Continual Learning: from zero to hero

how to build intelligent agents which never stop learning

Computer Science Department – University of Verona

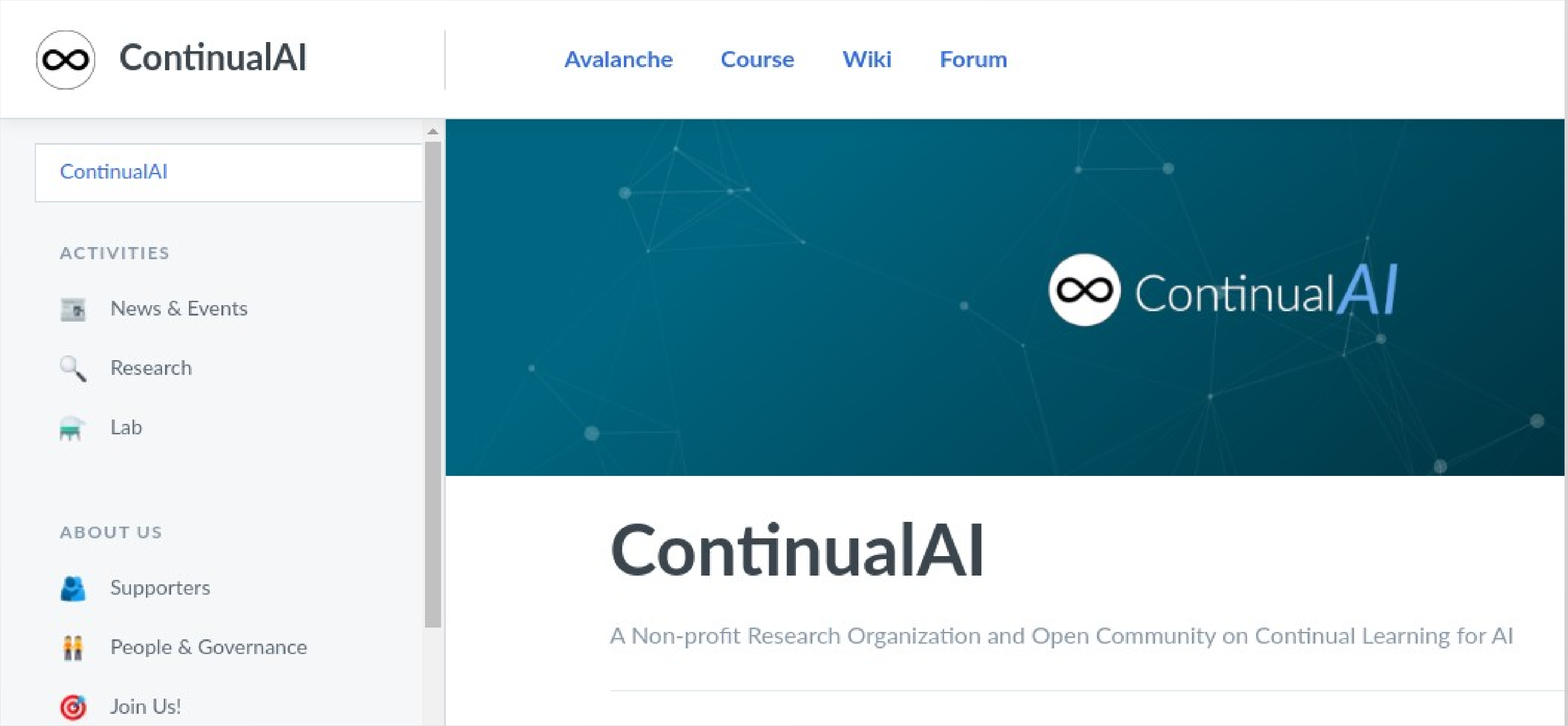
31 May 2022

Andrea Cossu, PhD student
@Scuola Normale Superiore
@University of Pisa
Computer Science Dept.
→ andrea.cossu@sns.it
→ https://www.andreacossu.com

PervasiveAI Lab (Unipi + CNR)
CIML group (Unipi)

