# Streaming Data Science


## Emanuele Della Valle

**Prof @ PoliMI**
**CRO & founder @ Motus ml**
**founder @ Quantia Consulting**


POLITECNICO
MILANO 1863


# Society as a whole is undergoing a data-centered revolution

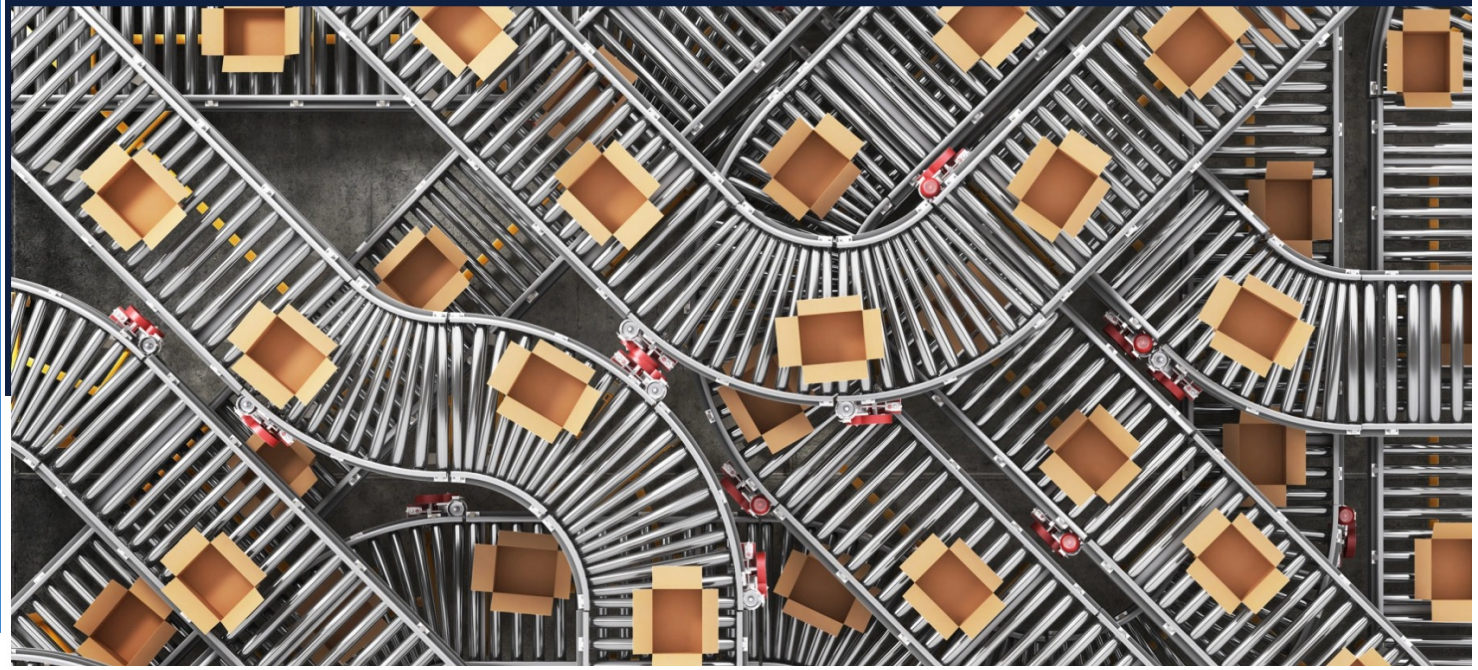
# But it's a world characterized by

ARTIFICIAL INTELLIGENCE

# Our weird behavior during the pandemic is messing with AI models

Machine-learning models trained on normal behavior are showing cracks — forcing humans to step in to set them straight.

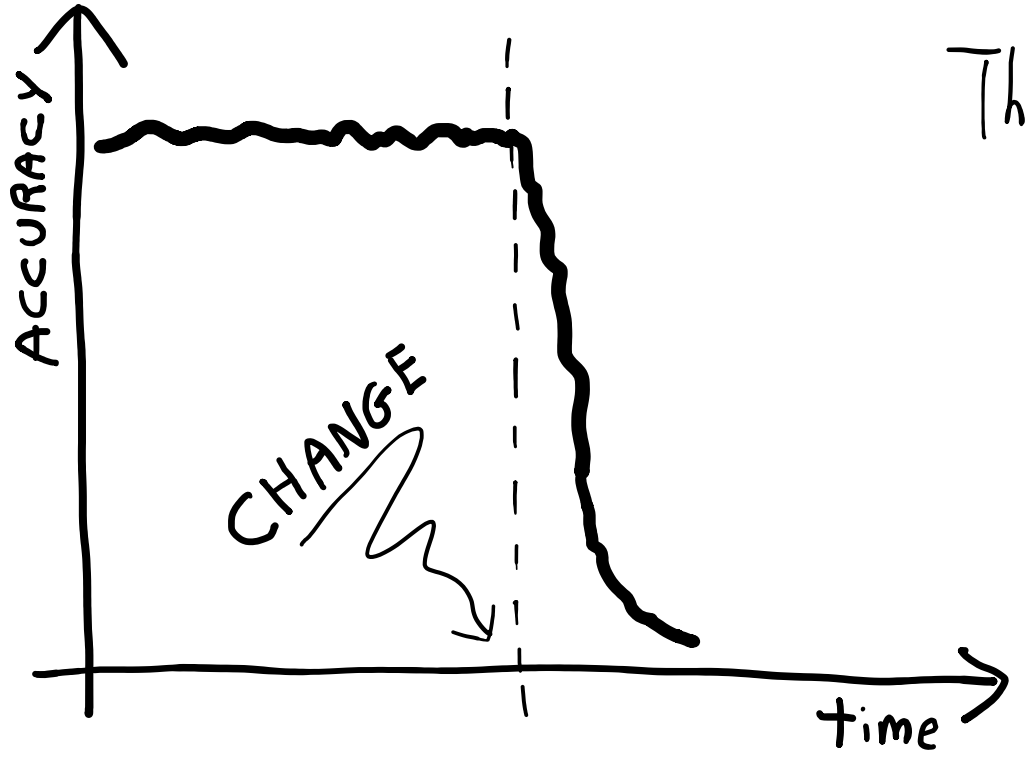
**By Will Douglas Heaven** May 11, 2020



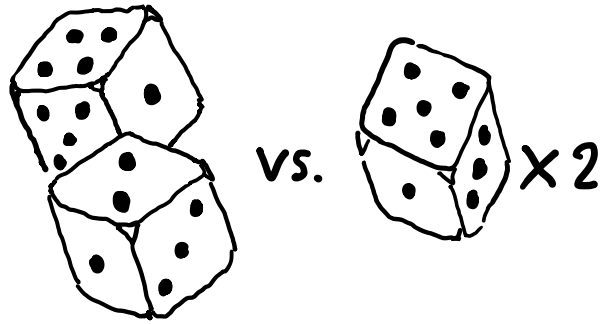
[src: https://www.technologyreview.com/2020/05/11/1001563/covid-pandemic-broken-ai-machine-learning-amazon-retail-fraud- humans-in-the-loop/]

