Self-Learning using Venn-Abers Predictors

Côme Rodriguez, *Vitor Martin Bordini*, Sébastien Destercke and Benjamin Quost

Heudiasyc lab

September 14, 2023

Plan

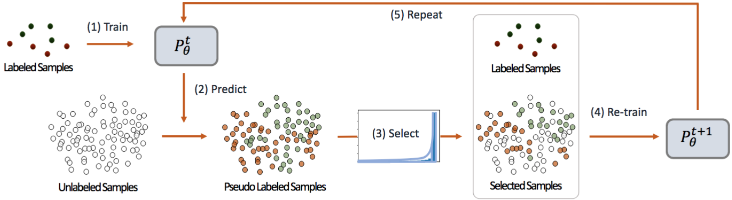
- Introduction
- How do we produce credal labels ? Venn-Abers predictors.
- How do we train on credal labels ? Optimistic loss.
- Experiments
- Conclusion and Perspectives

Human labelling problems

- Time-consuming.
- Need experts on some cases (e.g, medical applications).

A solution ? Self Learning [1]

- A model creates its own labels to train itself.
- How ?



Hard labels

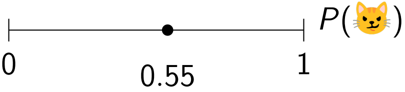
- Hard labels : one-hot encoding of each class.
- No uncertainty on prediction.
- Same label, different situations ?

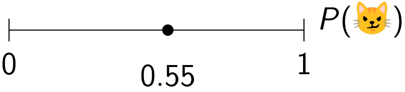
51% 🌧️ $\rightarrow$ ☂️ ?

90% 🌧️ $\rightarrow$ ☂️ ?

Soft labels

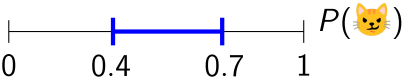
- Soft labels : a probability distribution.
- Uncertainty on prediction... Ambiguity ?

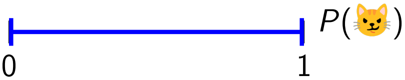




Credal Sets

- Credal Sets : sets of probability distributions.
- Uncertainty on uncertainty.
- Complete ignorance.



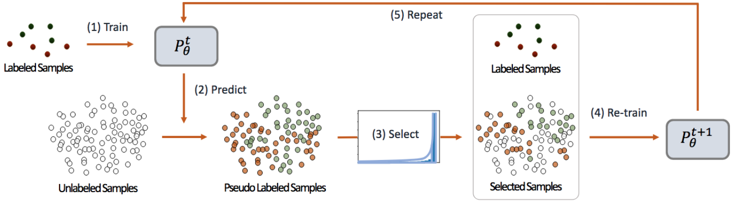


Calibration

- Model prediction reflects true probability ?
- Conformal Prediction notion : $P(y \in \Gamma^{\delta}) \geq 1 - \delta$.
- Venn-Abers notion : $P(y = 1 | h(x) = \alpha) = \alpha$.

Self-Learning from Credal Sets[1]

- A way to produce credal labels, ideally calibrated (Venn-Abers predictors).
- A way to integrate credal label in learning (optimistic loss).



Plan

- Introduction
- How do we produce credal labels ? Venn-Abers predictors.
- How do we train on credal labels ? Optimistic loss.
- Experiments
- Conclusion and Perspectives

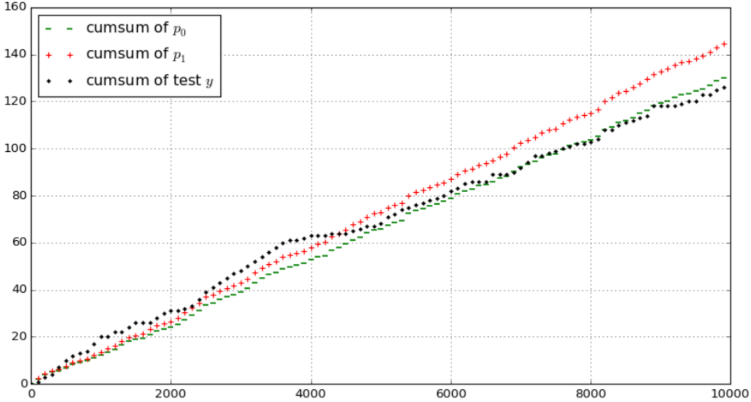
Inductive Venn-Abers Predictors (IVAP)

- Limited to binary problems.
- The algorithm builds a **calibrated credal set** $[\underline{p}, \overline{p}]$.

