapelle*, Divyat Mahajan*, Ioannis Mitliagkas & Simon Lacoste-Julien

Contributions

We introduce additive decoders: a simple architecture similar to object-centric decoders for which we can prove both disentanglement and extrapolation guarantees.

Motivations

- Explain why object-centric representation learning (OCRL) methods (such as Slot Attention [2]) perform disentanglement when they are trained only to reconstruct, without supervision.

Unidentifiability in representation learning

- Decoder: \( f: \mathbb{R}^d \rightarrow \mathbb{R}^l \) (\( d \gg l \))
- Encoder: \( g: \mathbb{R}^l \rightarrow \mathbb{R}^d \)
- Assume they solve the reconstruction problem, i.e. \( \mathbb{E} \| x - f(g(x)) \|^2 = 0 \)
- By taking \( f = f \circ \varphi \) and \( g = g \circ \psi \) where \( \varphi \) is some invertible map, we also solve the reconstruction problem.
- But \( f \) and \( g \) might have drastically different representations.
- This is a problem if we hope to learn a disentangled representation.
- Our solution: restrict \( f \) to be additive!
- Closely related to nonlinear independent component analysis (ICA) [1].

Disentanglement guarantee (identifiability)

Theorem (Informal): Assume that
1. Data generating process: \( x = \sum_{b \in \mathcal{B}} f^b(z_b) \) with \( z \sim \mathcal{N}(0, I) \)
2. Learned decoder is additive as well.
3. Ground-truth decoder is “sufficiently nonlinear” (see paper)
4. More regularity conditions... (the \( \mathcal{C} \) differentiation)
Then, \( \mathbb{E}_{z \sim \mathcal{N}(0, I)} [x - f(g(x))]^2 = 0 \) implies that, for all \( B \subseteq \mathcal{B} \),
\[ f^B(z) = f^B(z_{\pi(B)}) + \sum_{b \in \mathcal{B} \setminus B} f^b(z_{\pi(B)}) \]

Cartesian-Product Extrapolation

Corollary (Informal): Under the same assumptions as the previous theorem:
The learned block-specific decoders imitate the ground-truth ones!
- Latent factors can be dependent and have an almost arbitrary support.

Experiments: Datasets

Scalar Latent dataset: Balls move only along y-axis
- \( \dim(z) = 2 \)
- \( \mathcal{B} = \{(1,2), (3,4)\} \)

Block Latent dataset: Balls move only along both x, y axis
- \( \dim(z) = 4 \)
- \( \mathcal{B} = \{(1,2), (3,4)\} \)

Experiments: Disentanglement

Experiments: Extrapolation

Additive decoder
- Learn latent space
- Generated images

Non-additive decoder
- Learn latent space
- Generated images

Evaluation Metric

- Independent Case: \( z_{\mathcal{B}_A} \perp z_{\mathcal{B}_B} \)
- Dependent Case: \( z_{\mathcal{B}_A} \perp z_{\mathcal{B}_B} \)