POLITECNICO MILANO 1863

# Changes cause AI models to lose relevancy



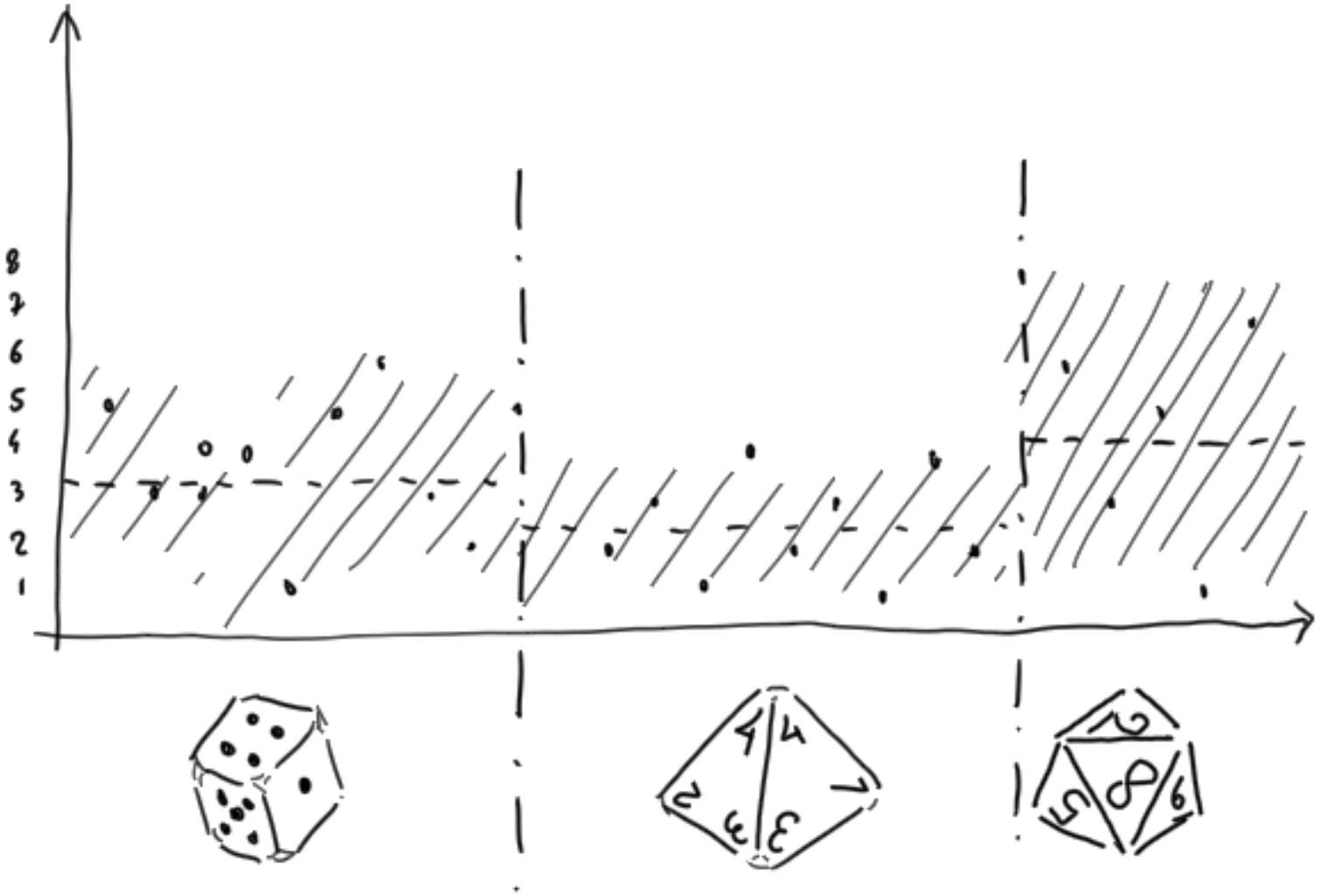

The **assumption**: data is **i**ndependent & **i**dentically **d**istributed




It does **not** hold

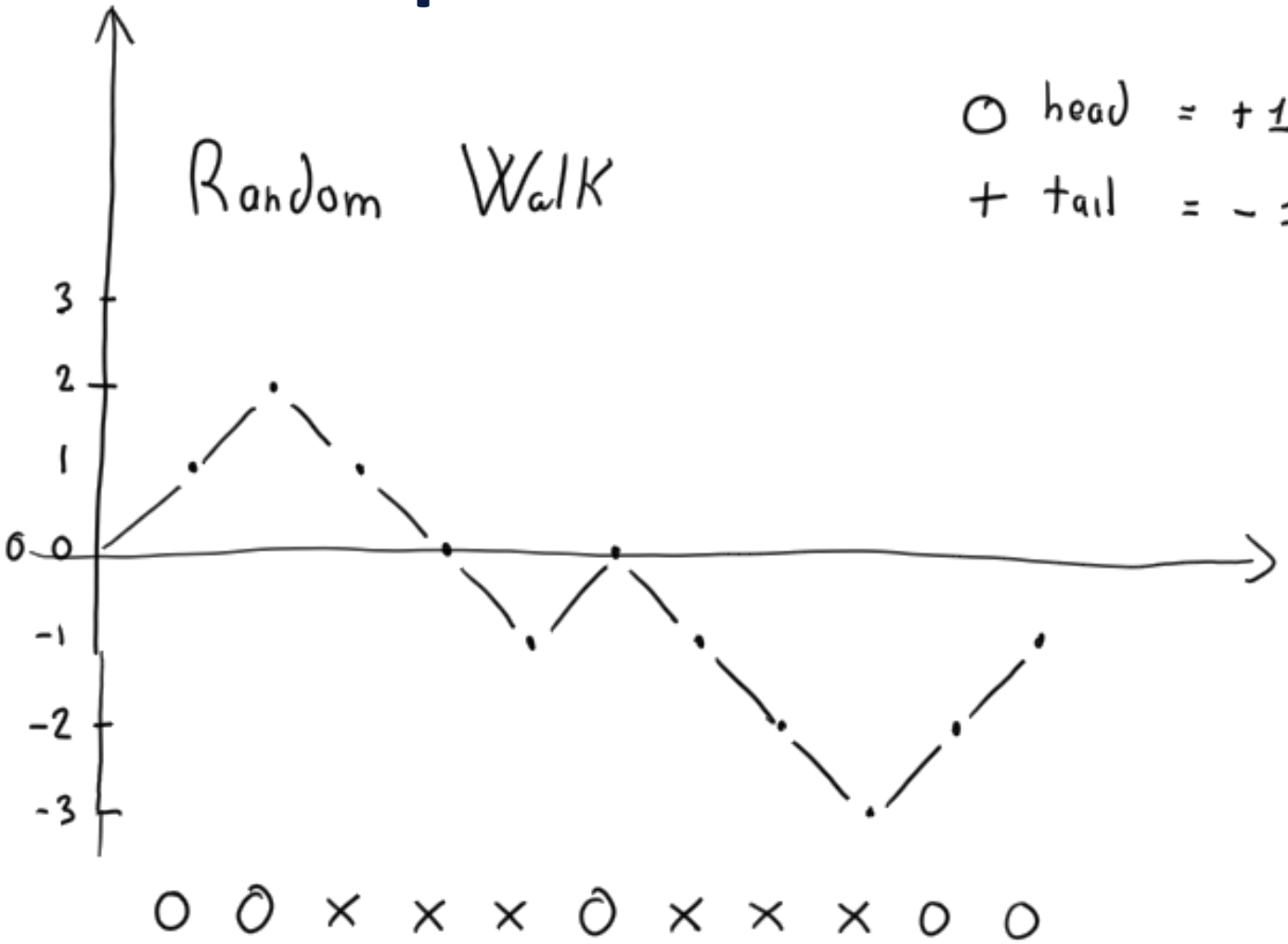
## i.i.d. assumption explained

# Non identically distributed data



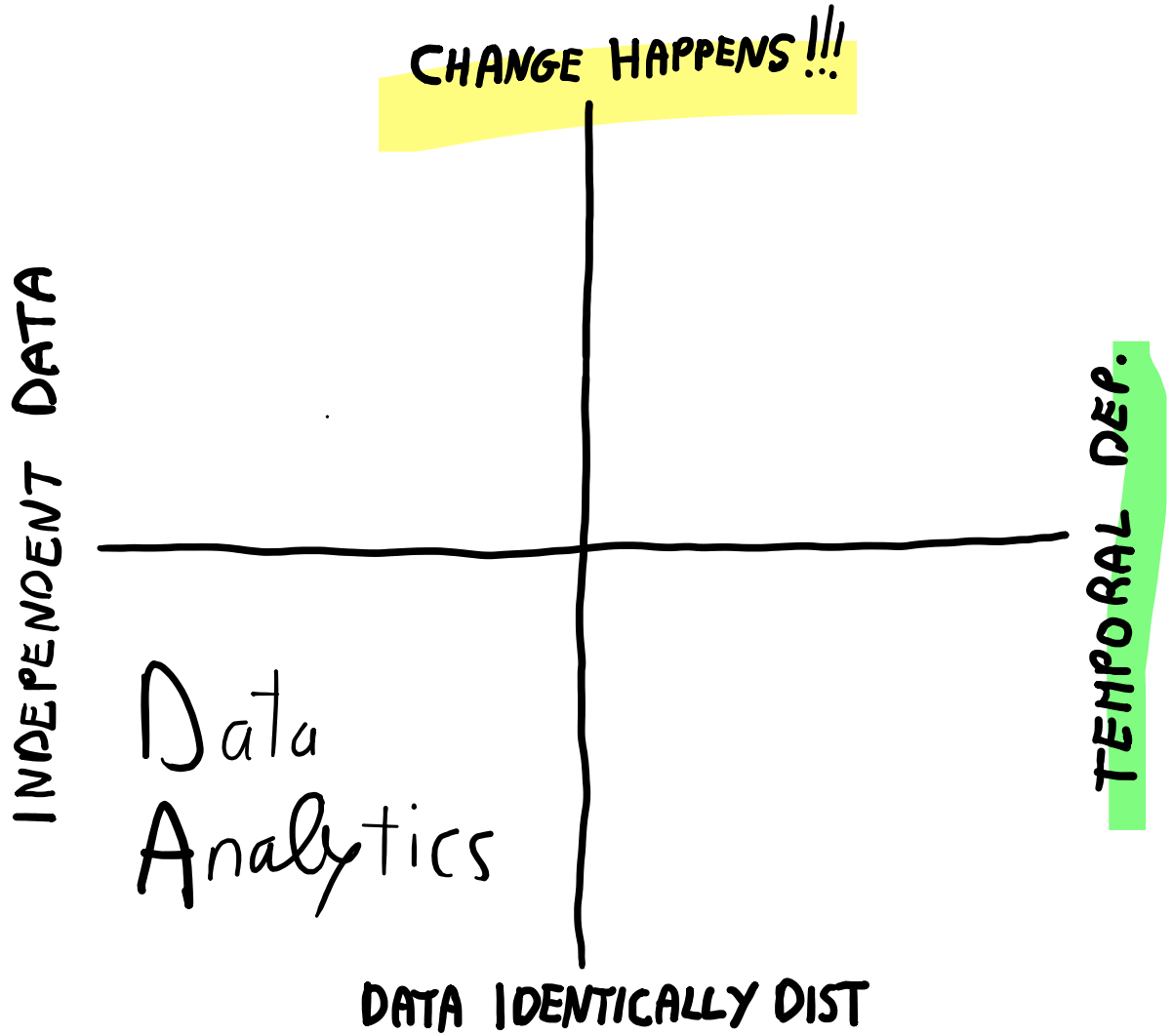

## i.i.d. assumption explained

# Non independent data

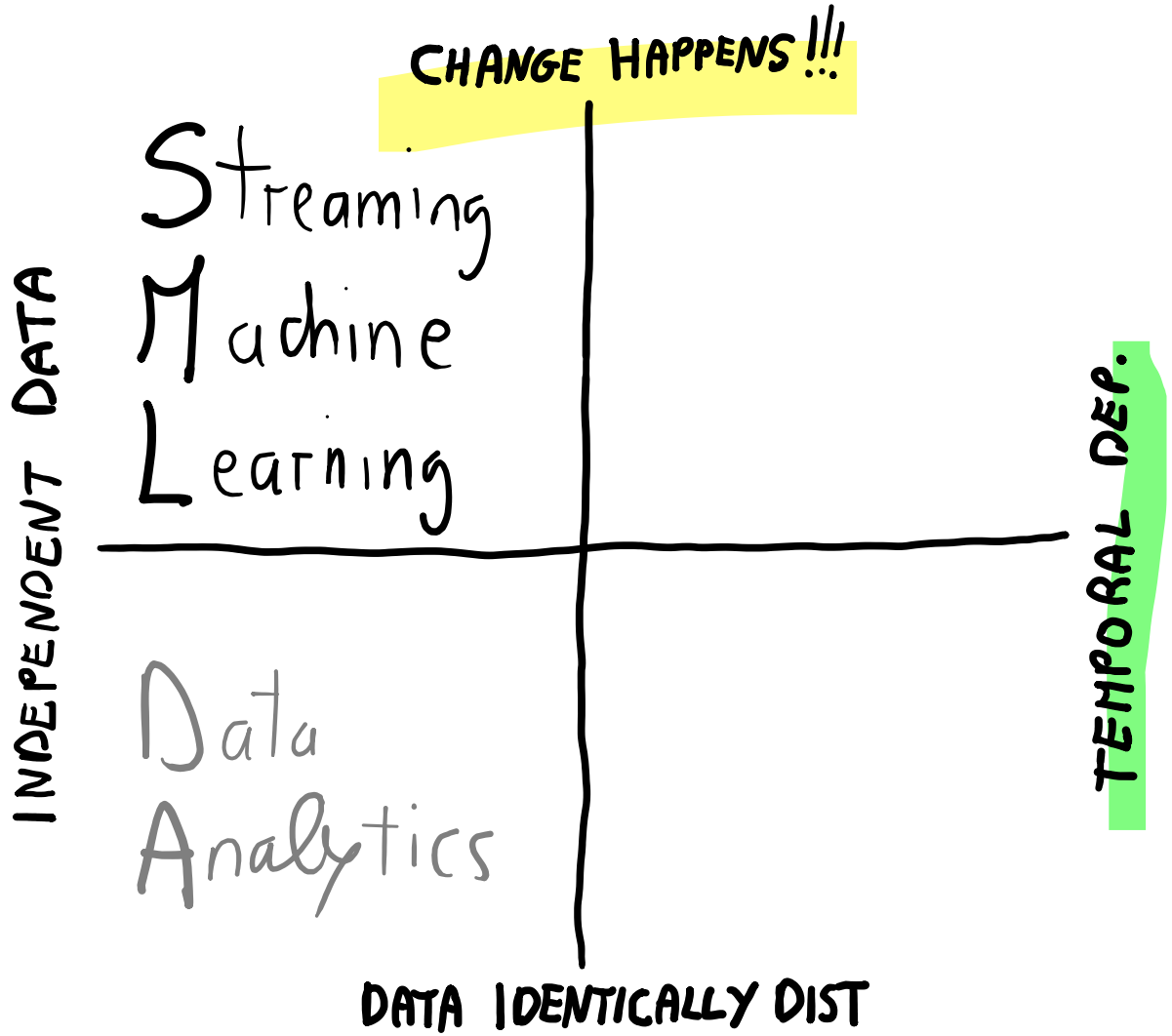


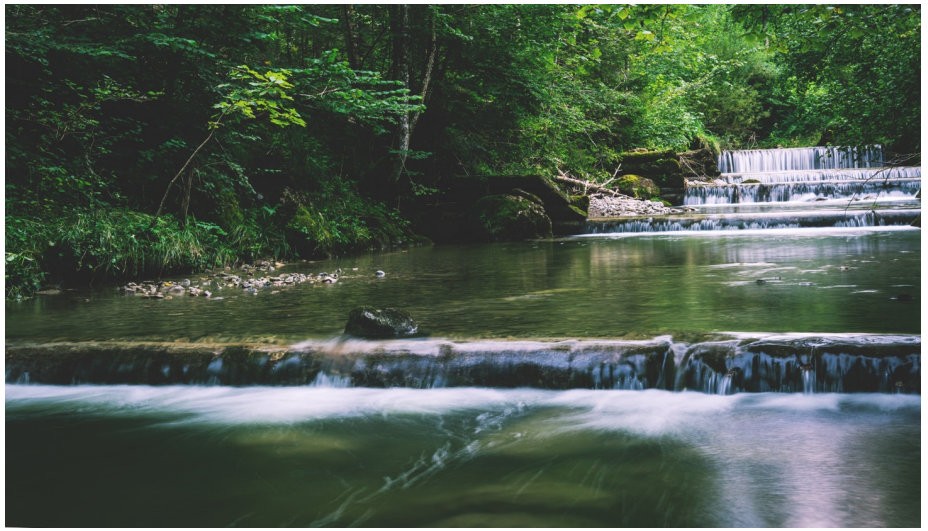

TEMPORAL DEP.

# State-of-the-art

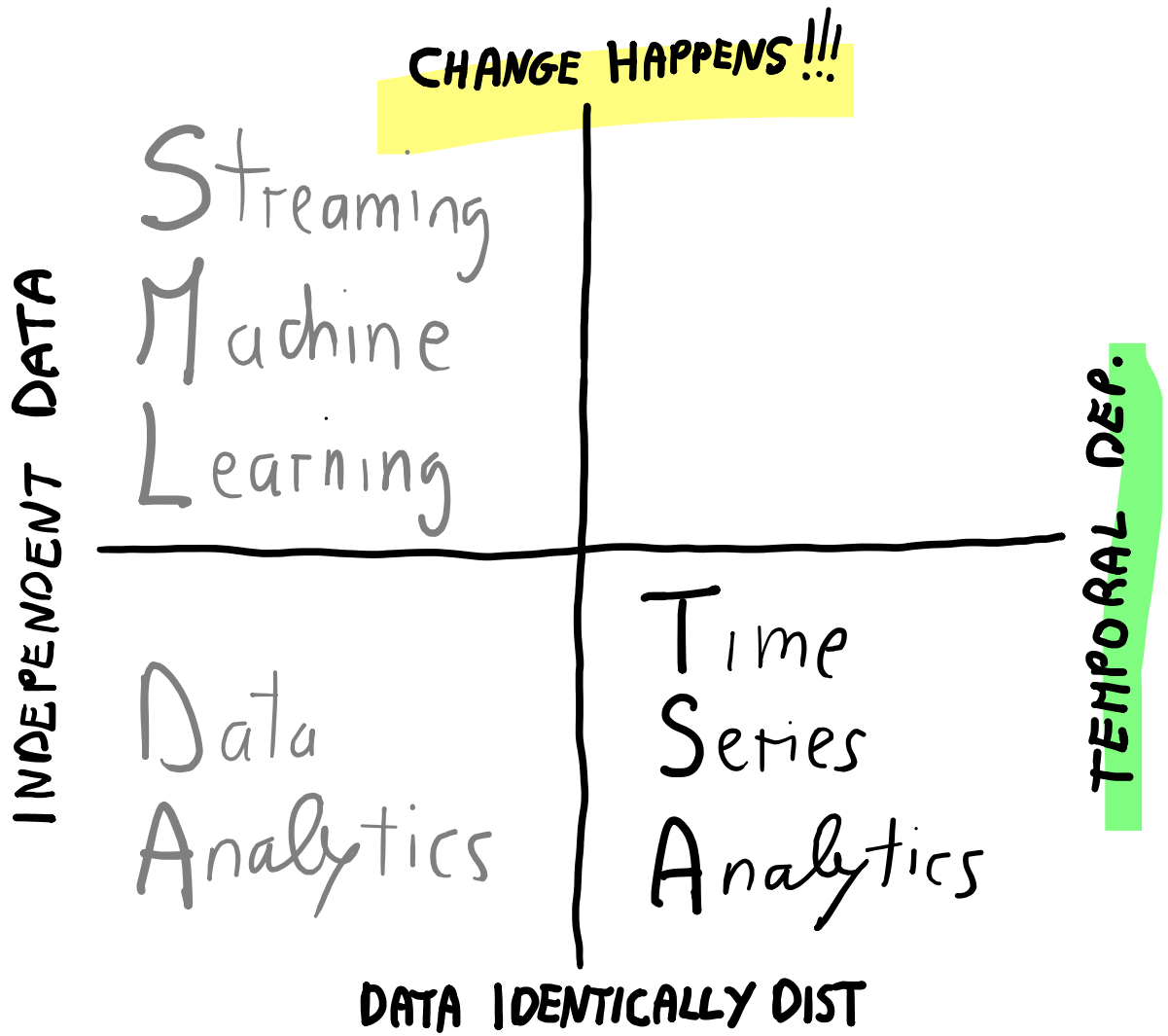


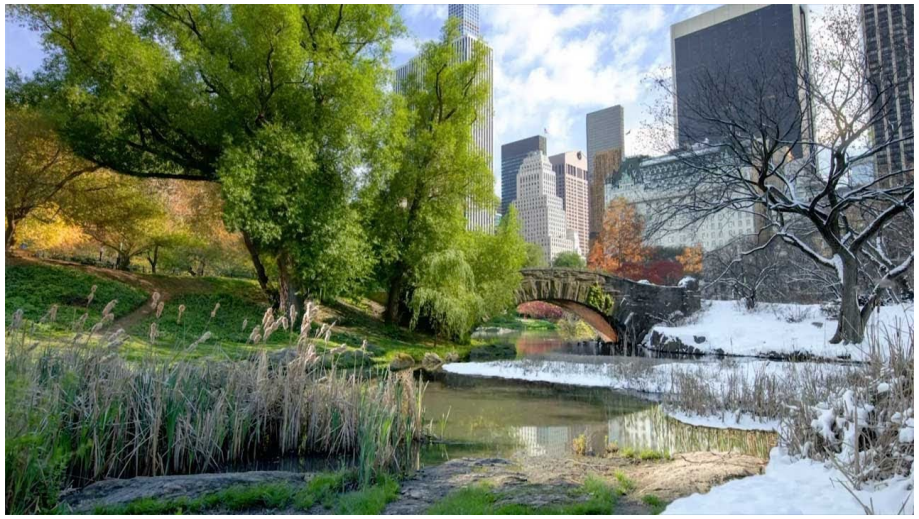
# State-of-the-art



# State-of-the-art




Explaining the past and forecast the future
of a continuous flow of data
without assuming data independence
with
**Time Series Analytics**

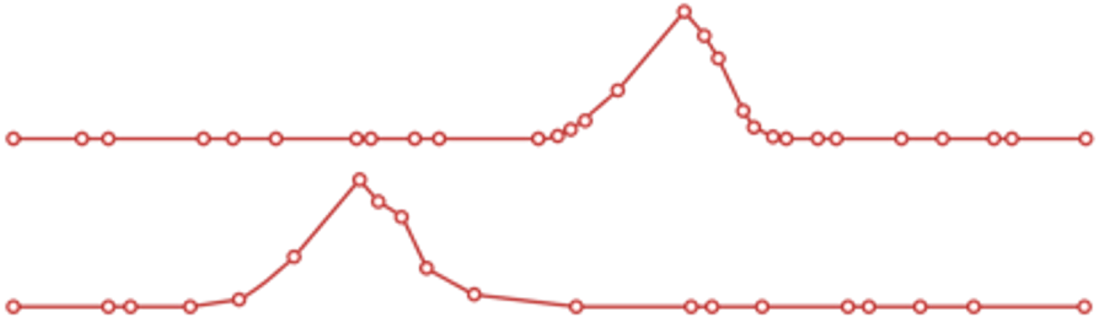
# Type of data

A ***time series*** is a sequence of observations on **one** (or more) **quantitative** variable **regularly** collected **over time**.