Board Member @ ContinualAI
Avalanche maintainer

https://www.continualai.org



Cooking (Supervised) Machine Learning

Ingredients

A dataset D composed by K paired elements
input sample *x* and target *y*

A model M (*we will consider artificial neural networks)*

A loss function L

Recipe

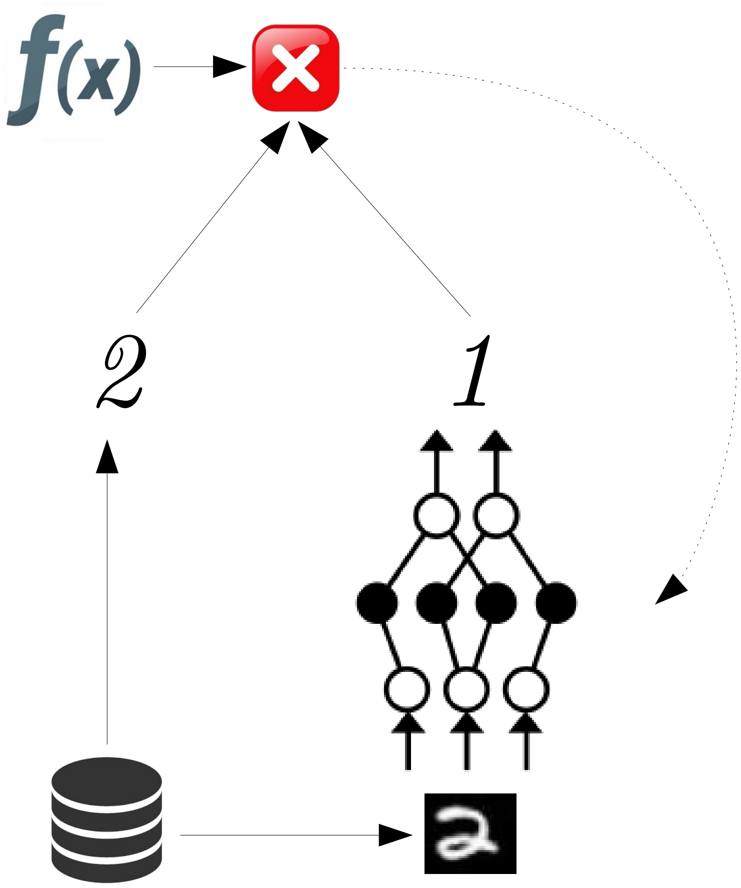
Train the model M by iterating on the dataset D

For each paired element
(start over when running out of elements):

- Compute model output OUT = M(x)
- Compute loss L = distance(OUT, y)
- Update model M

Result

Final model M ready to be used in the real-world!

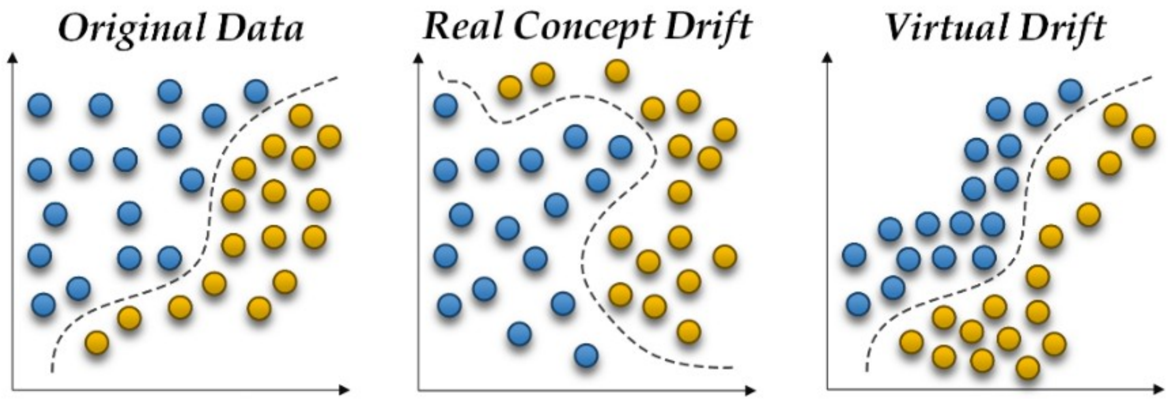


The power of Machine Learning

- Learning from data
  - no need to design ad-hoc, complex features
  - useful representations emerge during training (especially with deep networks)
- Impressive performance on a wide range of applications
  - language, vision, speech…
  - prediction, generation…
- Training is expensive, inference is cheap
- General paradigm for problem-solving, **if** you have enough data

