Novel Continual Learning Techniques on Noisy Label Datasets

AXYON.AI

Monica Millunzi[1,2], Lorenzo Bonicelli[1], Alberto Zurli[2], Alessio Salman[2], Jacopo Credi[2], Simone Calderara[1]

[1]University of Modena and Reggio Emilia, Italy
[2]Axyon AI

[1]{name.surname}@unimore.it
[2]{name.surname}@axyon.ai

AImage Lab
University of Modena and Reggio Emilia

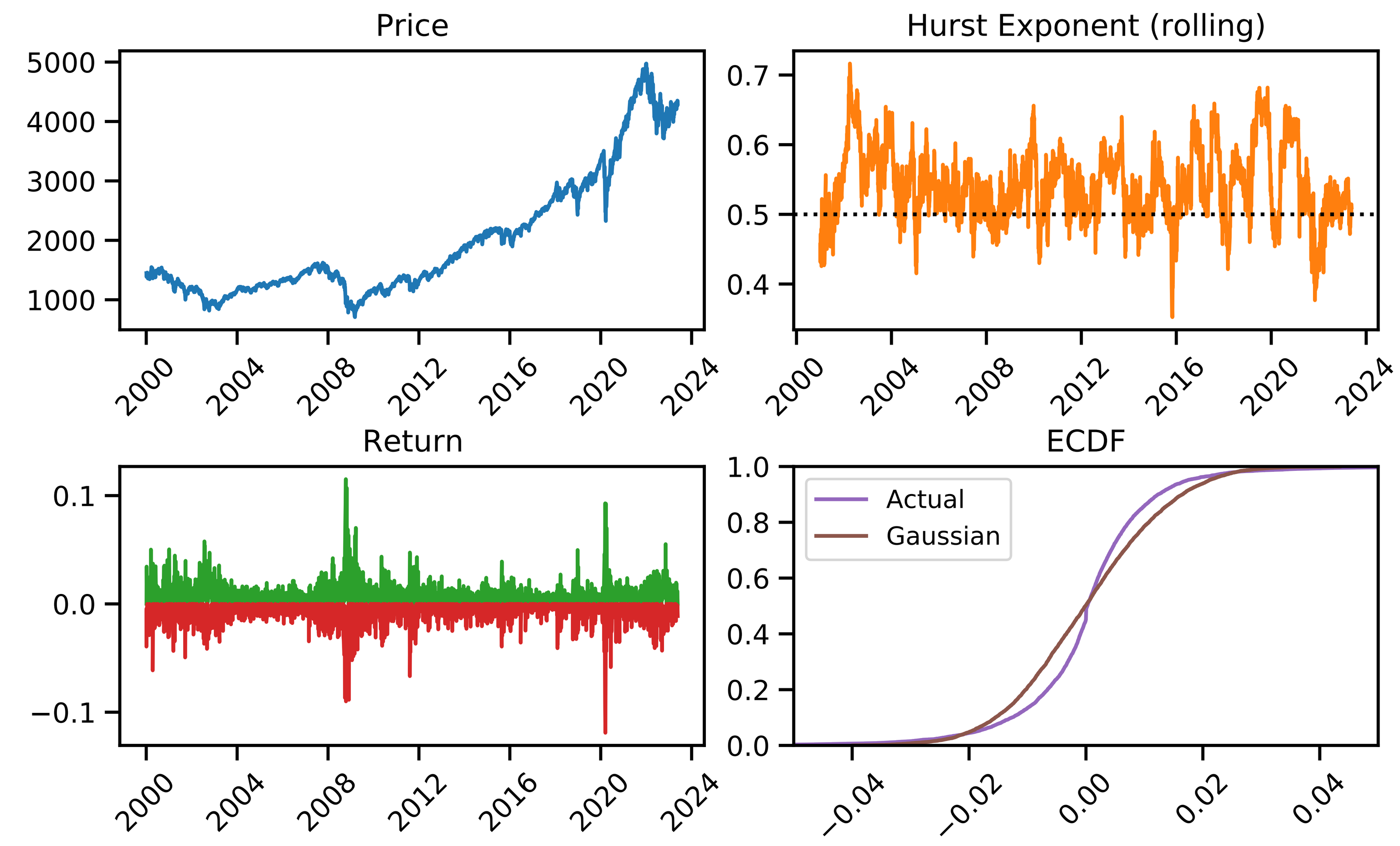
UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

Noise and Non-stationarity in Financial Machine Learning

Financial Machine Learning [2] differs from standard ML applications in many aspects, and in particular:

1. Financial asset prices are **non-stationary** time-series, and differencing does not fully address the problem:
   - Asset returns are (negatively) autocorrelated and heteroscedastic, exhibiting volatility clusters
   - Their distribution is non-Gaussian, with large kurtosis ("fat-tails")
2. Financial asset prices/returns exhibit a very poor signal-to-noise ratio, exposing ML models to overfitting.
   - E.g. Mean Hurst Exponent of daily S&P 500 closing price is around 0.54 (yearly lag).