# Events vs time series

The phenomenon happens and we observe them

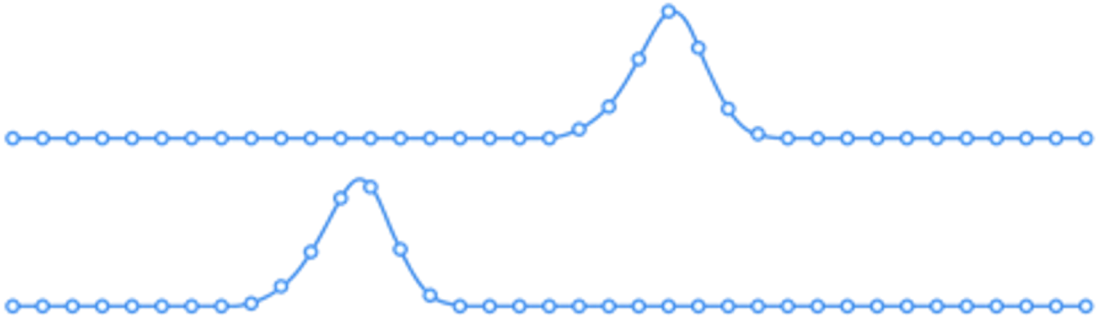
**irregular**ly



**Events**

We monitor a phenomenon

**regular**ly



**Time series**

# Events becomes time series via windowing

E.g.,

- The average trade price of Apple stock every 10 minutes over the course of a day
- The average response time for requests in an application over 1 minute intervals
- …

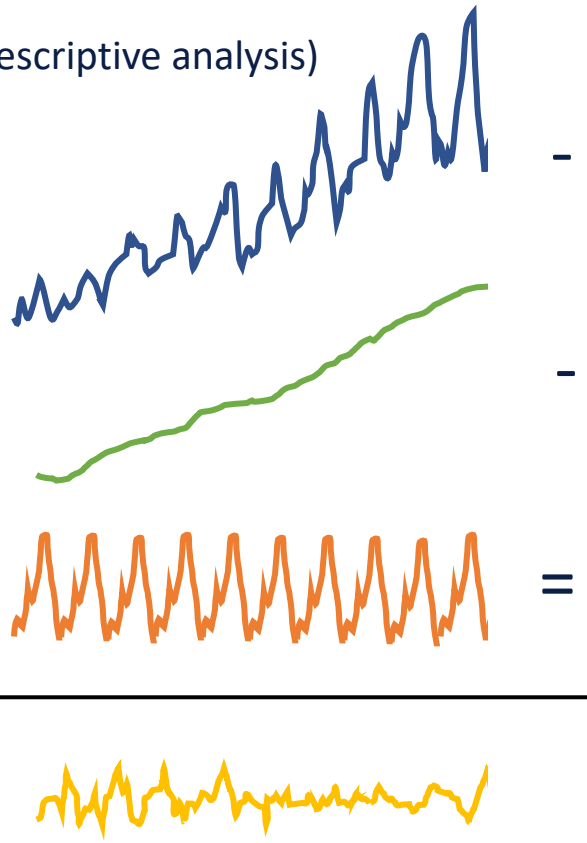
# Two purposes

**Modeling** (a.k.a., descriptive analysis)

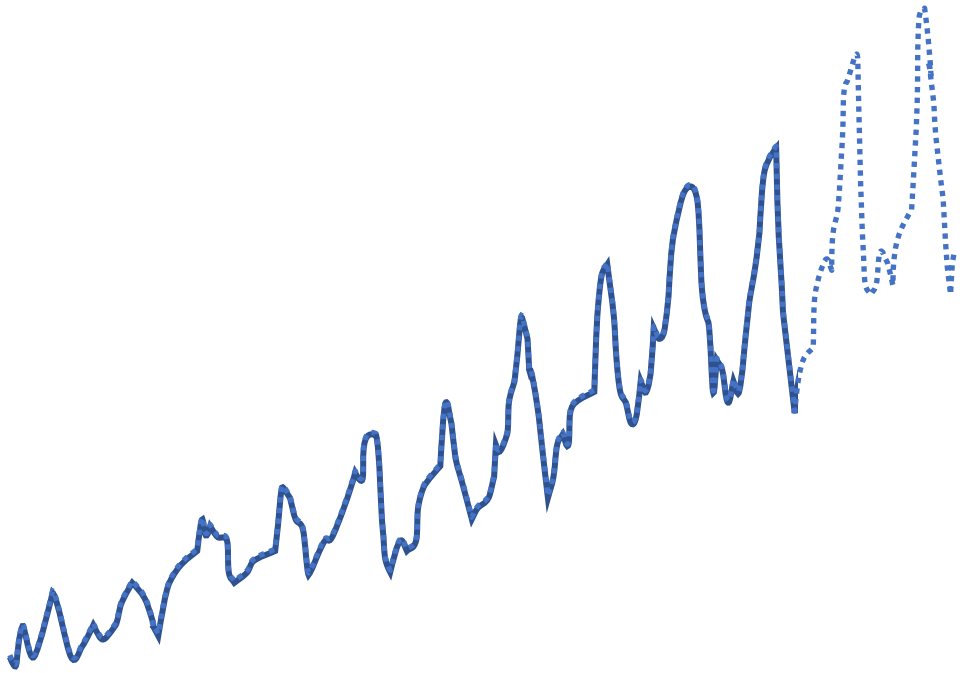
- Original
- Trend
- Seasonality

---

- Residual

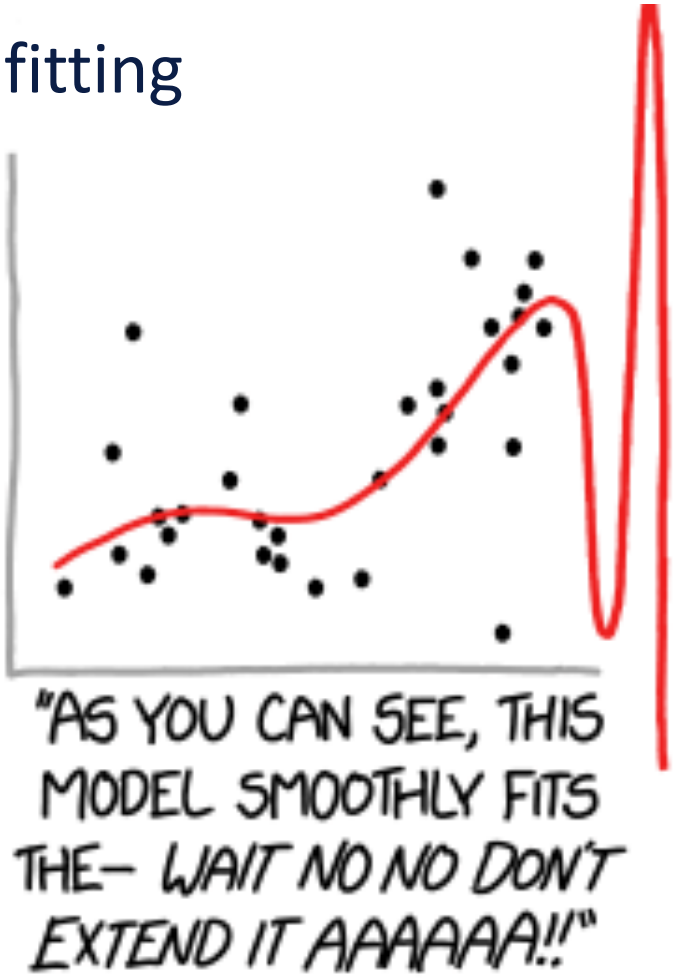


- **Forecasting** (a.k.a., predictive analysis)



# Traditional techniques
# Trend

Not fitting

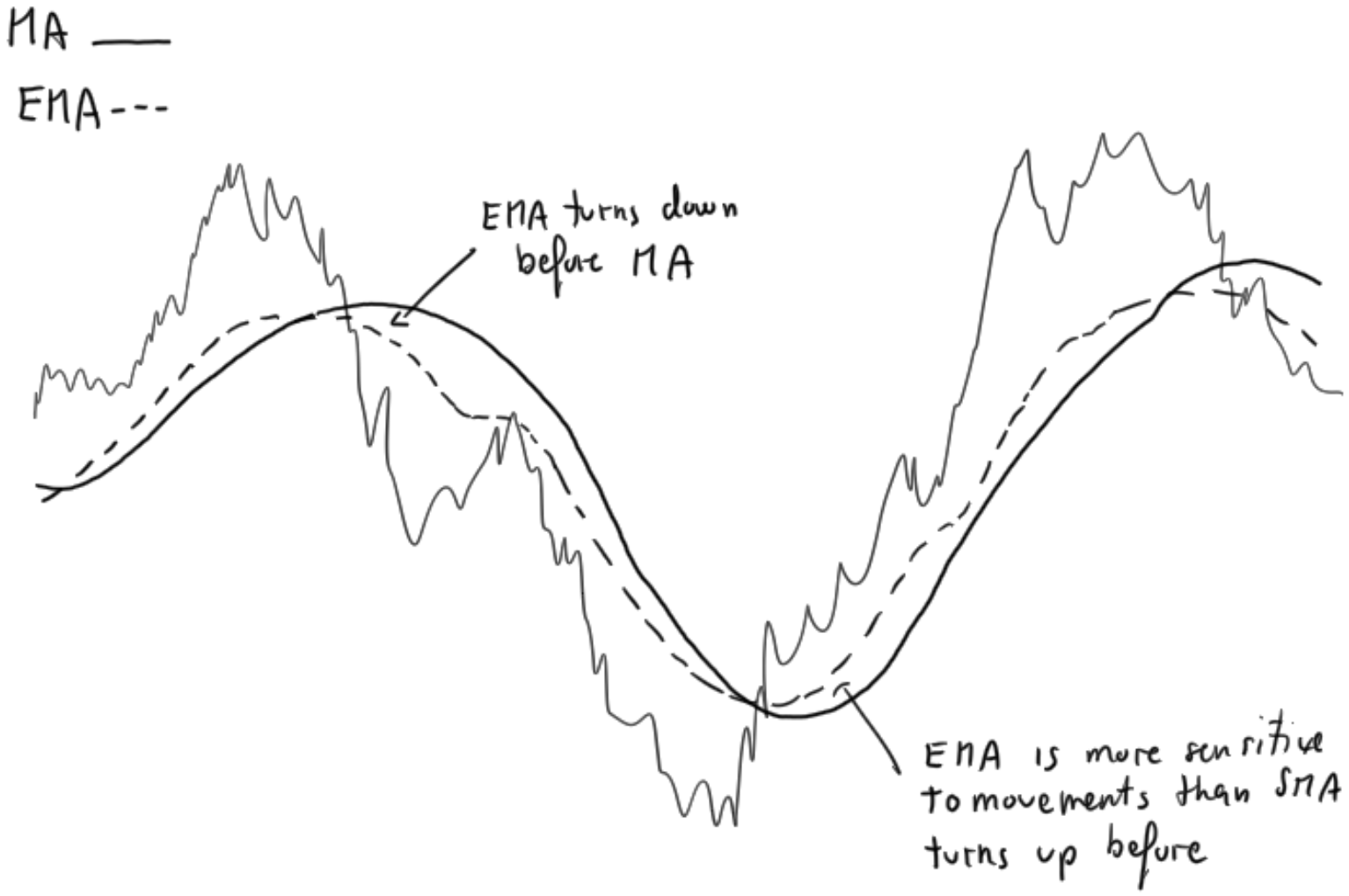



[src:; https://xkcd.com/2048/ ]

But **smoothing** or …

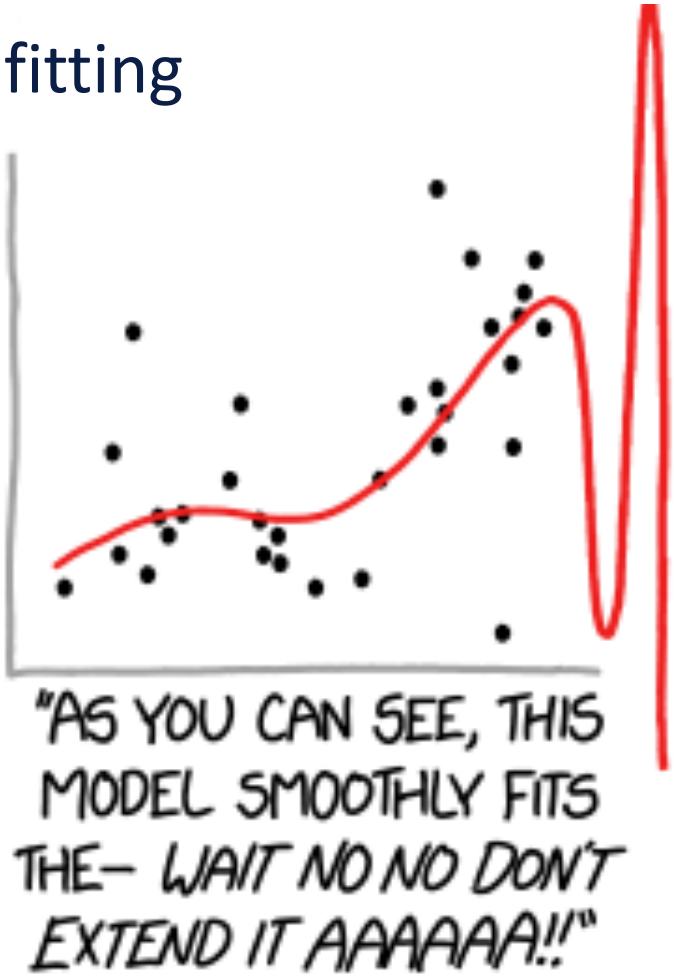
## Traditional techniques

# Capturing trend via smoothing



# Traditional techniques
# Trend

Not fitting

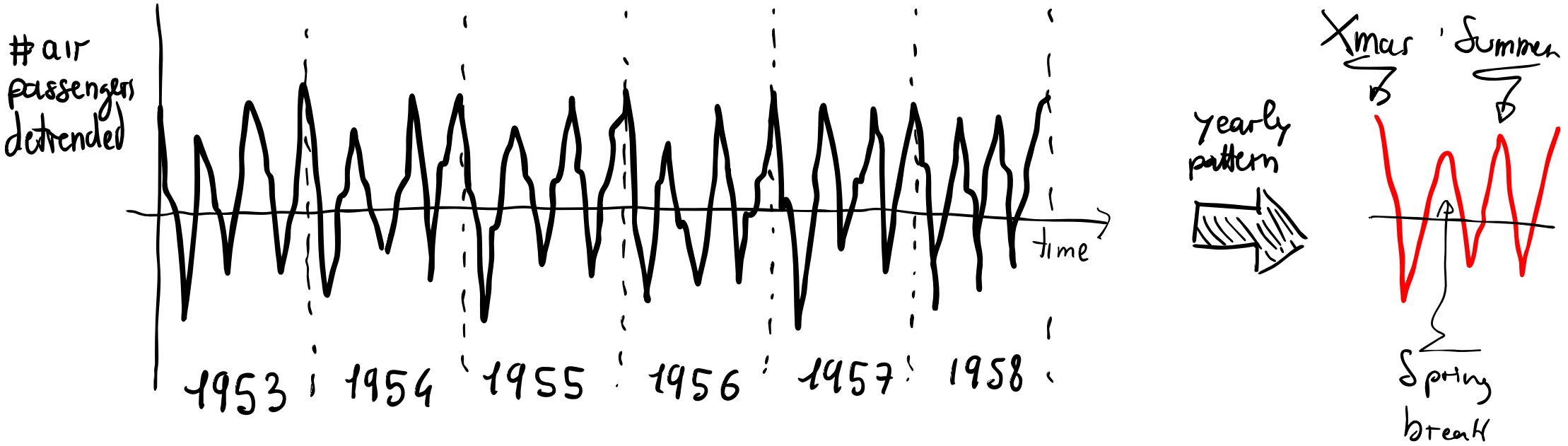



[src:; https://xkcd.com/2048/ ]

… but smoothing **or differencing**

# Traditional techniques
# Seasonality

- A repetitive pattern of a given duration, e.g., every year



## Traditional techniques

# Autoregressive moving-average (ARMA)

- Autoregressive

$$X_t = \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t$$

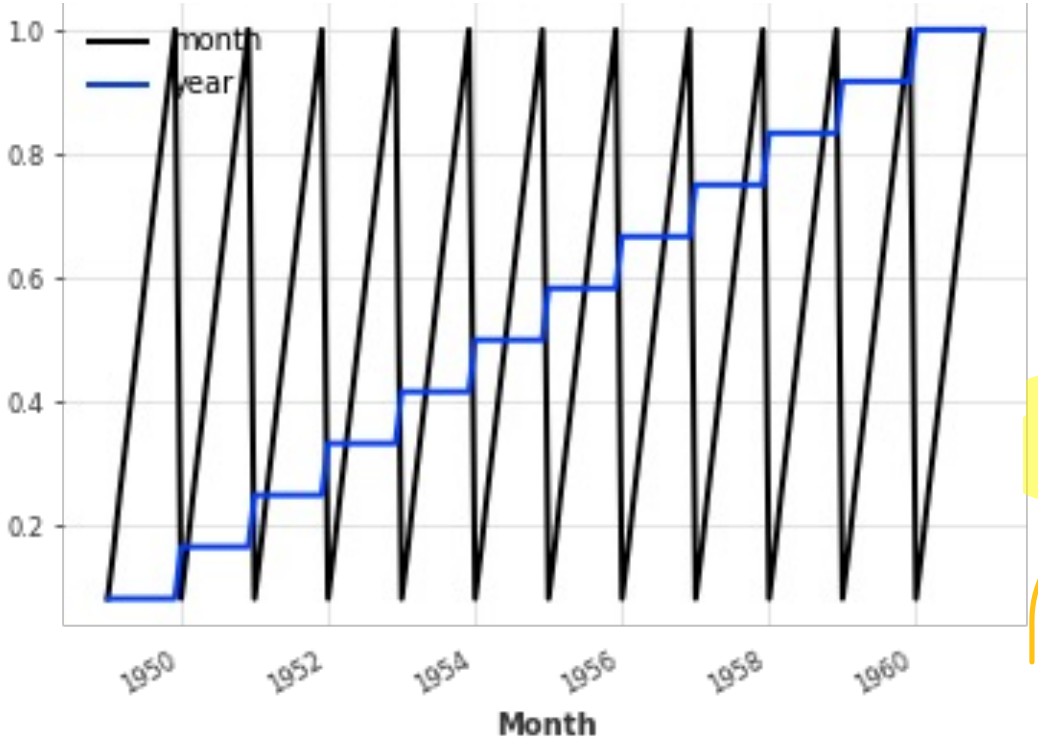
- Moving-average

$$X_t = \mu + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$$

- ARMA

$$X_t = \varepsilon_t + \sum_{i=1}^{p} \varphi_i X_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}.$$