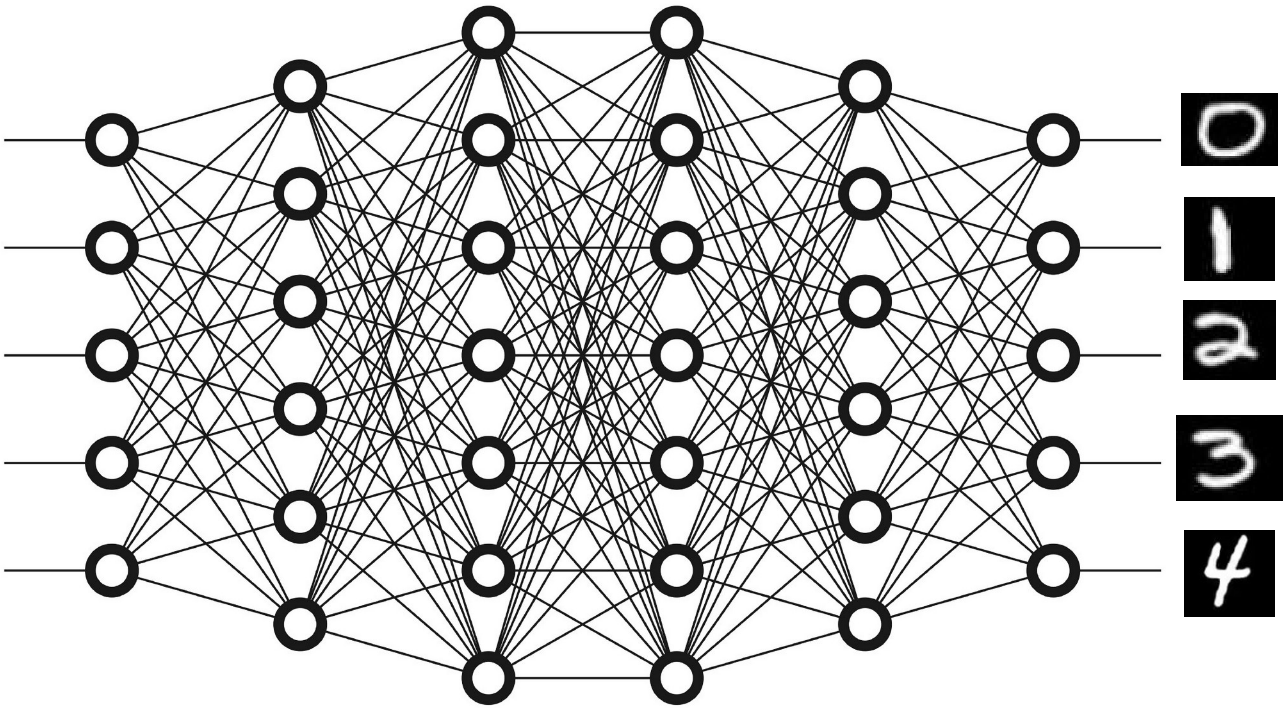
Non-stationary environments

- Data arrives **continually,** it is not entirely available at the beginning
- Data may change over time: drift!
  - Gradual / Abrupt
  - Permanent / Transient / Cyclical (recurrent concepts)
  - Real / Virtual
- Drift detection
  - active / passive
- Covariate Shift

