Navigating Complex Machine Learning Challenges in Streaming Data

ECML Tutorial 2024

Heitor Murilo Gomes[1*], Marco Heyden[2]
Maroua Bahri[3,4]
https://heymarco.github.io/ecml24-streamingchallenges/

* Corresponding author: heitor.gomes@vuw.ac.nz

https://capymoa.org/

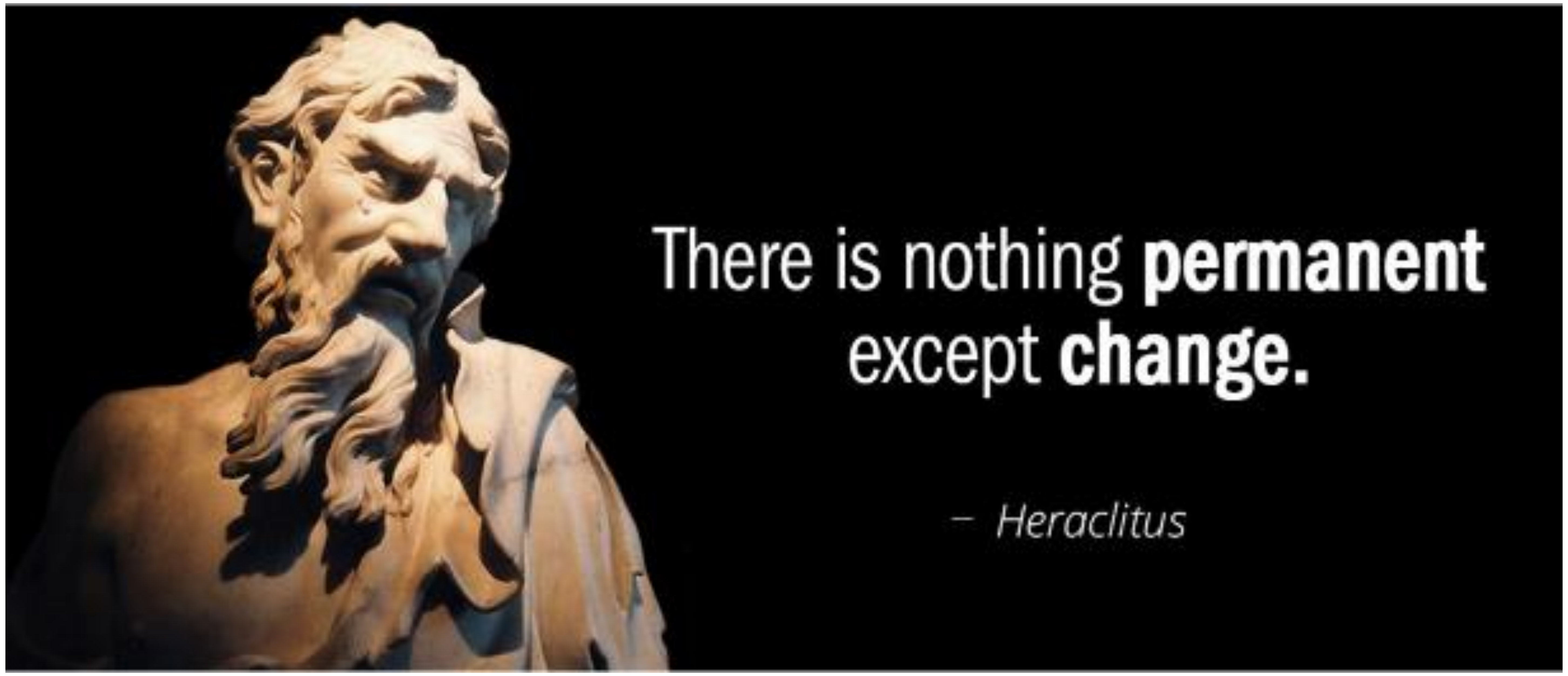
[1] Victoria University of Wellington, New Zealand, [2] KIT, Germany, [3] INRIA Paris, France,
[4] Sorbonne Université, France