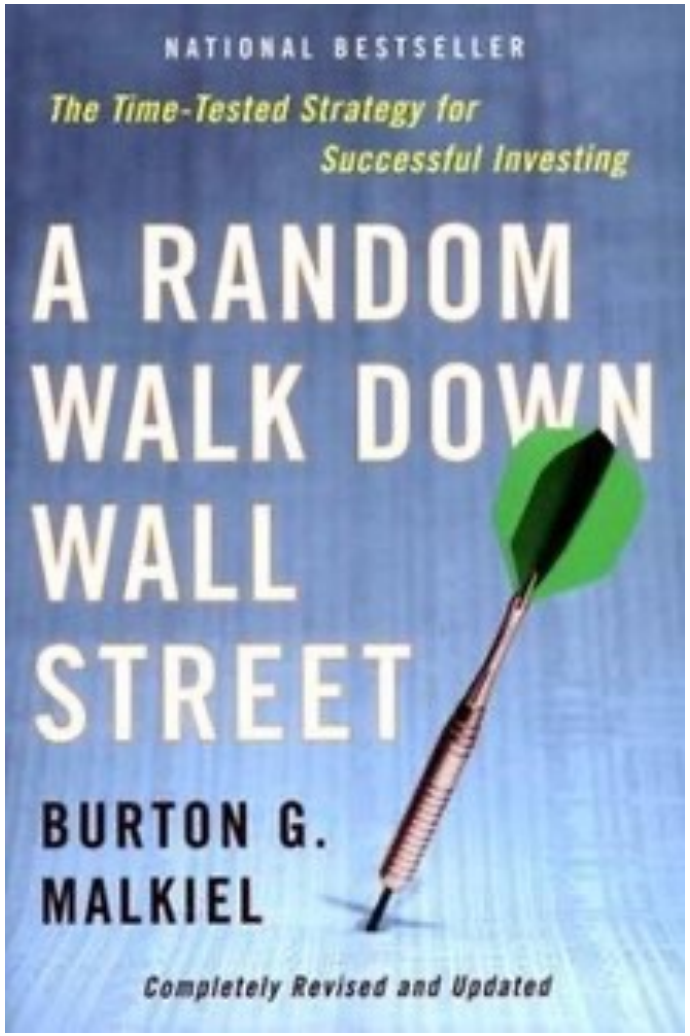
# Traditional techniques
# Using external data (covariates)




ARMAX: $X_t = \varepsilon_t + \sum_{i=1}^{p} \varphi_i X_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \sum_{i=1}^{b} \eta_i d_{t-i}.$

# Stock price forecasting

- The act of trying to determine the future value of a company stock (or other financial instrument) traded on an exchange
- To learn more https://en.wikipedia.org/wiki/Stock_market_prediction

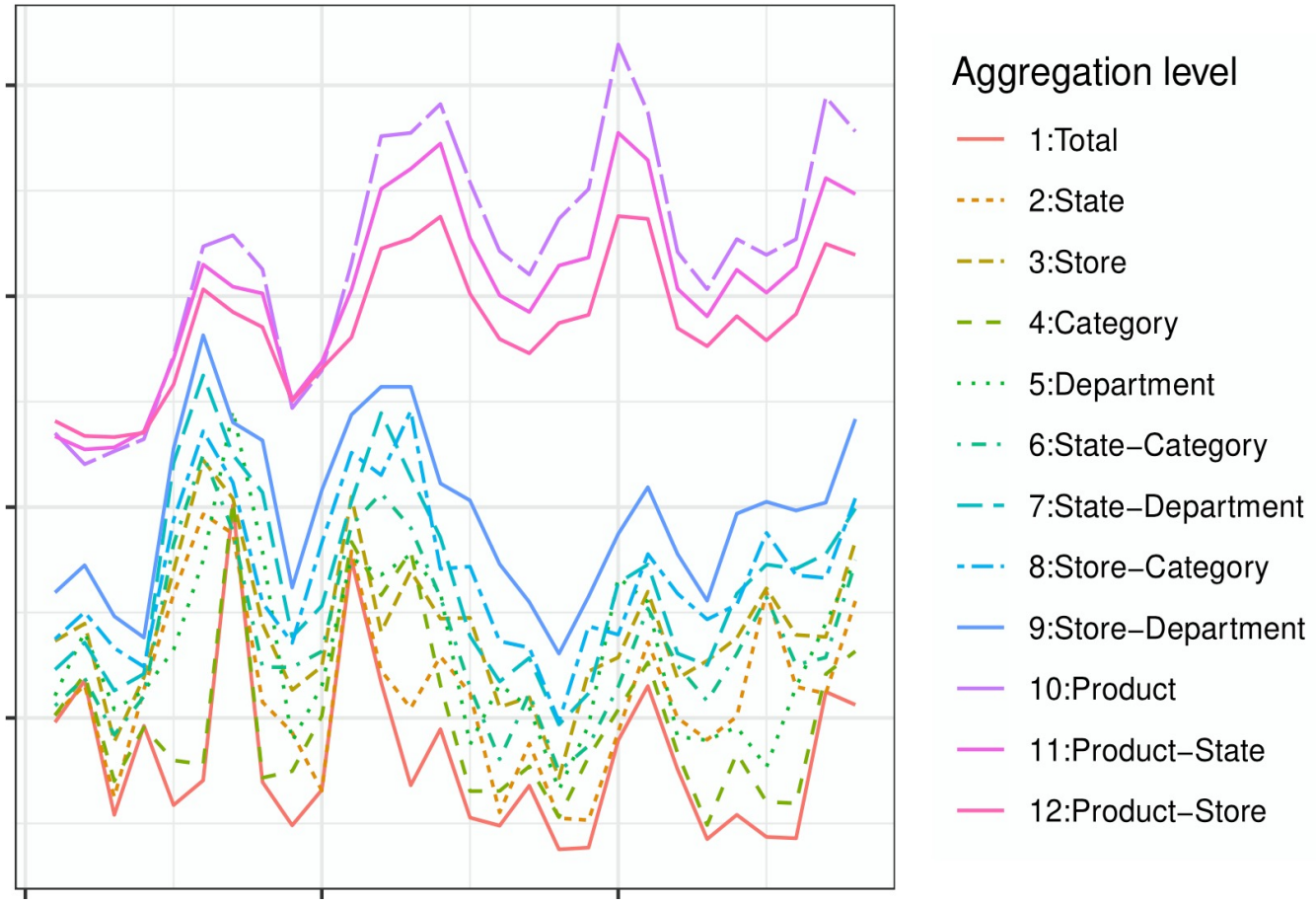


# Energy forecasting

- Get forecasts that reflect business realities (through repeatable, scalable, traceable, and defensible results), and plan future events (investments, trading, purchases …) with confidence
- E.g., SAS
  https://www.sas.com/en_us/software/energy-forecasting.html

# Unit sales forecasting for 1000s of products

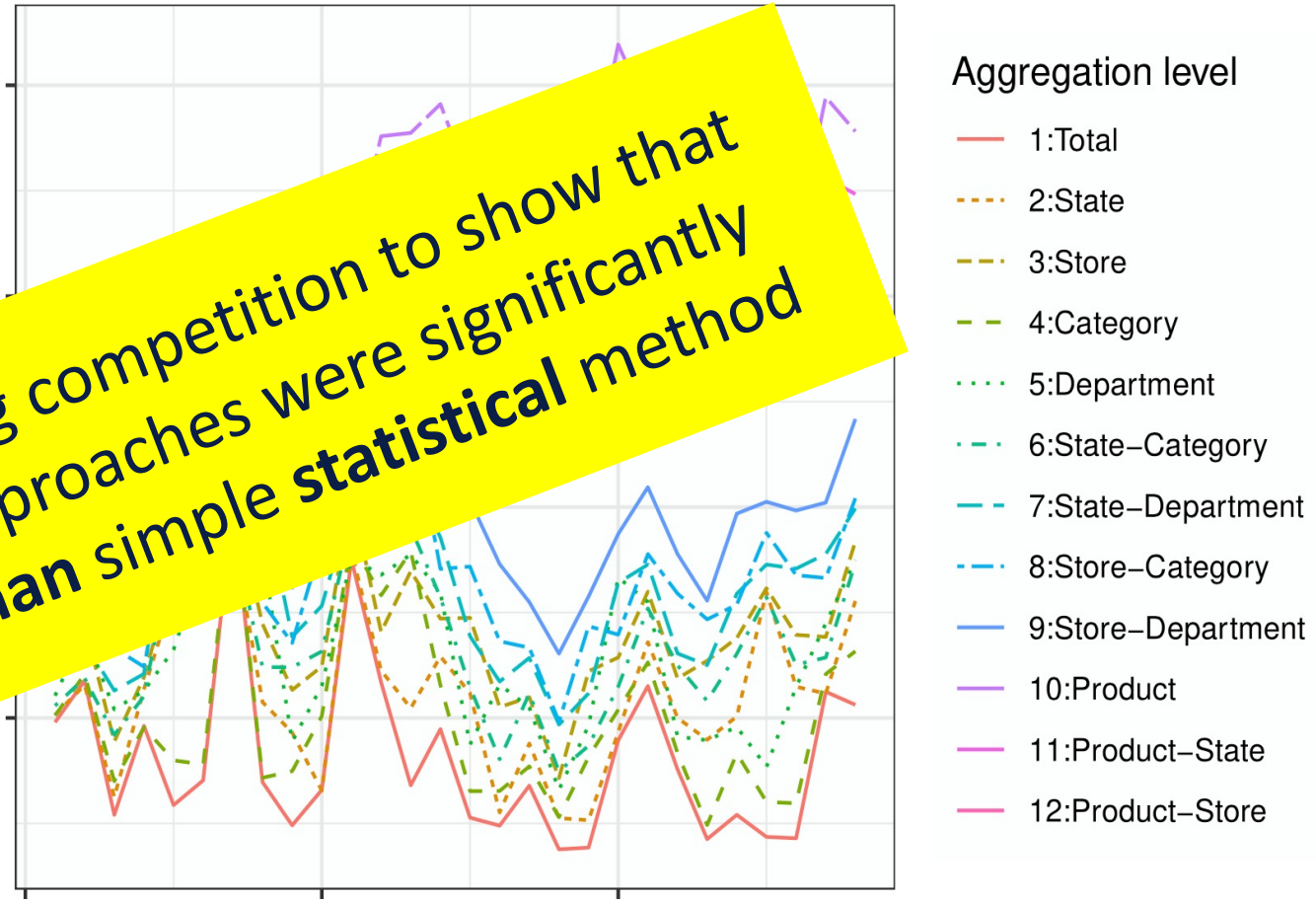
- Forecasting the unit sales of 3,049 Walmart products grouped by type (category and department) and selling location (stores and states).
- E.g., Makridakis's challenge https://mofc.unic.ac.cy/m5-competition/



# Unit sales forecasting for 1000s of products

- Forecasting the unit sales of 3,049 Walmart products grouped by type (category and department) and selling location (stores and state
- E.g., Makridakis'
  https://mofc

The first forecasting competition to show that two **ML**-based approaches were significantly **more accurate than** simple **statistical** method



# Advanced techniques
# Specialized Neural Network

POLITECNICO MILANO 1863

[src: https://arxiv.org/pdf/1905.10437.pdf ]

## N-BEATS: NEURAL BASIS EXPANSION ANALYSIS FOR INTERPRETABLE TIME SERIES FORECASTING


**Boris N. Oreshkin**
Element AI
boris.oreshkin@gmail.com

**Dmitri Carpov**
Element AI
dmitri.carpov@elementai.com

**Nicolas Chapados**
Element AI
chapados@elementai.com

**Yoshua Bengio**
Mila
yoshua.bengio@mila.quebec



ABSTRACT

We focus on solving the univariate times series point forecasting problem using deep learning. We propose a deep neural architecture based on backward and forward residual links and a very deep stack of fully-connected layers. The architecture has a number of desirable properties, being interpretable, applicable without modification to a wide array of target domains, and fast to train. We test the proposed architecture on several well-known datasets, including M3, M4 and TOURISM competition datasets containing time series from diverse domains. We demonstrate state-of-the-art performance for two configurations of N-BEATS for


## DeepAR: Probabilistic Forecasting with Autoregressive Recurrent Networks


**David Salinas, Valentin Flunkert, Jan Gasthaus**
Amazon Research
Germany
<dsalina,flunkert,gasthaus@amazon.com>



**Abstract**

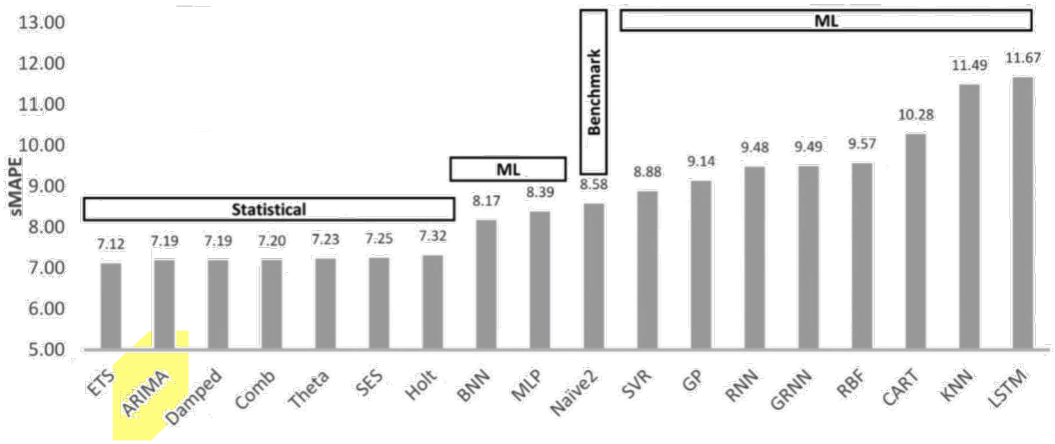
Probabilistic forecasting, i.e. estimating the probability distribution of a time series' future given its past, is a key enabler for optimizing business processes. In retail businesses, for example, forecasting demand is crucial for having the right inventory available at the right time at the right place. In this paper we propose DeepAR, a methodology for producing accurate probabilistic forecasts, based on training an auto-regressive recurrent network model on a large number of related time series. We demonstrate how by applying deep learning techniques to forecasting, one can overcome many of the challenges faced by widely-used classical approaches to the problem. We show through extensive empirical evaluation on several real-world forecasting data sets accuracy improvements of around 15%


[src: https://arxiv.org/pdf/1704.04110.pdf ]