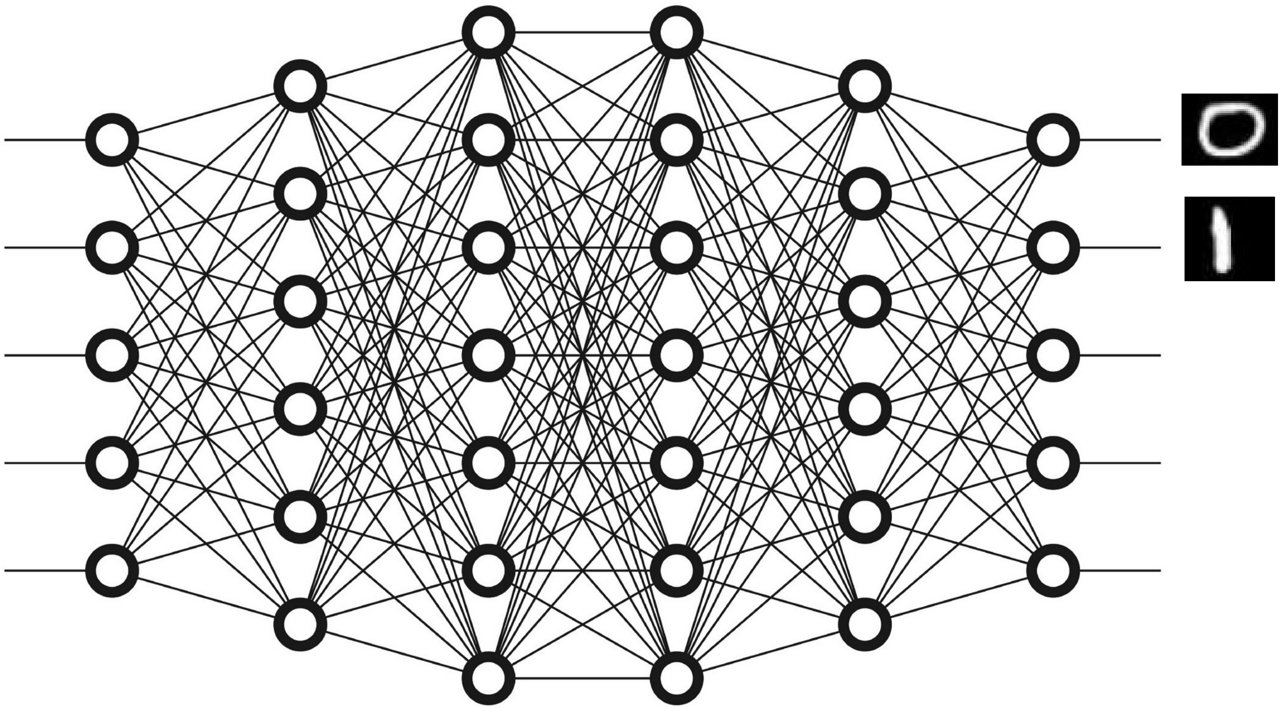


Pesaranghader et. al., "McDiarmid Drift Detection Methods for Evolving Data Streams*: A review*", *IJCNN, 2018*

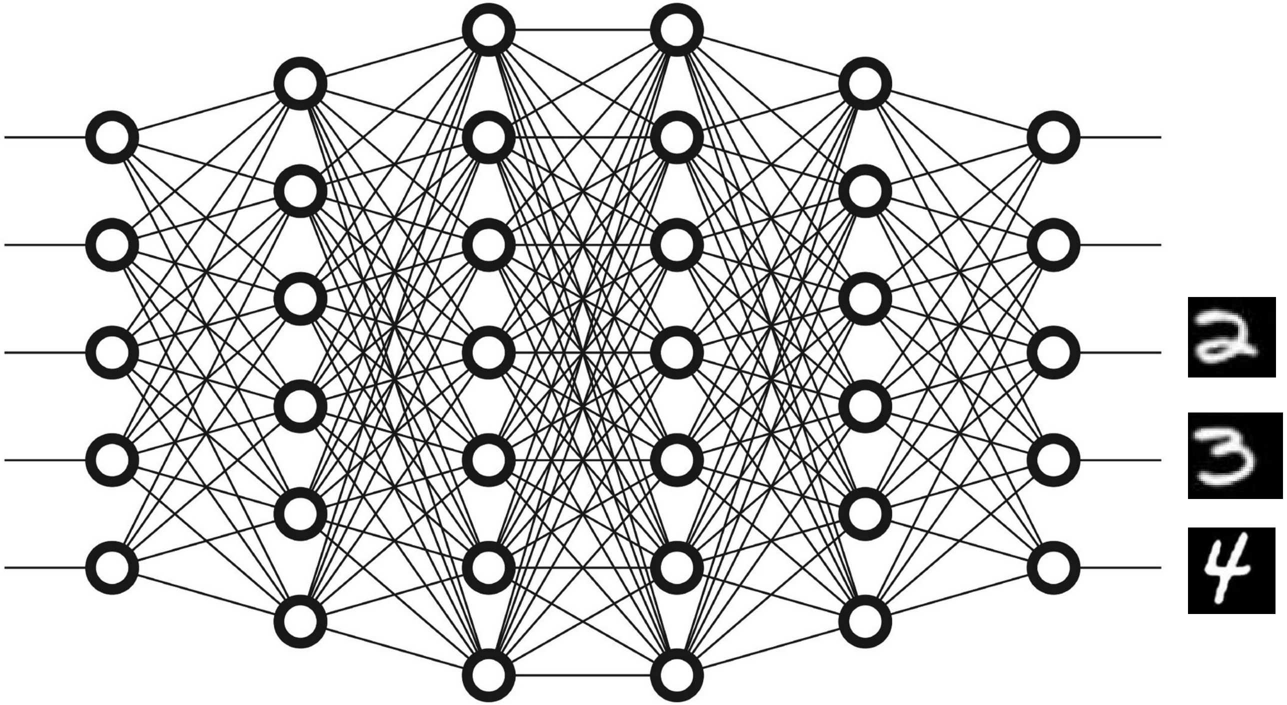
Training continually:
Objective: classifying correctly the digits 0, 1, 2, 3, 4