Concept Drifts

Evolving Stream Learning

- The world is dynamic… **changes occur all the time**

- These **changes affect** our **machine learning models**



Evolving Stream Learning

- The world is dynamic… **changes occur all the time**

- These **changes affect** our **machine learning models**

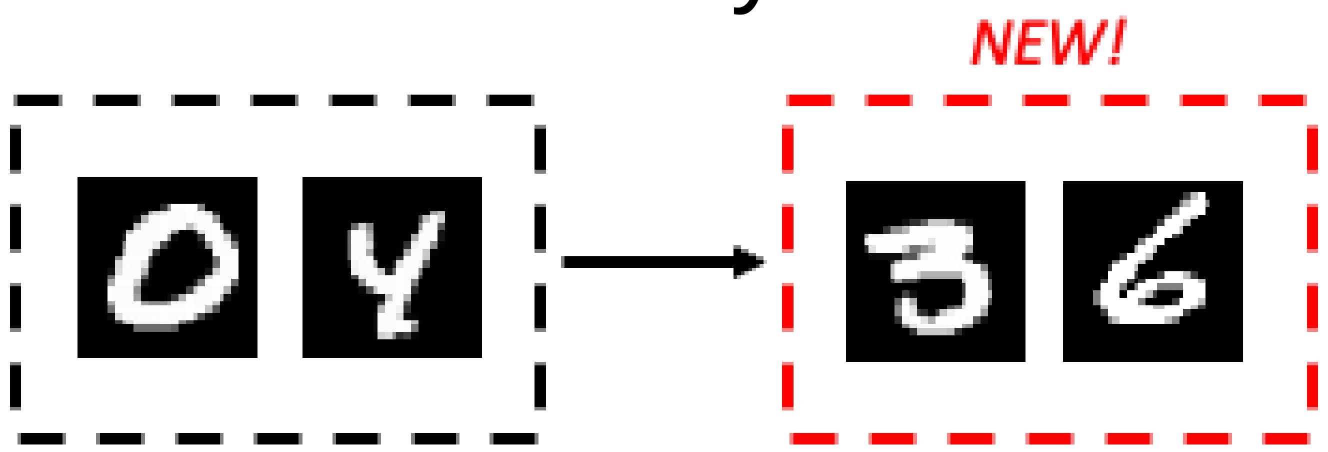
Ideally, we would like to…