# Be aware of trade-offs!

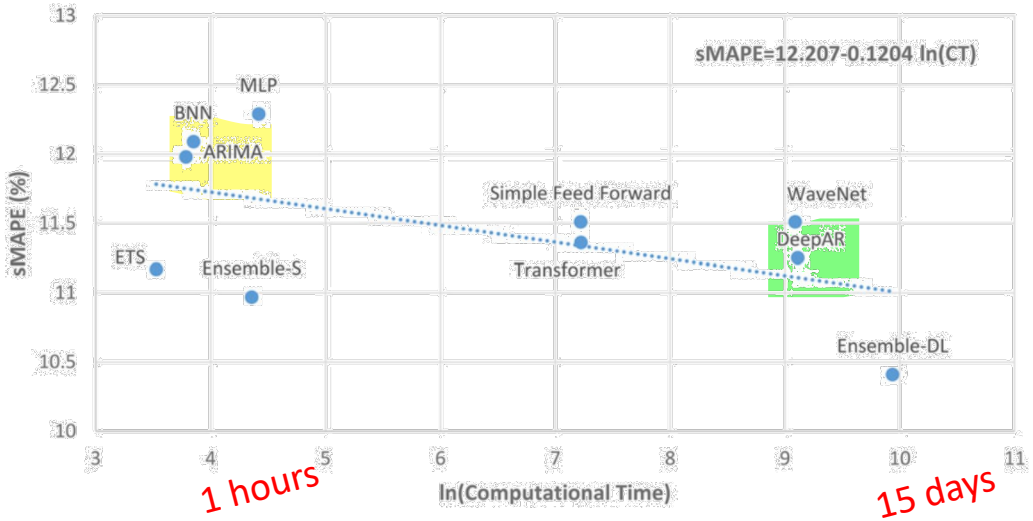
## 2018

**Statistical** methods largely outperform ML-based methods




## 2022

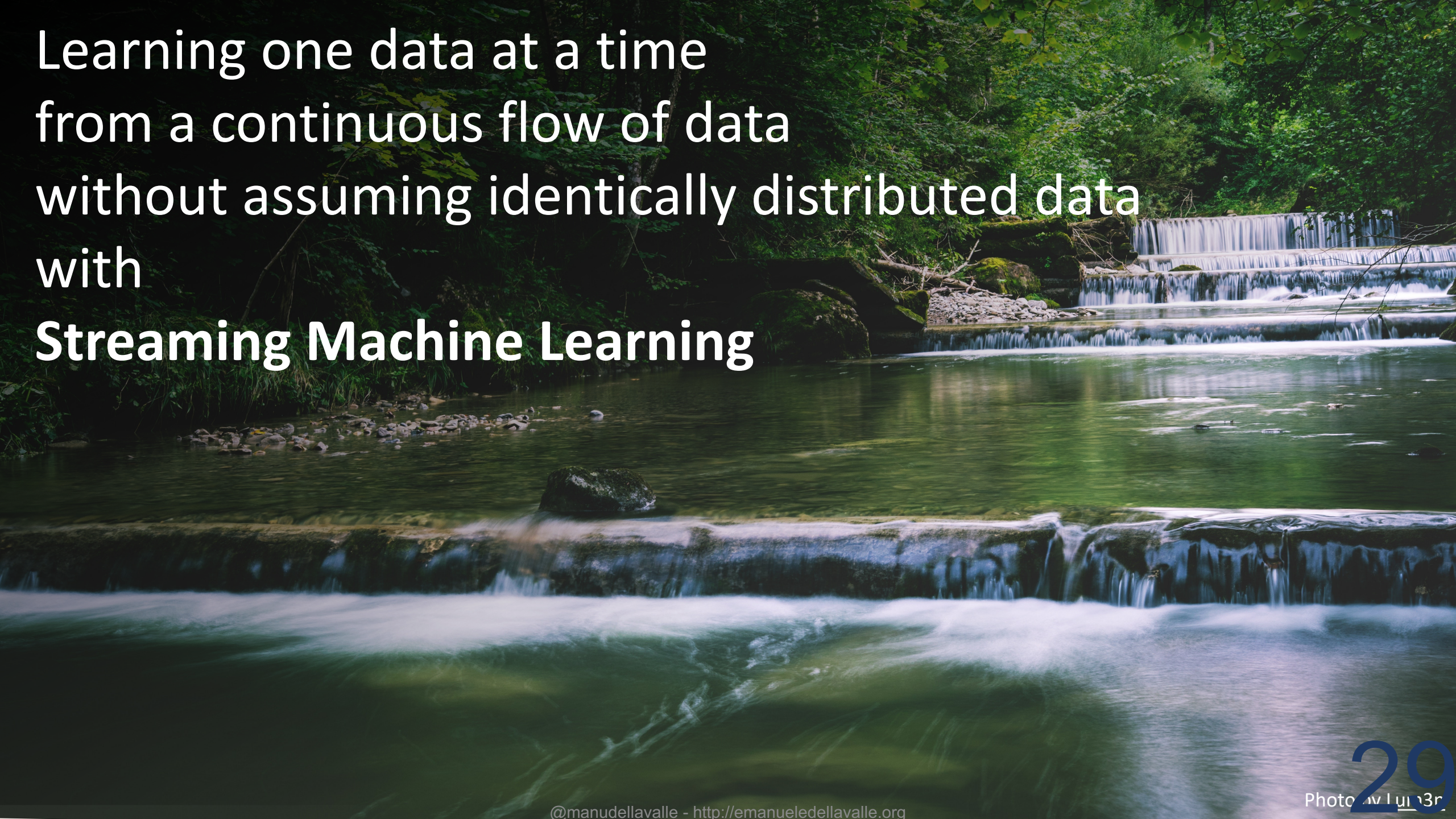
**ML-based** methods (i.e., deep learning ones) outperform statistical methods *but* at a vast *computational cost*




[src: https://www.tandfonline.com/doi/epdf/10.1080/01605682.2022.2118629 ]

# Learning Characteristics

- Time series analytics is not inherently adaptive nor focused on real-time learning
- Models are (almost always) retrained from scratch as new data points are received.


Learning one data at a time
from a continuous flow of data
without assuming identically distributed data
with
**Streaming Machine Learning**

# Type of data

data arrives in a **continuous stream**, potentially in real-time, assuming **data** points are **independent but** data stream can evolve over time (i.e., the data are **not identically distributed**)

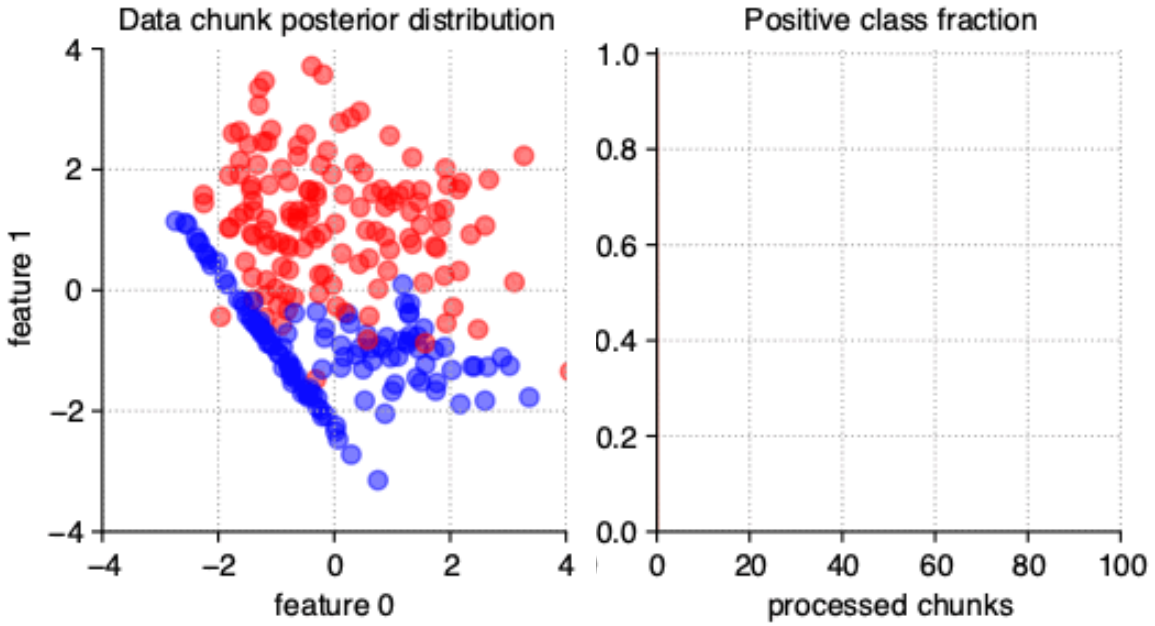
# Purpose

- extract **insights** from data **as** it arrives and discard the data
- often with
  - a low-latency requirement
  - a minimal computational footprint
- classification is the main application
- forecasting and clustering are also in scope


[src: http://www.motusml.com ]

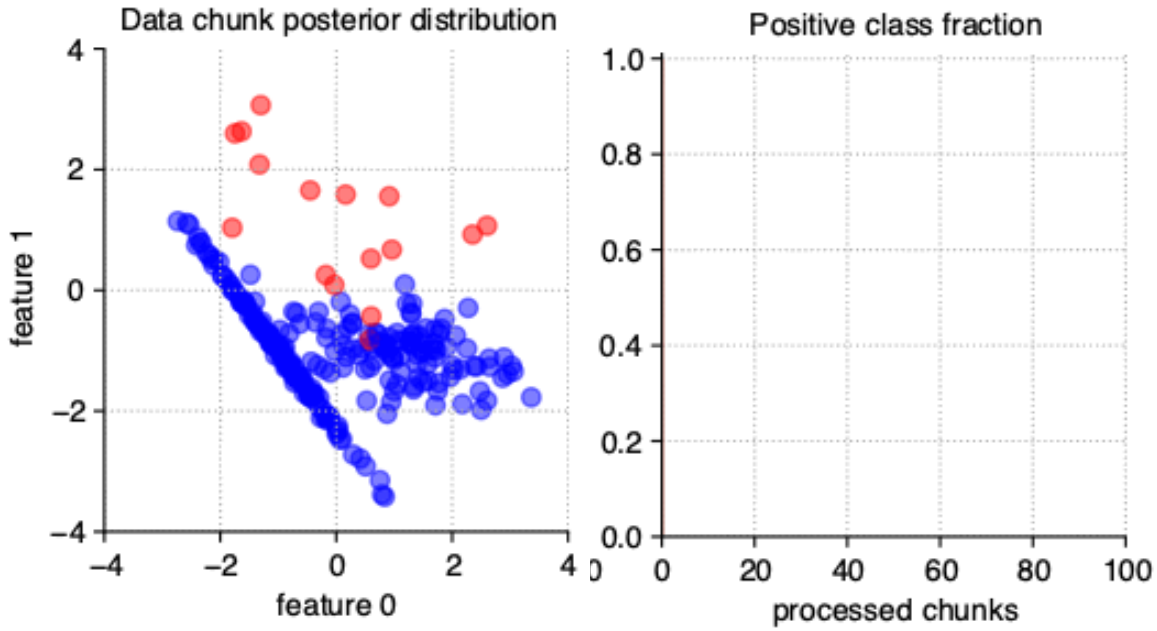
# Data arrives in a continuous stream

As “easy” as a **balanced** one




[src: https://stream-learn.readthedocs.io/en/latest/_images/stationary.gif ]

As “complex” as a **Dynamically imbalanced** one



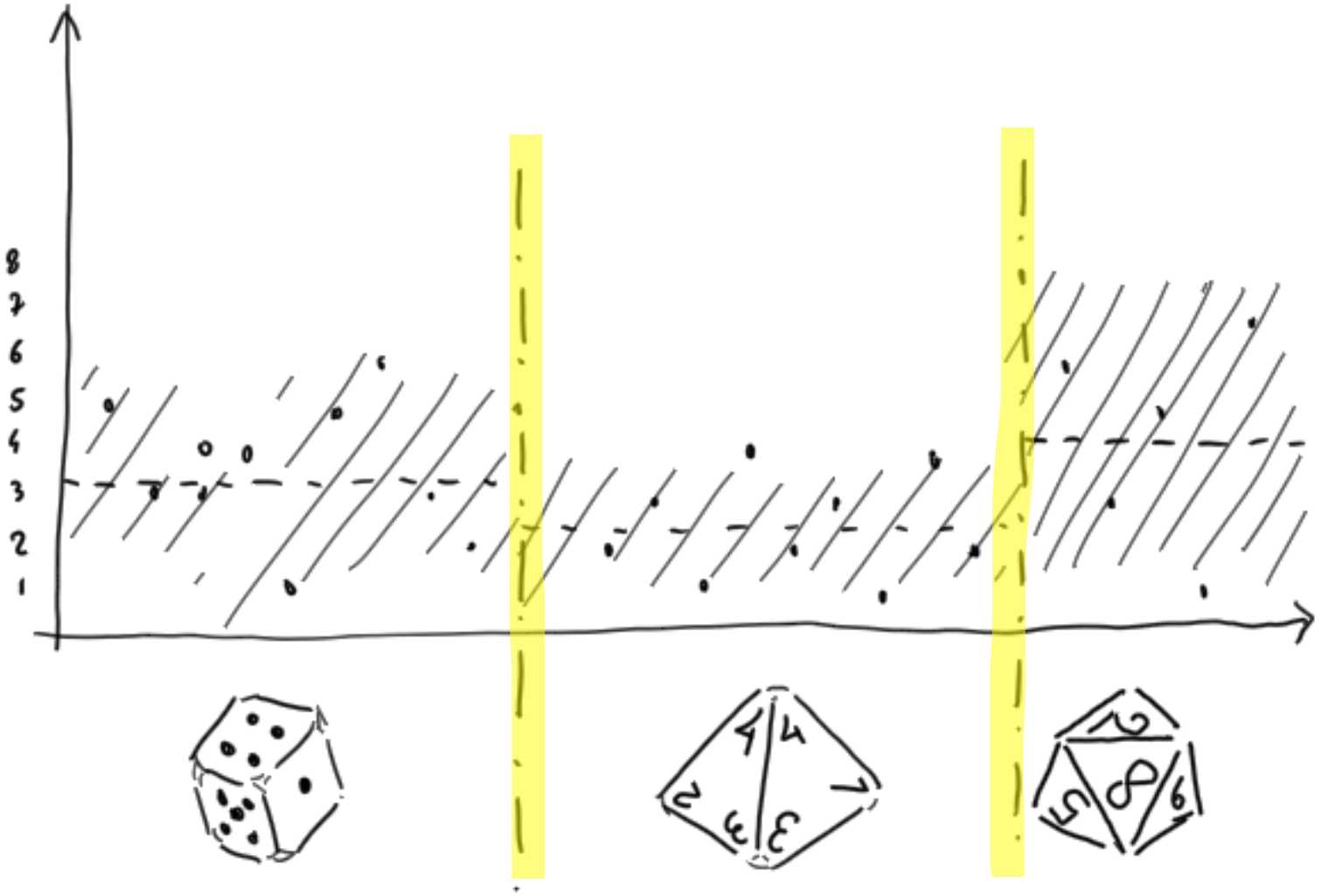

[src: https://stream-learn.readthedocs.io/en/latest/_images/dynamic-imbalanced.gif ]

# Hoeffding Trees

- **Decision Tree built incrementally**
  - A data point at a time
  - Memory only stores the model
- There are **theoretical guarantees** that the final tree is identical (with high probability) to a tree built using a batch decision tree algorithm
- Split if $H(x_1) - H(x_2) > \varepsilon = \sqrt{\frac{R^2 * \log(1/\delta)}{2N}}$


[src: https://www.bilibili.tv/en/video/4787736782838275 ]

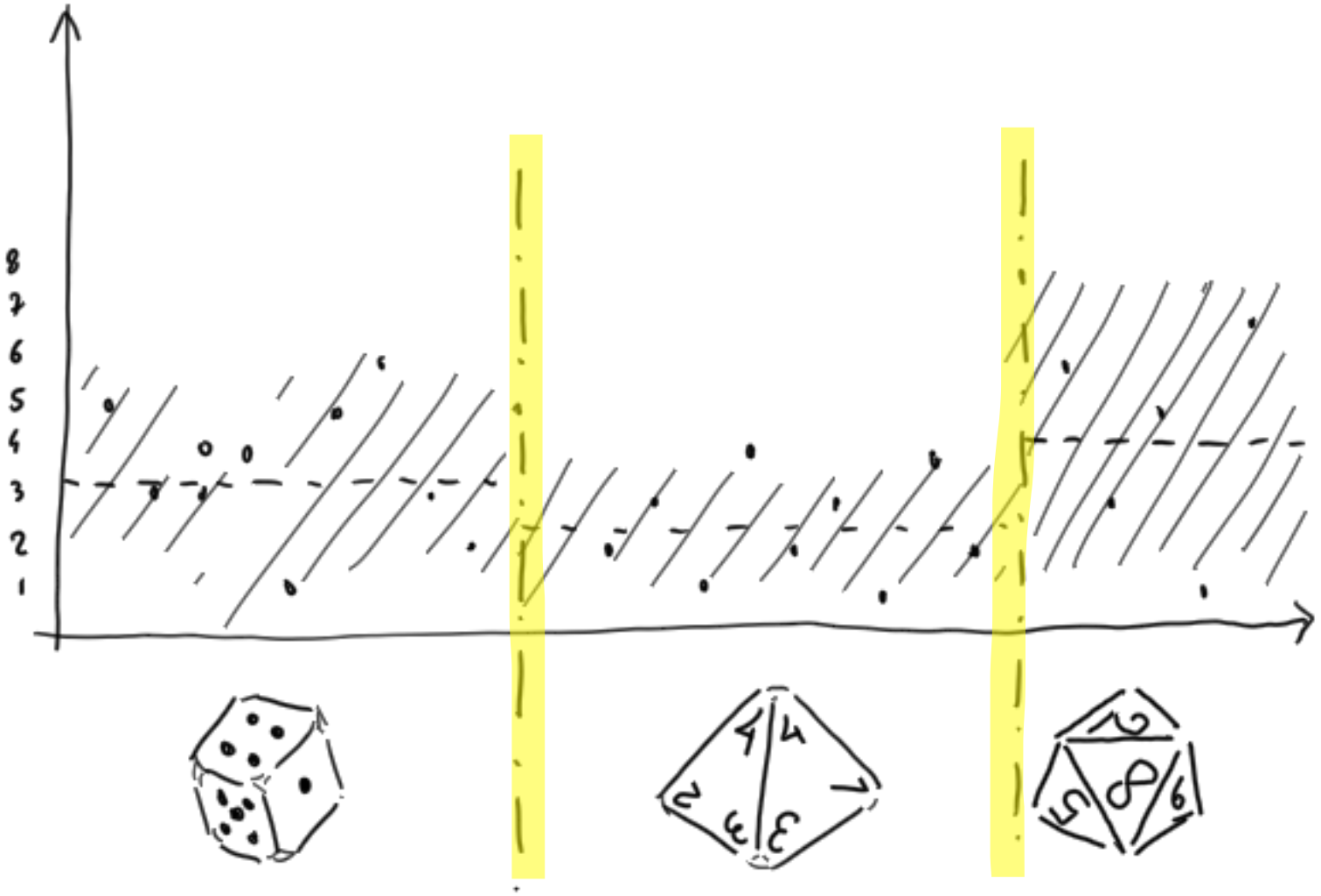
# Recall: Non identically distributed data



