1. Neuro-Symbolic Reasoning

We introduce Neuro-Symbolic Continual Learning (NeSy-CL), a new machine learning problem, where the machine has to:

1. Learn over a sequence of NeSy tasks;
2. Acquire high-quality concepts, avoiding reasoning shortcuts;
3. Preserve the knowledge on the concepts and labels.

2. Continual Learning

- Learning is divided in tasks
- Fine-tuning: the models leads to catastrophic forgetting.
- CL Strategies make use of memory and/or regularization to prevent it.

No method makes use of prior knowledge.

3. Neuro-Symbolic Continual Learning

We extend it to a knowledge on the concepts and labels. For each task, the knowledge on the concepts and labels.

Theorem (3.2): A model with parameters $\theta$ attains maximal likelihood, i.e. $\theta \in \Theta(K, D)$, if and only if, for all $(x, y) \in D$, it holds that $p_{\theta}(C | x) = \mathbb{K}[y | x]$.

Example: Consider MNIST-Addition with only $\theta_1$: $0 + 1 = 1$ and $0 + 2 = 2$. Then:

- $0 \rightarrow 0$;
- $1 \rightarrow 1$;
- $2 \rightarrow 2$;
- $0 \rightarrow 1$;
- $1 \rightarrow 0$;
- $2 \rightarrow 1$.

are two optimal solutions

Mitigation with concept supervision. Q3: We show that only few is sufficient!

4. Problem Statement

We denote input $x \in \mathbb{R}^d$, concepts $c \in \mathbb{N}$, labels $y \in \mathbb{N}$, and prior knowledge $K$. Data are distributed according to:

$$p(y \mid C, X, Y, K_0) := p(y \mid C, K_0) \cdot p(C \mid X) \cdot p(Y)$$

Example: MNIST-Addition consists of sums between digits, i.e. $0 + 1 = 1$.

We extend it to a class/concept-incremental NeSy-CL benchmark.

5. Reasoning with DeepProbLog

DeepProbLog = probabilistic logic programming + neural predicates:

$$p_{\theta}(y \mid x, K) = \sum_c p_{\theta}(y \mid c, x) \cdot p_{\theta}(c \mid x)$$

$$p_{\theta}(y \mid c) = \frac{1}{Z(c, K)} \cdot I(c, y) \mid K$$

For each task, we learn the parameters $\theta$ via maximum likelihood.

6. NeSy predictors fall prey of Reasoning Shortcuts

A Reasoning Shortcut is an optimal solution with incorrect concepts.

Theorem (3.2): A model with parameters $\theta$ attains maximal likelihood, i.e. $\theta \in \Theta(K, D)$, if and only if, for all $(x, y) \in D$, it holds that $p_{\theta}(C | x) = \mathbb{K}[y | x]$.

Example: Consider MNIST-Addition with only $\theta_1$: $0 + 1 = 1$ and $0 + 2 = 2$. Then:

- $0 \rightarrow 0$;
- $1 \rightarrow 1$;
- $2 \rightarrow 2$;
- $0 \rightarrow 1$;
- $1 \rightarrow 0$;
- $2 \rightarrow 1$.

are two optimal solutions

Mitigation with concept supervision. Q3: We show that only few is sufficient!