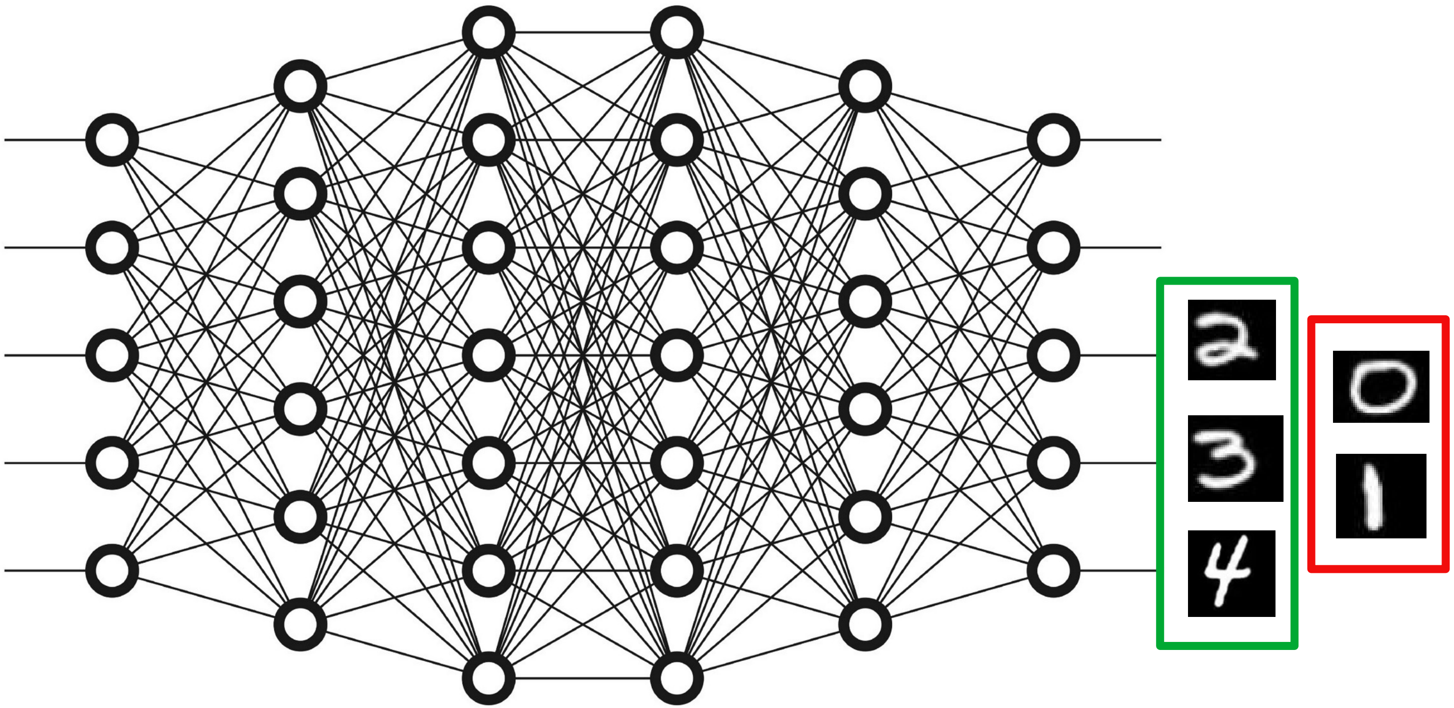
Training continually
first experience: classify digits 0, 1



Training continually
second (experience): classify digits 2, 3, 4



Test time (evaluation)
measure performance on the entire test data



60% final test accuracy

Catastrophic Forgetting

- Performance deteriorates on previous tasks once training on new information
- **Stability-Plasticity** dilemma:
  - Model needs plasticity to acquire new knowledge
  - Model needs plasticity to avoid forgetting previous information
- Modern neural networks are all towards **plasticity!**

The Continual Learning challenge

"A Continual (lifelong) learning system is defined as an **adaptive** algorithm capable of learning from a **continuous stream** of information, with such information becoming **progressively available** over time and where the **number of tasks** to be learned [...] are not predefined.
Critically, the accommodation of new information should occur **without catastrophic forgetting** or interference"

Parisi et. al. "*Continual lifelong learning with neural networks: A review*", *Neural Networks, 2019*