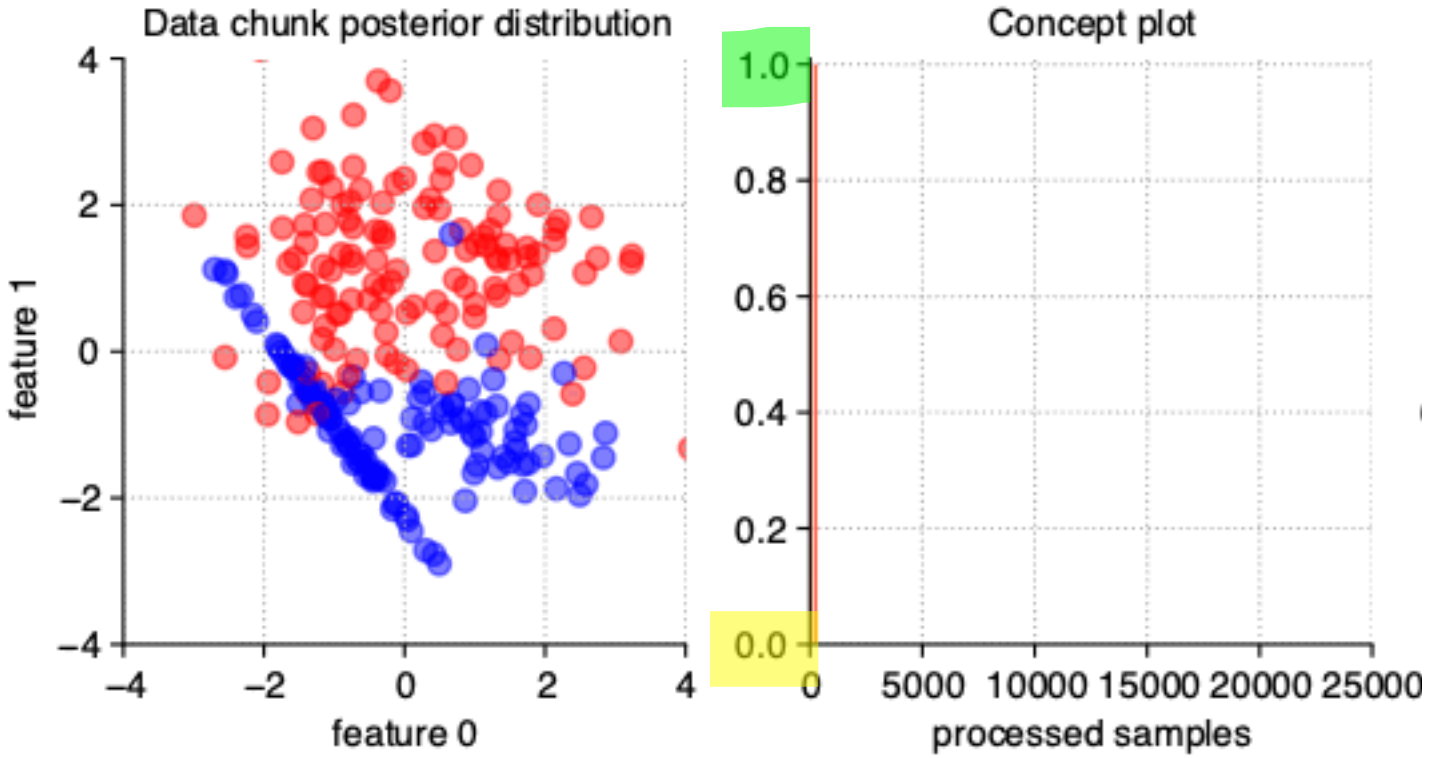
DRIFTs
(a.k.a., change points in TSA)

# Precisely, a Data Drift (i.e., a change in X*)

*** X is the input distribution**

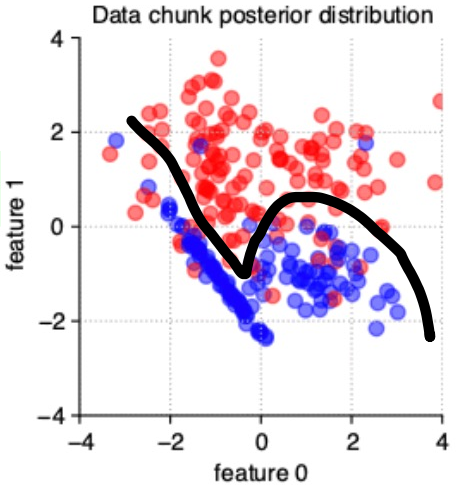


# Concept Drift is a change in X → y !!!

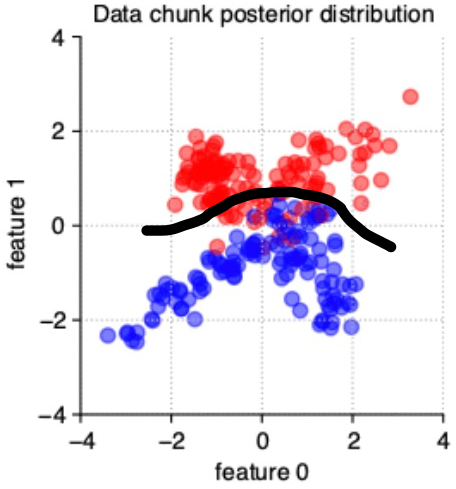



[src: https://stream-learn.readthedocs.io/en/latest/_images/incremental.gif ]

Concept 1




Concept 0

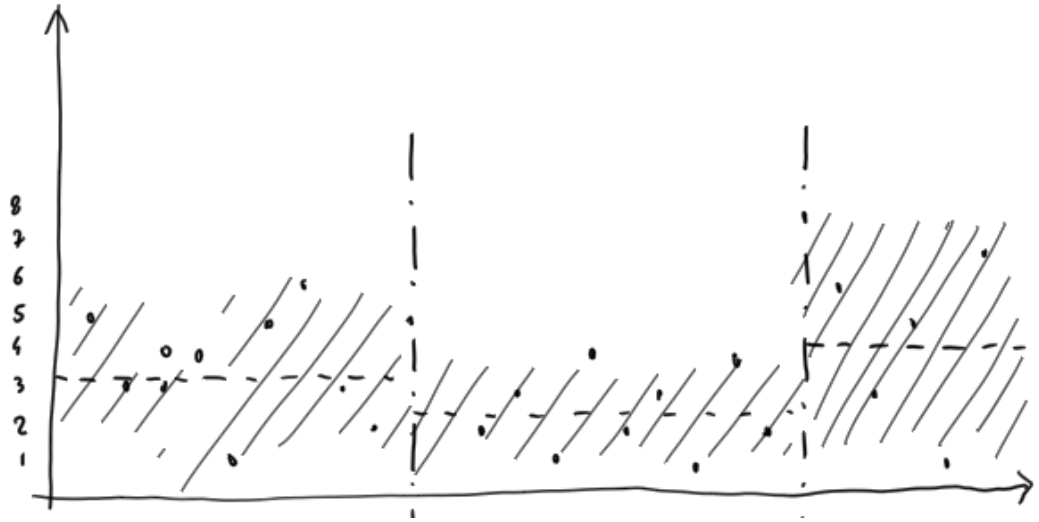


## Techniques

# Drift detectors

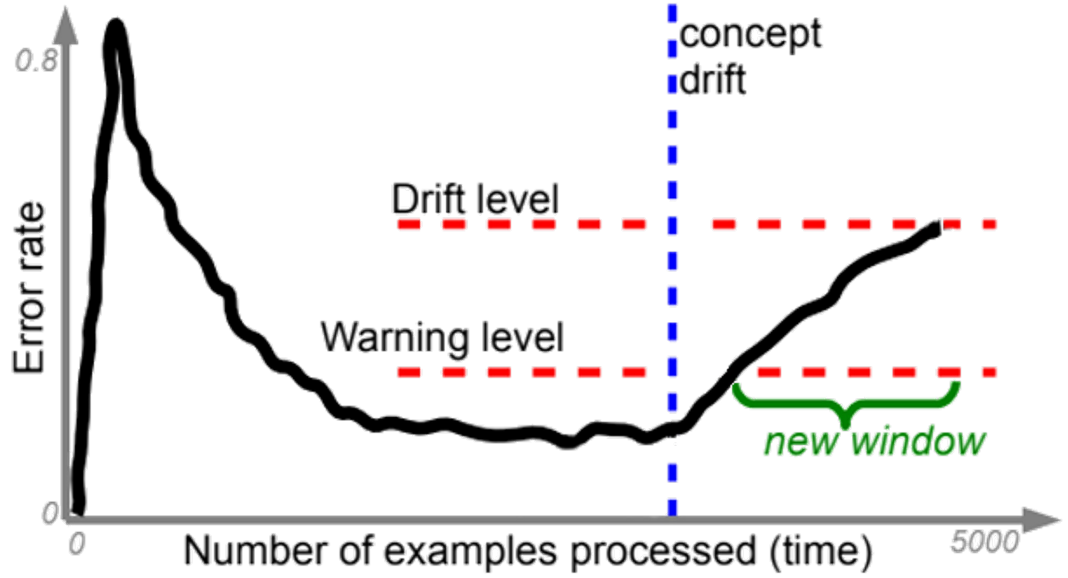
### Monitoring the input distribution

Is there a statistically significant **difference between** the distribution of the **recent X and** the one of the **old X**?



### Monitoring the prediction error

Is there a statistically significant **growth** between the recent and old **errors**?



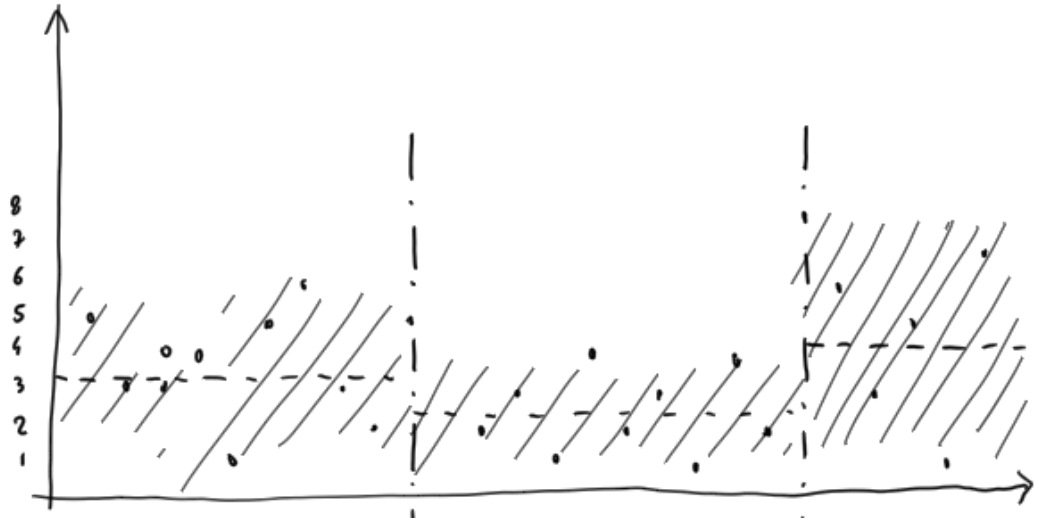
## Techniques

# Drift detectors

ADWIN!

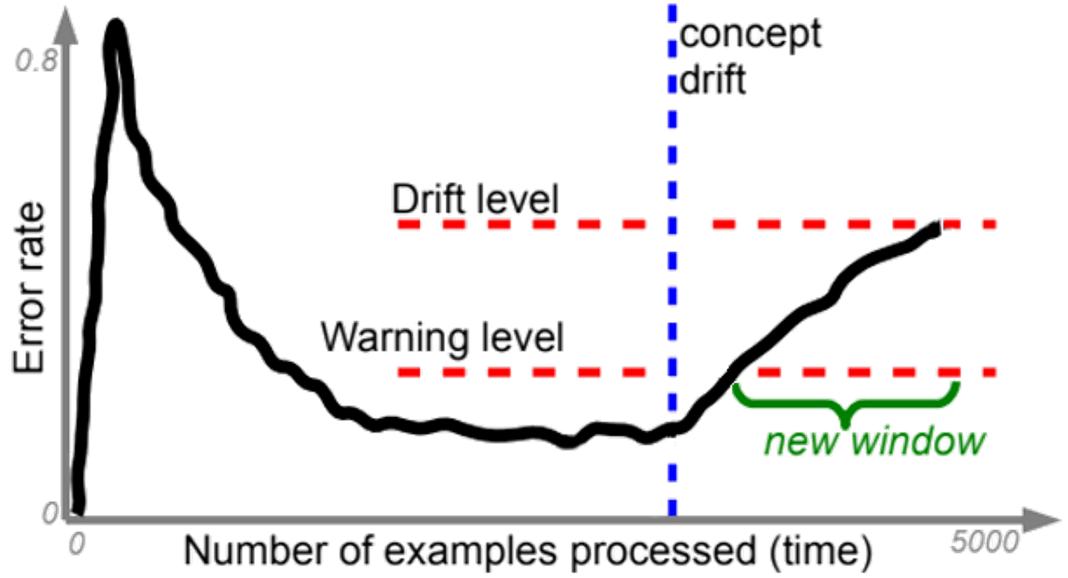
### Monitoring the input distribution

Is there a statistically significant **difference between** the distribution of the **recent X and** the one of the **old X**?



### Monitoring the prediction error

Is there a statistically significant **growth** between the recent and old **errors**?