Example: S&P 500 Index



Learning with Noisy Labels

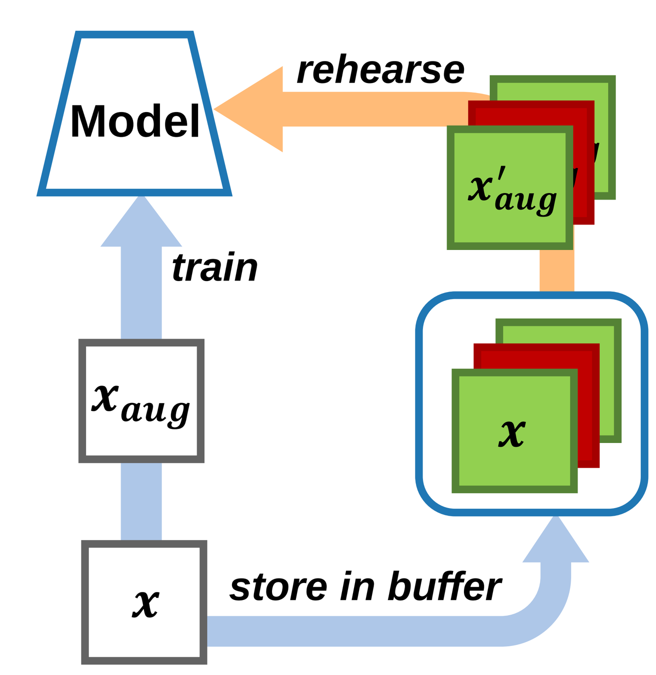
Data come from a **noisy distribution** $\tilde{\mathcal{D}} = \{(x_i, \tilde{y}_i)\}_{i=1}^{N}$, with $\tilde{\mathcal{Y}}$ being the noisy label space

Continual Learning

- Learning from a **sequence** of tasks $\{\mathcal{D}_1, \dots, \mathcal{D}_T\}$
- **Experience Replay** (ER): train with current data stream $\mathcal{D}_t$ and a buffer $\mathcal{M}$ of past data

$$\theta^* = \arg\min_{\theta} \mathop{\mathbb{E}}_{(\mathbf{x},\tilde{y})\sim\mathcal{D}_t} \left[\mathcal{L}(f(\mathbf{x}), \tilde{y})\right] + \mathcal{L}_R$$