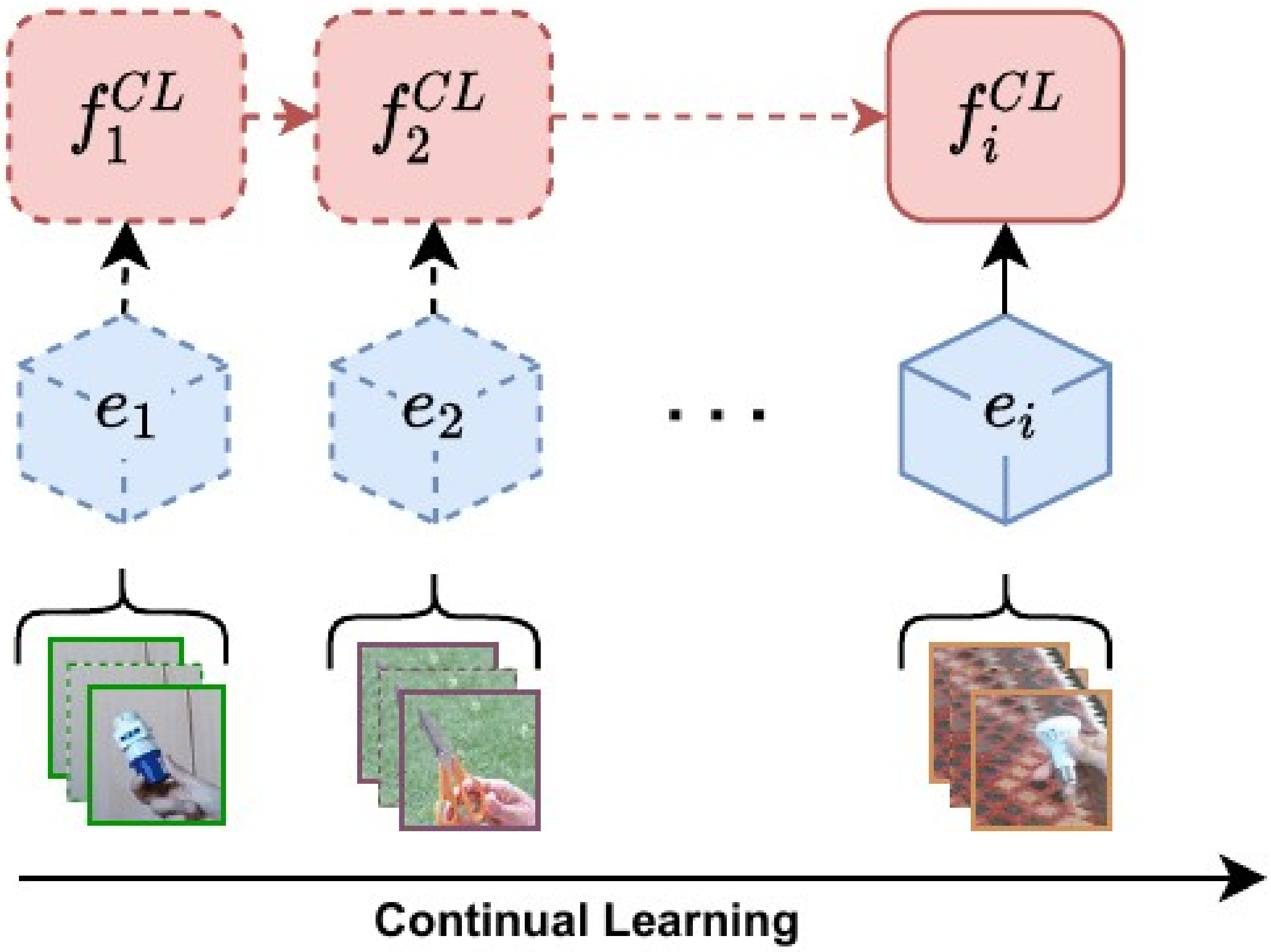
SCUOLA NORMALE SUPERIORE

UNIVERSITÀ DI PISA

The “grand” view of Continual Learning: Towards Sustainable Artificial Intelligence