## Techniques
# Hoeffding Adaptive Tree (HAT)

**PROBLEM**:

- When a **concept drift** occurs, part of the decision tree may **lose accuracy**

**SOLUTION**:

- When **ADWIN detects** a concept drift, start **grow**ing **alternative branches**
- **When** they become **more accurate** than existing ones, **switch them** with the old ones

[src: https://giphy.com/gifs/dance-dancing-groot-JwTqLNfrx4OPe ]

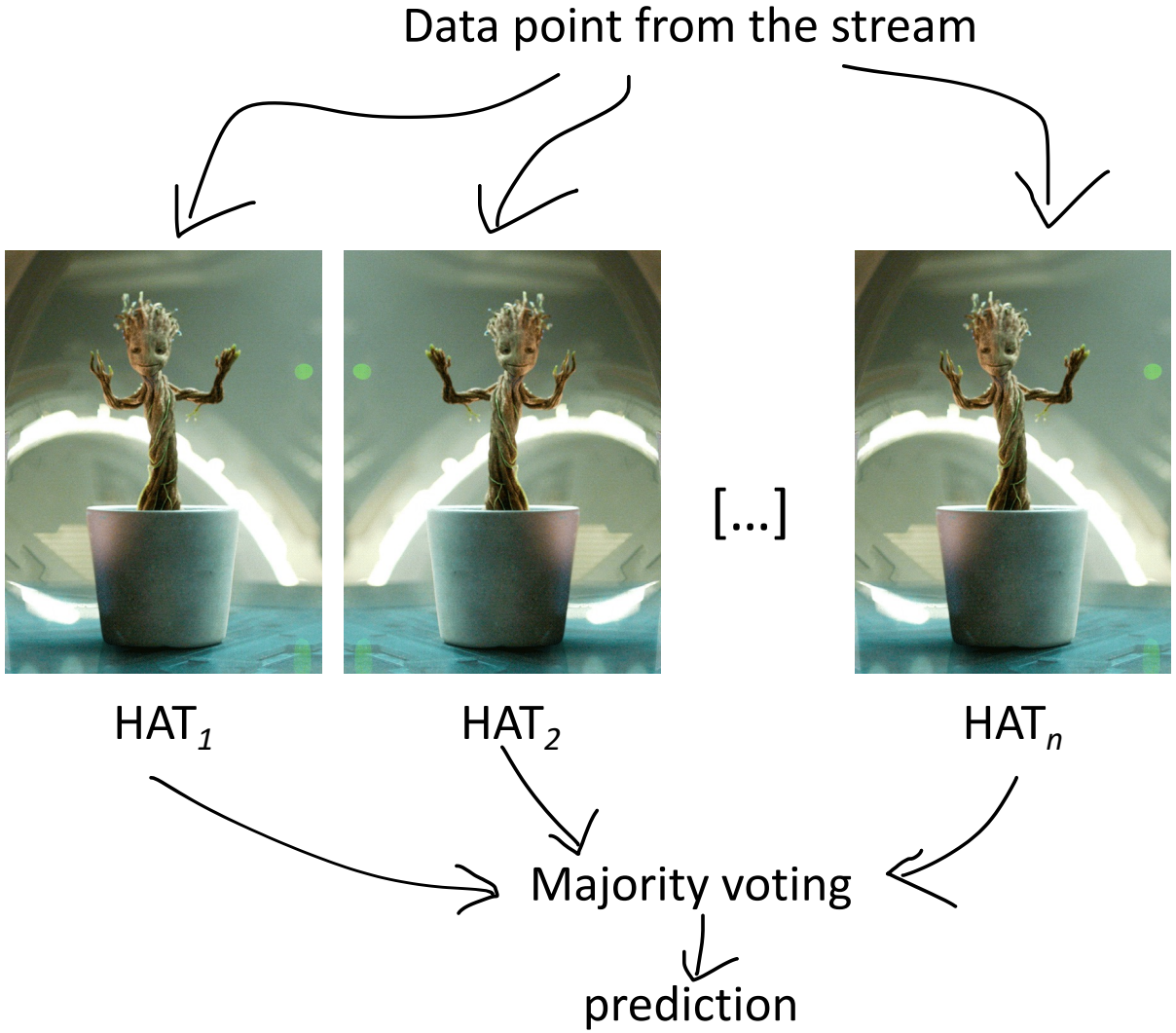
# Techniques
# Ensembles!

Via **Bagging**

- **Combine *n* HAT** to produce an output with a lower variance
- E.g., **A**daptive **R**andom **F**orest
  (see animation to the right)

Via **Boosting**

- Fits HAT models **sequentially on** the **residuals** of the previous HAT model
- **E.g., Ada**ptive **Boost**ing



# Predictive maintenance

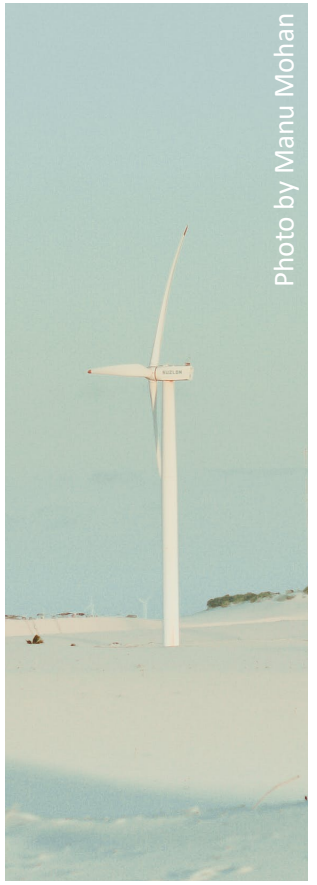
- Model and **predict** the **wear** of **equipment** to estimate
  - if it will require **maintenance** in a given timeframe (classification task)
  - its the **remaining useful life** (forecasting task)

- To learn more
  https://www.linkedin.com/posts/alberto-fassio-raiway_sima-ai-edgecomputing-activity-7118114440931123201-zYs2

# Precise continuously learnt forecasts for renewable energy

- Prediction of energy production from renewable sources to **minimize imbalance costs** associated with market operations
- **Forecasts** are **continuously** and directly **adapted** in real-time to:
  - **specific performance** of each solar panel or wind turbine,
  - actual operating **conditions**,
  - local **environment**al variables and
  - specific **weather forecast**s for each plant.
- To learn more motus ml spin off POLITECNICO DI MILANO





# Learning Characteristics

**S**treaming **M**achine **L**earning is

- **adaptive** and designed to handle
  - changing data distributions, a.k.a., data drifts or virtual concept drift
  - Concept drifts, a.k.a., changes in X → y
- well-suited for applications requiring **immediate response to** incoming data **changes**

[src: https://gifer.com/en/2rci ]

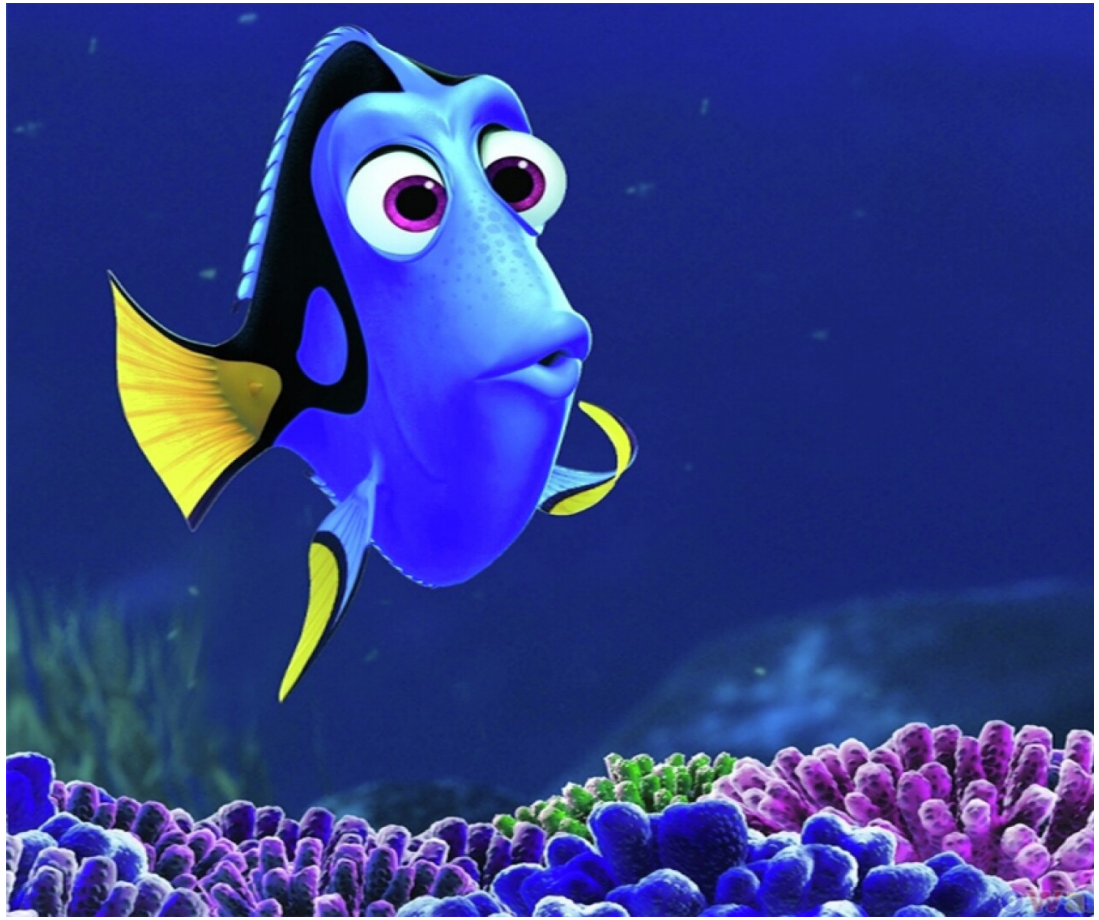
# Reflection

QUESTION

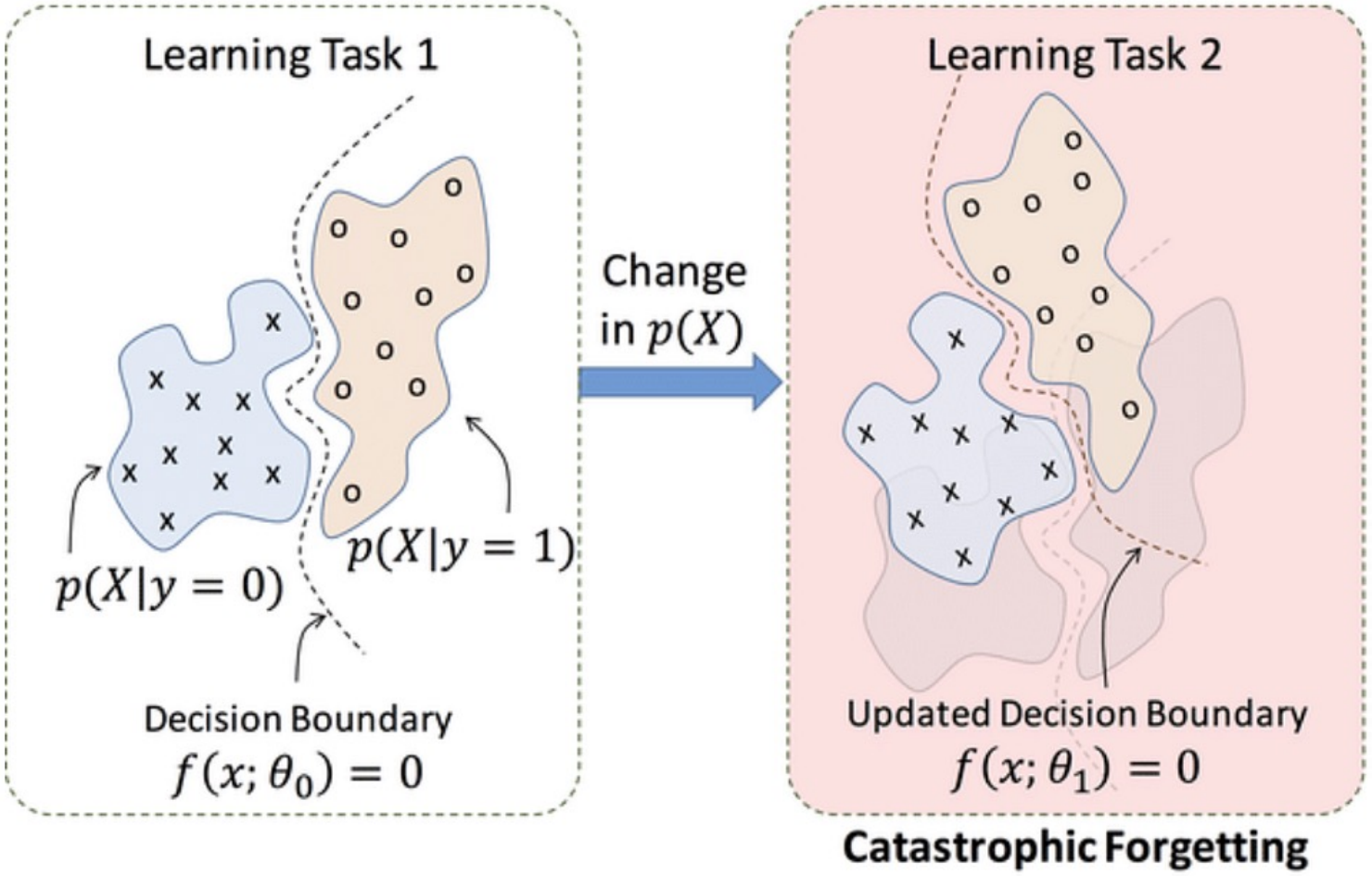
- At what cost does SML immediately react to changes?

ANSWER

- It forgets!

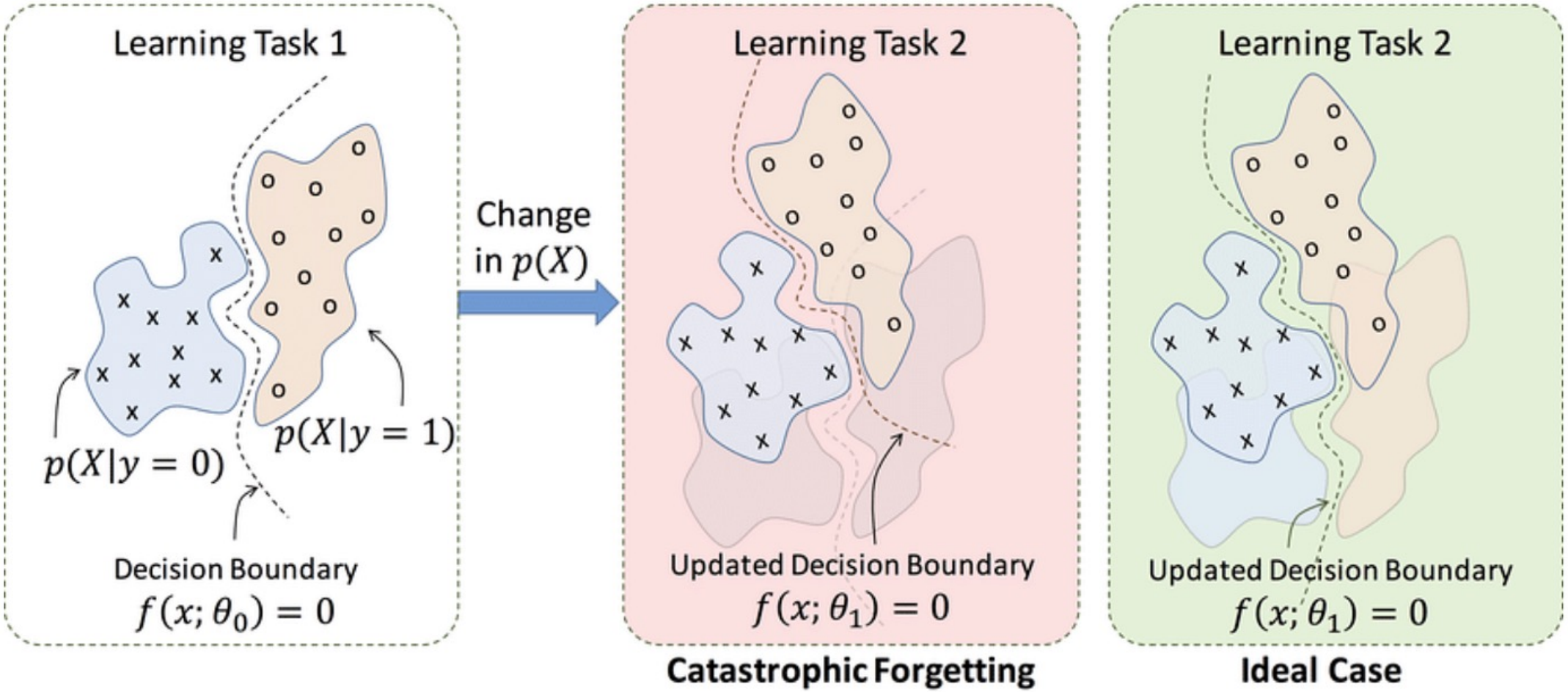


# Catastrophic forgetting




[src: https://arxiv.org/pdf/1903.06070.pdf ]

# Catastrophic forgetting (cont.)




[src: https://arxiv.org/pdf/1903.06070.pdf ]

# The plasticity-stability dilemma in AI

**Plasticity** is the ability to adapt to changes


Photo by Antoni Shkraba

**Stability** is the ability to retain what learned while adapting


Photo by Aleksey Kuprikov

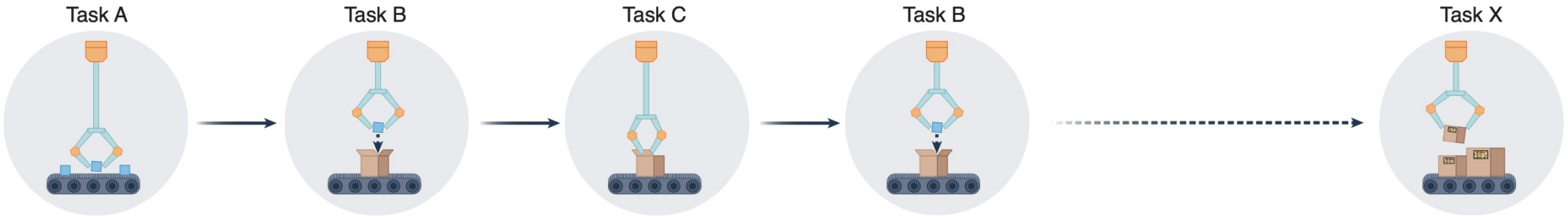



Photo by Lum3n

## Lifelong learning
# definition & an example

- Learning from a sequence of training episodes intermixed with situations that require applying (recently or previously) learned skills