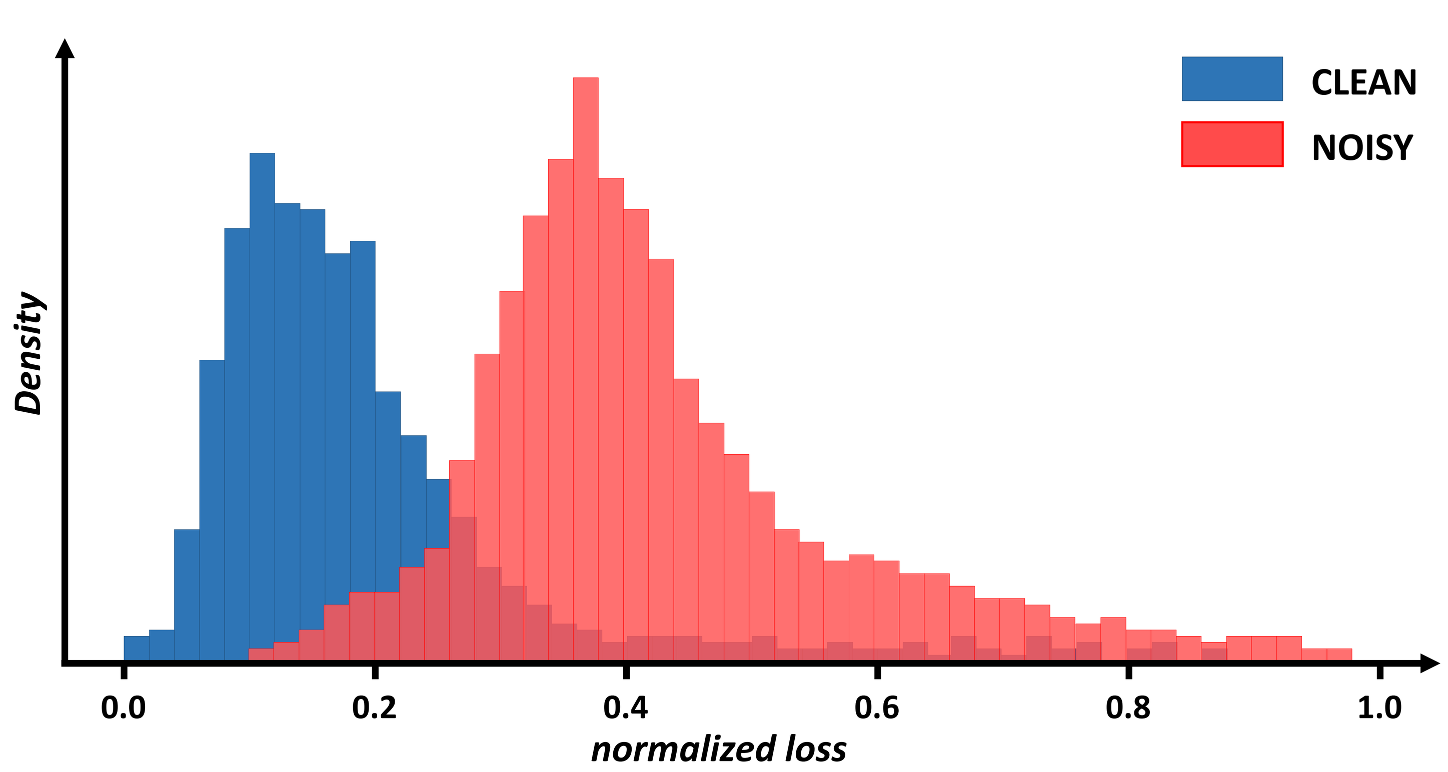
$$\mathcal{L}_R = \mathop{\mathbb{E}}_{(\mathbf{x}_r,\tilde{y}_r)\sim\mathcal{M}} \left[\mathcal{L}(f_\theta(\mathbf{x}_r), \tilde{y}_r)\right]$$



Problem formulation and Experiments

PROBLEM: Samples from the noisy label space $\tilde{\mathcal{Y}}$ are stored inside the buffer $\mathcal{M}$

- Exploit **small-loss** criterion [3] to identify *clean* and *noisy* examples
- Fill the *replay memory* $\mathcal{M}$ with the clean examples only, selected via Gaussian Mixture Model (GMM) or Oracle



Method	**Split-N-CIFAR-10**			
Noise rate (symmetric)	0%	20%	40%	60%
Multitask	91.69	82.02	72.04	54.83
Finetuning	19.66	18.83	18.02	15.99
ER-ACE [1]	71.15	53.82	37.43	22.87
ER-ACE w/ Oracle	-	51.10	39.06	23.57
ER-ACE w/ GMM (**OURS**)	-	52.90	37.95	24.93

Table 1: Final Average Accuracy [↑] of ER with Asymmetric Cross Entropy (ER-ACE) combined with two different techniques to identify noisy samples and prevent storing them inside the memory buffer; comparison with some baseline methods.

References

[1] Caccia, Lucas, et al. "New insights on reducing abrupt representation change in online continual learning.", In arXiv preprint arXiv:2203.03798 (2022).

[2] Coqueret, Guillaume. "Machine Learning in Finance: From Theory to Practice: by Matthew F. Dixon, Igor Halperin, and Paul Bilokon, Springer (2020). ISBN 978-3-030-41067-4. Paperback." (2021): 9-10.

[3] Gui, Xian-Jin, Wei Wang, and Zhang-Hao Tian. "Towards understanding deep learning from noisy labels with small-loss criterion." arXiv preprint arXiv:2106.09291 (2021)

Acknowledgements

This work was supported by Axyon AI SRL and has received funding through the Decreto Ministeriale n° 352 of 09/04/2022, under the Recovery and Resilience Facility (PNRR).

Finanziato dall'Unione europea
NextGenerationEU