The algorithm

1. Train a model on the training set $\mathcal{D}_T$.
2. Compute scores $h(x_0), ..., h(x_k)$ on the calibration set $\mathcal{D}_C$.
3. Compute two isotonic regressions : one for $\{(h(x_0), y_0), ..., (h(x_k), y_k), h(x, 0)\}$ and other for $\{(h(x_0), y_0), ..., (h(x_k), y_k), (h(x), 1)\}$ to obtain functions $f_0(h(x))$ and $f_1(h(x))$ respectively.
4. The output is $[\underline{p}, \overline{p}] := [f_0(h(x)), f_1(h(x))]$

Graphical representation [2]



Plan

- Introduction
- How do we produce credal labels ? Venn-Abers predictors.
- How do we train on credal labels ? Optimistic loss.
- Experiments
- Conclusion and Perspectives

Loss function

- Classical Loss is usually between two probabilities
- How do I adapt it to Credal Sets ?

Loss function

- Classical Loss is usually between two probabilities
- How do I adapt it to Credal Sets ?
- Optimistic approach : take the minimum over all values in the set,i.e, $\mathcal{L}_{min}(K, h_\theta(x)) = \min_{p \in K} \mathcal{L}(p, h_\theta(x))$.
- Usually an easier problem to solve than taking the maximum, for example.

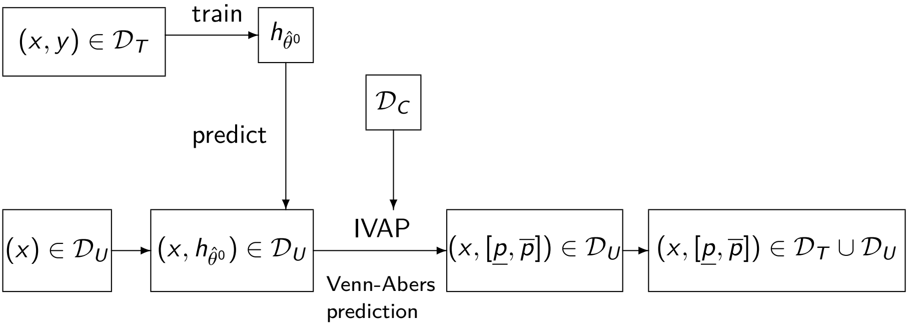
Our approach

$$\boxed{(x, y) \in \mathcal{D}_T} \xrightarrow{\text{train}} \boxed{h_{\hat{\theta}^0}}$$

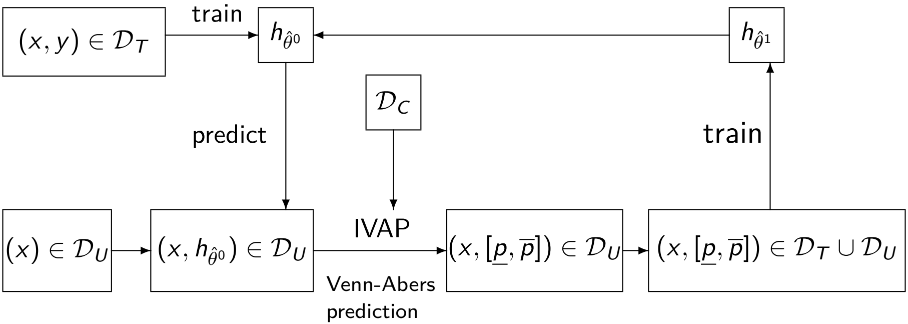
Our approach

$$
\begin{tikzcd}
(x, y) \in \mathcal{D}_T \arrow[r, "\text{train}"] & h_{\hat{\theta}^0} \arrow[d, "\text{predict}"'] \\
(x) \in \mathcal{D}_U \arrow[r] & (x, h_{\hat{\theta}^0}) \in \mathcal{D}_U
\end{tikzcd}
$$