- E.g., a robotic arm

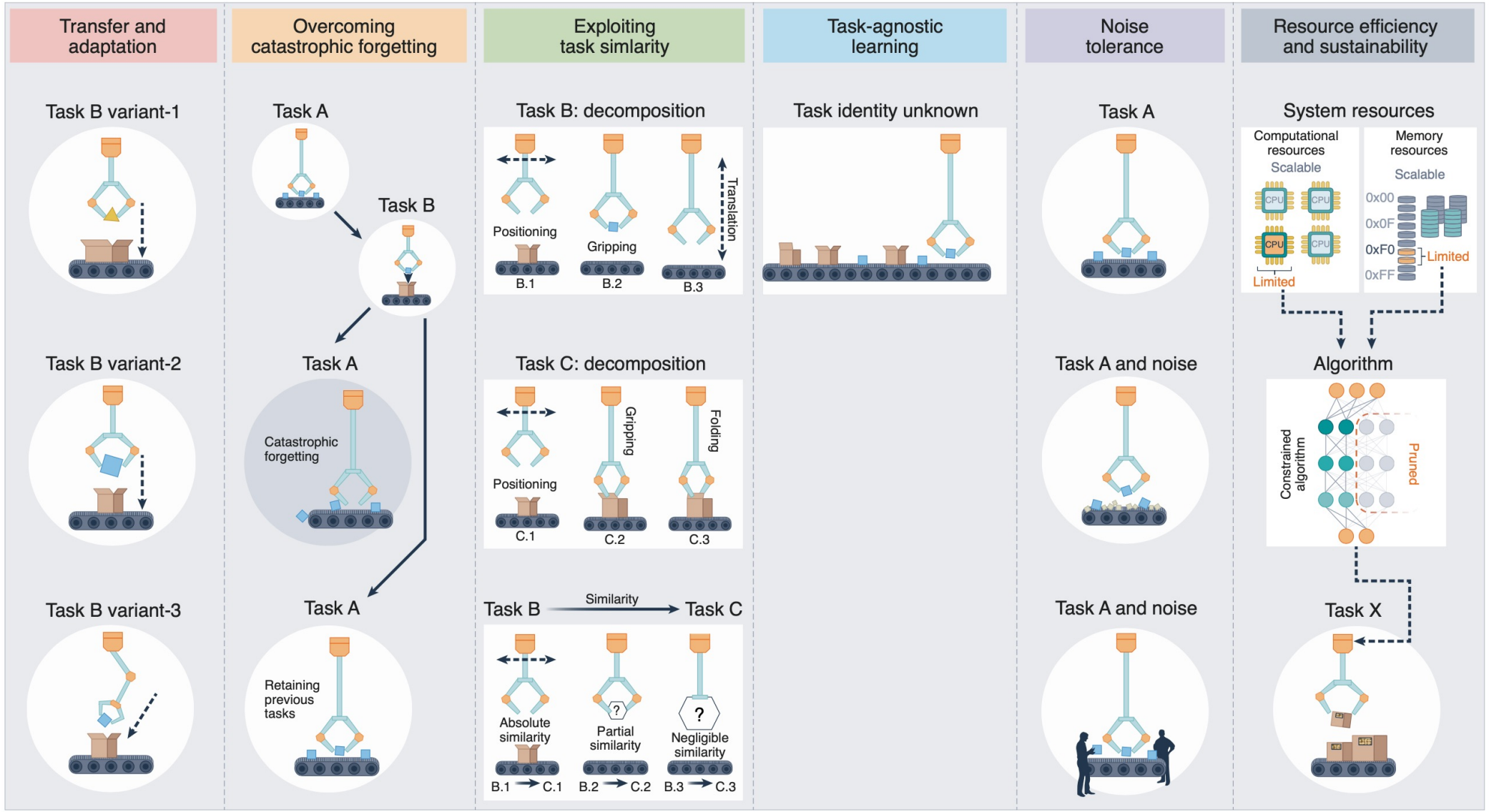



[src: https://www.nature.com/articles/s42256-022-00452-0 ]

# Type of data

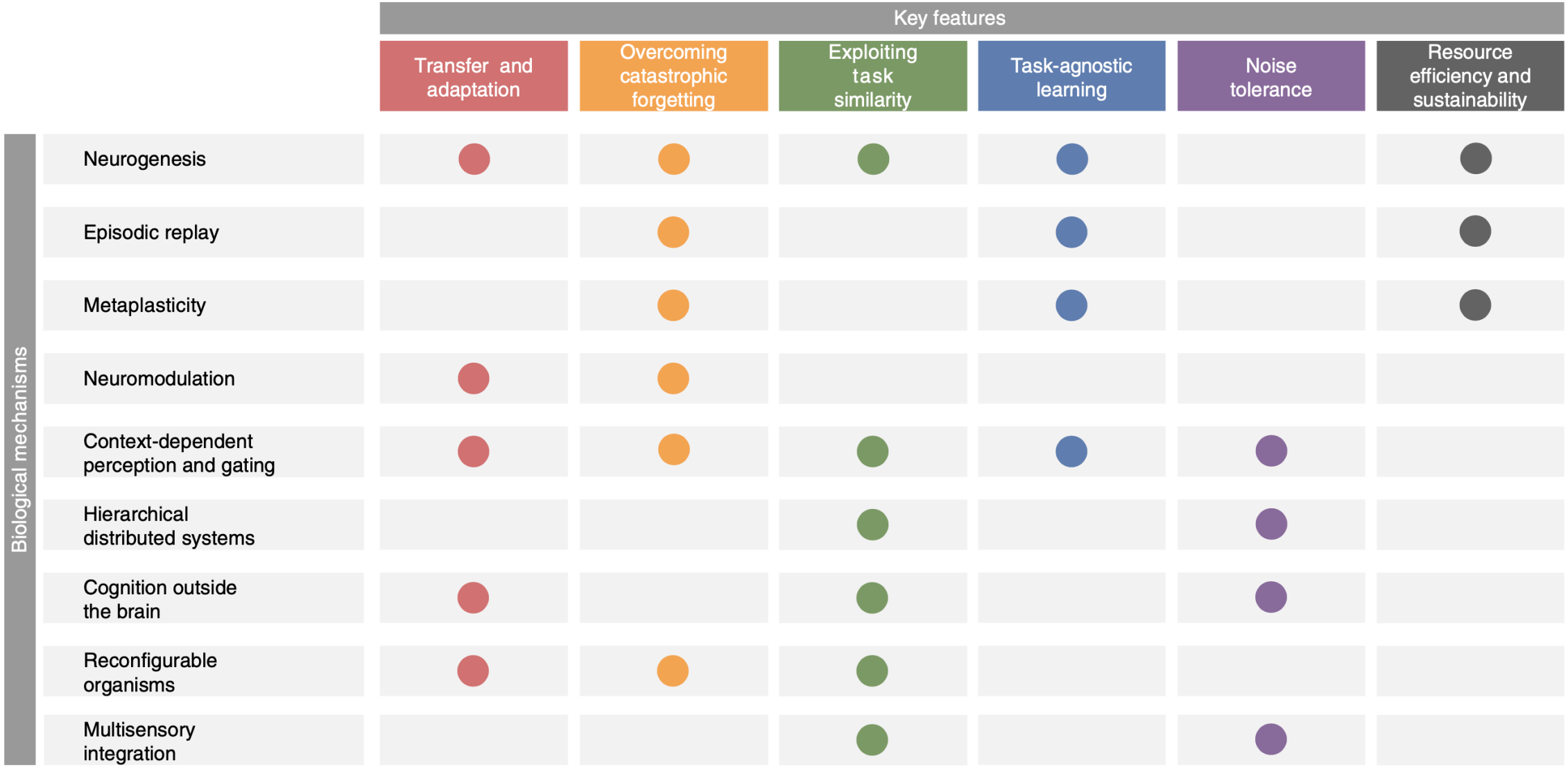
data arrives in **manually drafted experiences** (a.k.a., **tasks**), where training and test episodes intermix, assuming **data** points are **i.i.d. only within the same experience**. Data in two experiences are normally distributed differently

# Lifelong learning
# Key features




[src: https://www.nature.com/articles/s42256-022-00452-0 ]

# Lifelong learning
# Biological underpinnings




[src: https://www.nature.com/articles/s42256-022-00452-0 ]

# Continual AI Purpose

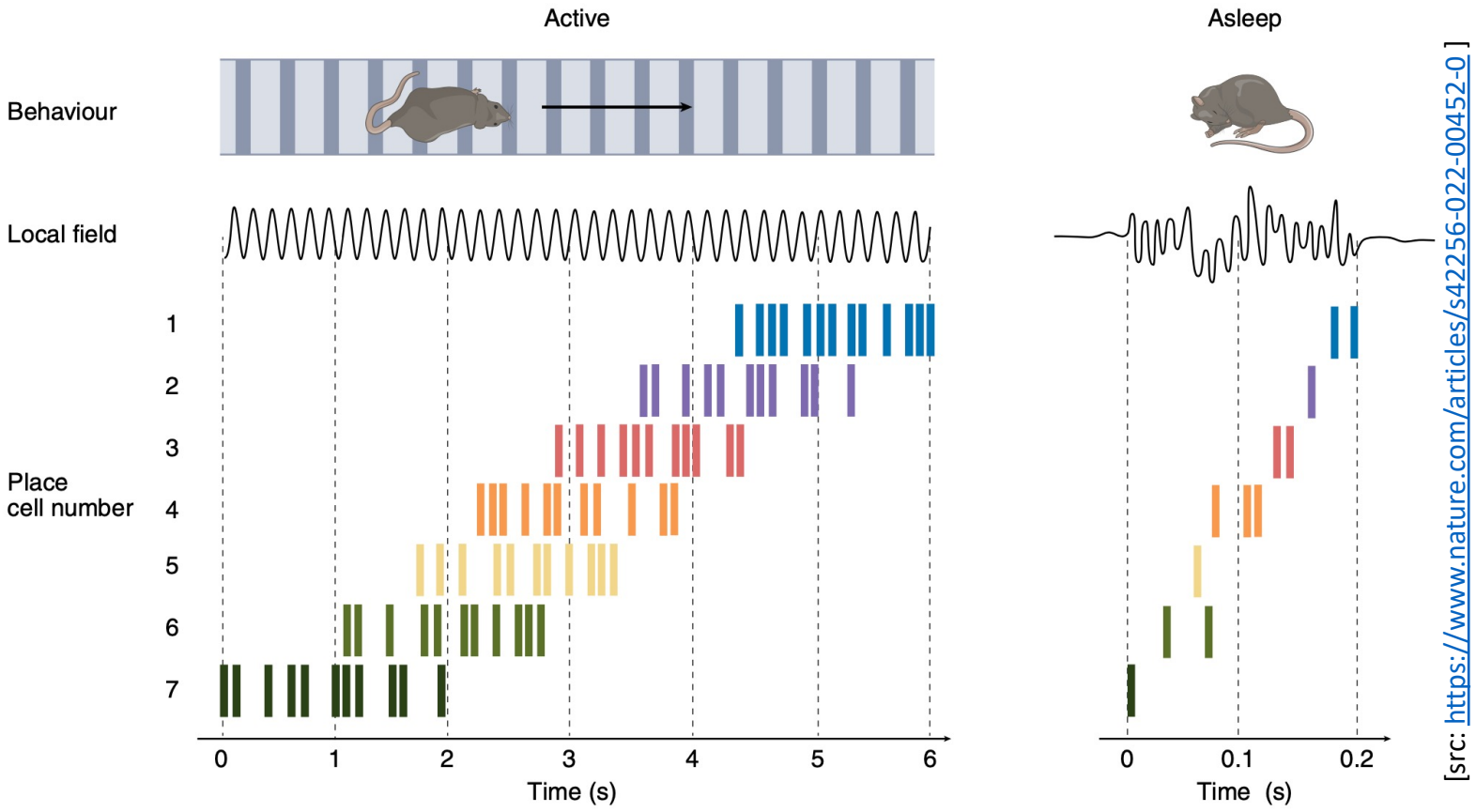
- It is used primarily to **avoid catastrophic forgetting** while learning from new experiences
- It **balances** between **stability and plasticity** potentially via task similarity
- Eventually, it also learns when task identity is unknown, tolerates noise, and is more prone to resource constraint environments


[src: https://www.continualai.org/ ]

# Techniques
# Rehearsal strategies (a.k.a., episodic replay)

Alleviate catastrophic forgetting by

- **storing past data** in an episodic memory system
- **regularly replaying past data** with real samples drawn from the new task
- **learning from the** obtained **mix** of replayed and real samples

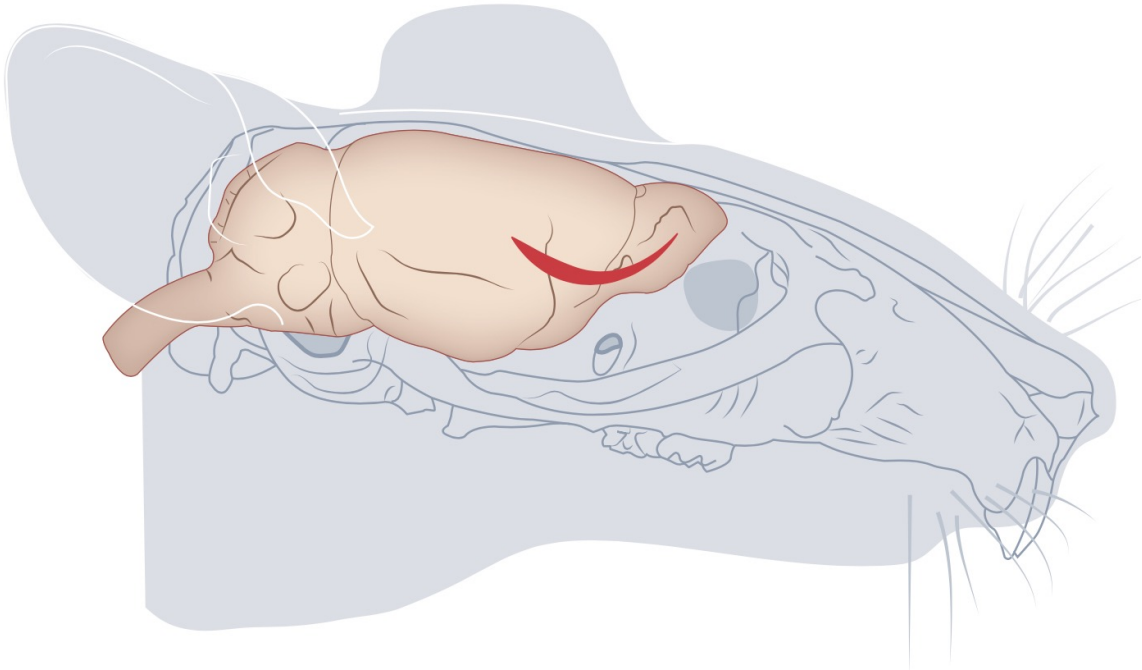



[src: https://www.nature.com/articles/s42256-022-00452-0 ]

# Techniques
# Architectural strategies (a.k.a., neurogenesis)

Alleviate catastrophic forgetting by

- choosing specific deep learning architectures, layers, activation functions, and/or
- weight-freezing and **dynamic architectures** (e.g., adding new layers)



In mice, olfactory inter**neurons** are **produced and** subsequently **migrate** to the olfactory bulb to scale up the number of **new memories** that can be encoded and **stored without catastrophic forgetting** of previously consolidated memories

[src: https://www.nature.com/articles/s42256-022-00452-0 ]

# Learning Characteristics

**C**ontinual **L**earning

- ensures that models can **adapt to new information** while **retaining previously learned knowledge**
- **focus**es on long-term model **stability**

# Overall conclusion

In summary, ***streaming machine learning*** is designed for ***real-time data stream processing*** and model adaptation***, time series analytics*** is primarily ***retrospective*** and focuses on historical data analysis, while ***continual learning*** is concerned with the ***long-term adaptation*** of models to new data while avoiding forgetting essential knowledge.

The **choice** of approach **depends**
**on** the specific **requirements** of the application **and**
the nature of the **data** being analyzed.

# Streaming Data Science


## Emanuele Della Valle

**Prof @ PoliMI**
**CRO & founder @ Motus ml**
**founder @ Quantia Consulting**


POLITECNICO
MILANO 1863