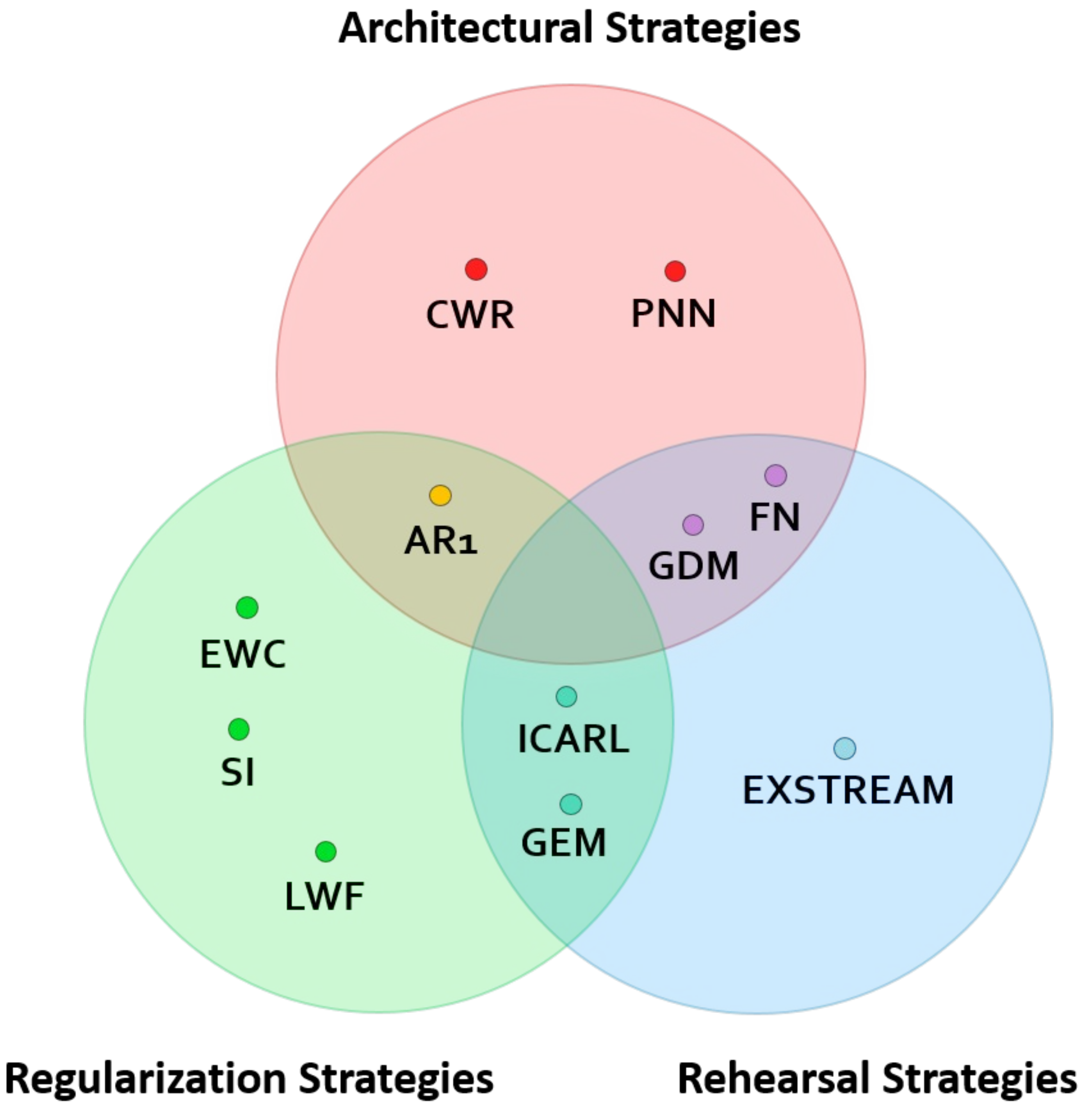
ContinualAI

- Keep the information inside the model up-to-date
  - No need to re-train from scratch every time (save CO2)
- No need for enormous amount of data all at once
  - build your dataset over time
- Work in resource constrained environment (e.g., edge computing and IoT)
- Support data privacy
- Fix machine bias (model “patches”)
- Prone to forgetting – extra computation to mitigate it
- Reduced predictive performance



Continual Learning: a stream of experiences

Lomonaco et. al., “Avalanche: an End-to-End Library for Continual Learning”, CLVision workshop, CVPR 2021



D. Maltoni and V. Lomonaco, “Continuous Learning in Single-Incremental-Task Scenarios”, Neural Networks

SCUOLA NORMALE SUPERIORE

UNIVERSITÀ DI PISA

ContinualAI

Continual Learning strategies: how to mitigate catastrophic forgetting

Evaluation of a CL agent

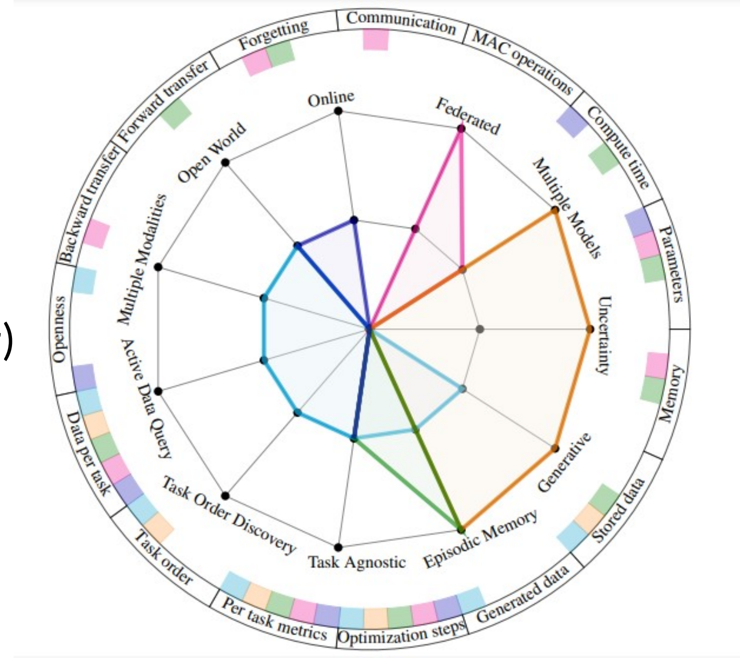
At the end of each experience, measure:

Performance metrics (e.g., accuracy)

- Accuracy on the current experience (training/validation)
- Accuracy on the future experiences (forward transfer)
- Accuracy on the past experiences (forgetting/backward transfer)

System / computational cost metrics

- Training time, memory consumption,

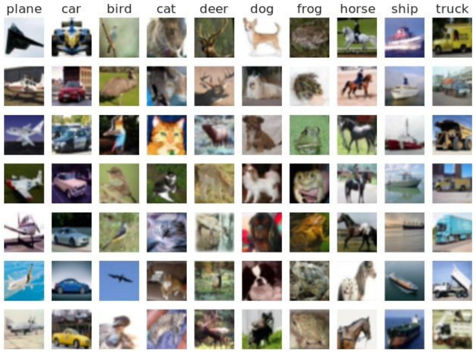


Mundt et. al., "CLEVA-Compass: A Continual Learning Evaluation Assessment Compass to Promote Research Transparency and Comparability", ICLR 2022

Continual Learning scenarios and benchmarks I

Dataset = MNIST, CIFAR-10/100, ImageNet, … (strong focus on Computer Vision)

- New Classes (NC) / Class-incremental
  - Split Dataset
- New Instances (NI) / Domain-incremental
  - Permuted Dataset, Rotated Dataset, …
  - CORe50 NI → new objects from different exposures



Continual Learning scenarios and benchmarks II

- Class-incremental with Repetition (CIR) / New Instances and Classes (NIC)
  - Towards a more “natural” way of learning: encounter both new classes and new/old instances
- Task-Free / Online / Streaming
  - See samples (few at a time) only once
  - no information about where they come from (as previous scenario, usually)
  - no information about task boundaries
  - Usually built out of NC scenario

References and credits

- **References**

G. I. Parisi, R. Kemker, J. L. Part, C. Kanan, and S. Wermter, “Continual lifelong learning with neural networks: A review”, Neural Networks, vol. 113, pp. 54–71, May 2019, doi: 10.1016/j.neunet.2019.01.012

A. Cossu, M. Ziosi, and V. Lomonaco, “Sustainable Artificial Intelligence through Continual Learning”, Proceedings of the 1st International Conference on AI for People: Towards Sustainable AI, 2021. https://eudl.eu/doi/10.4108/eai.20-11-2021.2314097

A. Pesaranghader, H. L. Viktor, and E. Paquet, “McDiarmid Drift Detection Methods for Evolving Data Streams”, in 2018 International Joint Conference on Neural Networks (IJCNN), pp. 1–9. doi: 10.1109/IJCNN.2018.8489260.

D. Maltoni and V. Lomonaco, “Continuous Learning in Single-Incremental-Task Scenarios”, Neural Networks, vol. 116, pp. 56–73, 2019, doi: 10/gk79q3

V. Lomonaco, L. Pellegrini, A. Cossu, A. Carta, G. Graffieti, T. L. Hayes, M. De Lange, M. Masana, J. Pomponi, G. M. van de Ven, M. Mundt, Q. She, K. Cooper, J. Forest, E. Belouadah, S. Calderara, G. I. Parisi, F. Cuzzolin, A. S. Tolias, S. Scardapane, L. Antiga, S. Ahmad, A. Popescu, C. Kanan, J. van de Weijer, T. Tuytelaars, D. Bacciu, and D. Maltoni, “Avalanche: an End-to-End Library for Continual Learning”, in 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Jun. 2021, pp. 3595–3605. doi: 10.1109/CVPRW53098.2021.00399.

- **Image credits**

Slide 1.
https://miro.medium.com/max/1024/1*LsMzwScAxN-0b3kOc2eMxw.jpeg
Slide 2.
https://static.thenounproject.com/png/1551919-200.png (neural network)
http://wfarm3.dataknet.com/static/resources/icons/set113/ee242415.png (data)
https://openclipart.org/image/2400px/svg_to_png/110/molumen-red-square-error-warning-icon.png (error)
https://www.iconexperience.com/_img/g_collection_png/standard/512x512/function.png (function)
Slide 7-10:
https://iq.opengenus.org/content/images/2018/10/ANN1.jpg
Slide 11.
https://alzheimergadfly.net/wp-content/uploads/2018/07/Brain-eraser-e1533486845740.png

Time to code with Avalanche!

https://avalanche.continualai.org/

Avalanche *powered by* ContinualAI

But first, ~~let me take a selfie~~ any questions?