Our approach



Our approach



Plan

- Introduction
- How do we produce credal labels ? Venn-Abers predictors.
- How do we train on credal labels ? Optimistic loss.
- Experiments
- Conclusion and Perspectives

Setup

- Probability model : Neural Network (1 hidden layer) since it is poorly calibrated.
- Three different strategies : Self-Learning with hard labels(SL),soft labels (SLSL) and Venn-Abers (SLVA)

Dataset	number of neurons (hidden layer)	λ
Breastcancer	5	0.01
Digits	10	0.01
Australian	4	0.005
Banknote	2	0.01
Heart disease	5	0.005
Adult	10	0.001

Performance

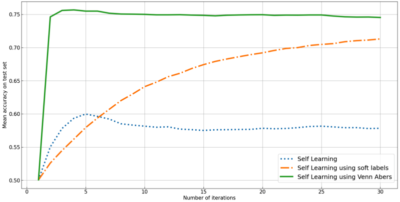
- Performance of SLVA is better in most cases.

Dataset	$\overline{a}$ (mean over iterations) SL	SLSL	SLVA	a_{30} (accuracy at 30 iterations) SL	SLSL	SLVA
Breastcancer	0.953	0.951	**0.959**	0.961	0.962	**0.968**
Digits	**0.851**	0.817	0.838	**0.882**	0.838	0.876
Australian	0.815	0.789	**0.827**	0.857	0.851	**0.859**
Banknote	0.907	0.881	**0.926**	0.980	0.976	**0.986**
Heart disease	0.775	0.768	**0.782**	0.815	**0.820**	0.818
Adult	0.578	0.653	**0.741**	0.578	0.713	**0.745**

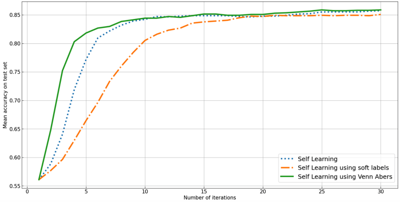
Results

- Faster convergence for SLVA.

Adult



Australian



Plan

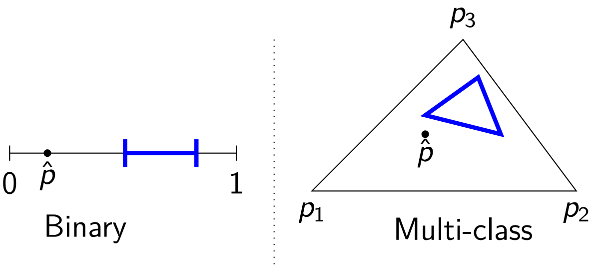
- Introduction
- How do we produce credal labels ? Venn-Abers predictors.
- How do we train on credal labels ? Optimistic loss.
- Experiments
- Conclusion and Perspectives

Conclusion

- Self-learning is promising, but can be biased.
- Adding uncertainty into our prediction leads to better performance because the model tends to be less biased.
- IVAP also guarantees a calibrated result.

Perspectives

- Venn-Abers is limited to binary problems, how to overcome that ? Venn Predictors,
- Optimization problem no longer on a interval, but on a polytope on the $\mathcal{R}^{|Y|}$ space.



References

[1] Cascante-Bonilla, P., Tan, F., Qi, Y., Ordonez, V. : Curriculum labeling : Revisiting pseudo-labeling for semi-supervised learning. In : Proceedings of the AAAI Conference on Artificial Intelligence. vol. 35, pp. 6912–6920 (2021)

[2] Tocaccelli, P. : Tutorial on venn-abers prediction. Conformal and Probabilistic Prediction with Applications (Jun 2017), https://cml.rhul.ac.uk/copa2017/presentations/VennTutorialCOPA2017.pdf