(1) **Detect, understand** and **react to changes** in the data

(2) **Learn new concepts** without **forgetting old concepts**

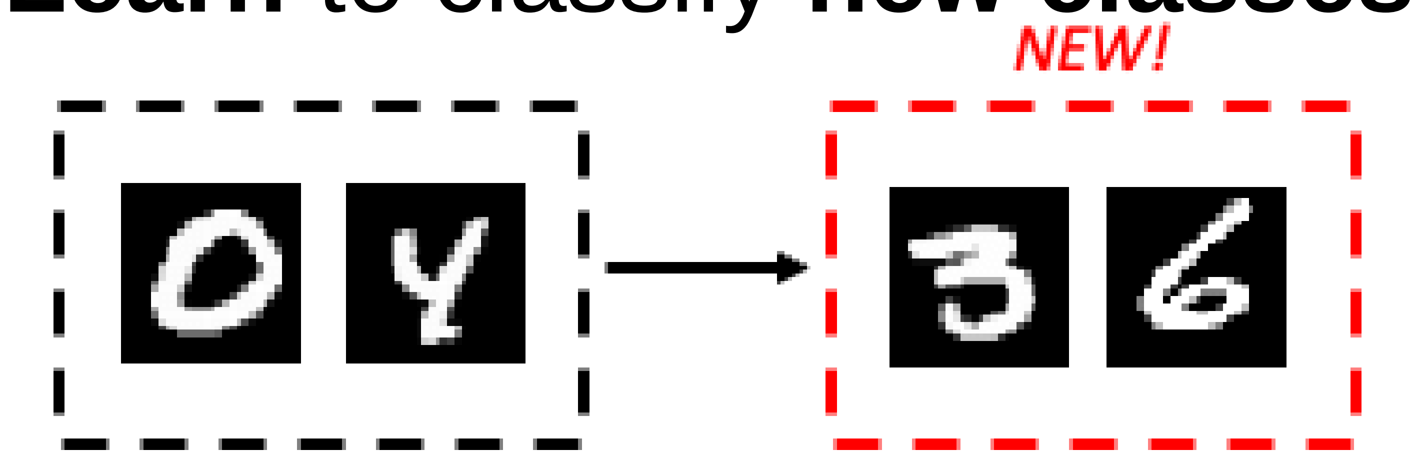
Some Examples

Learn to classify **new classes**

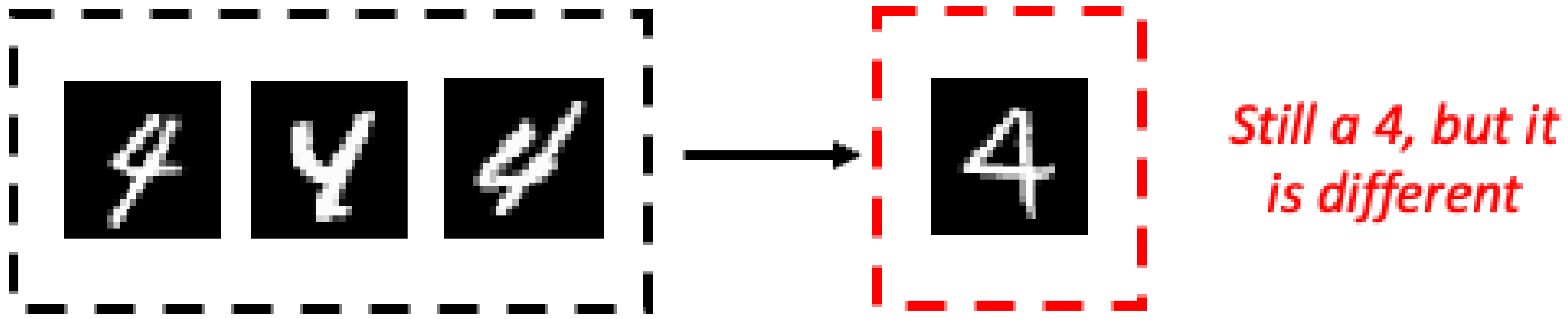


Some Examples

Learn to classify **new classes**

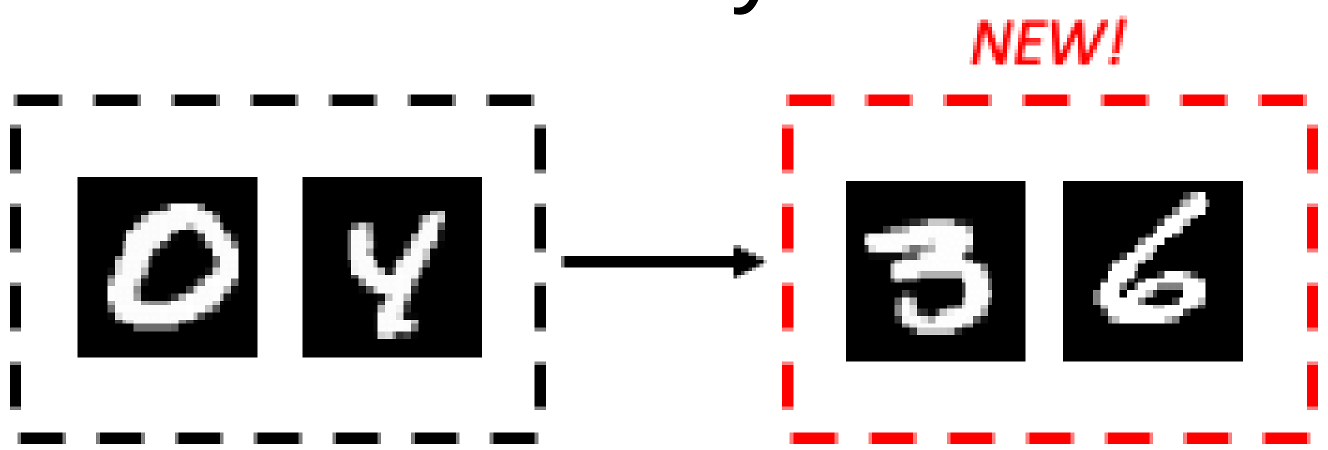


Update model to accommodate for **changes within existing classes**

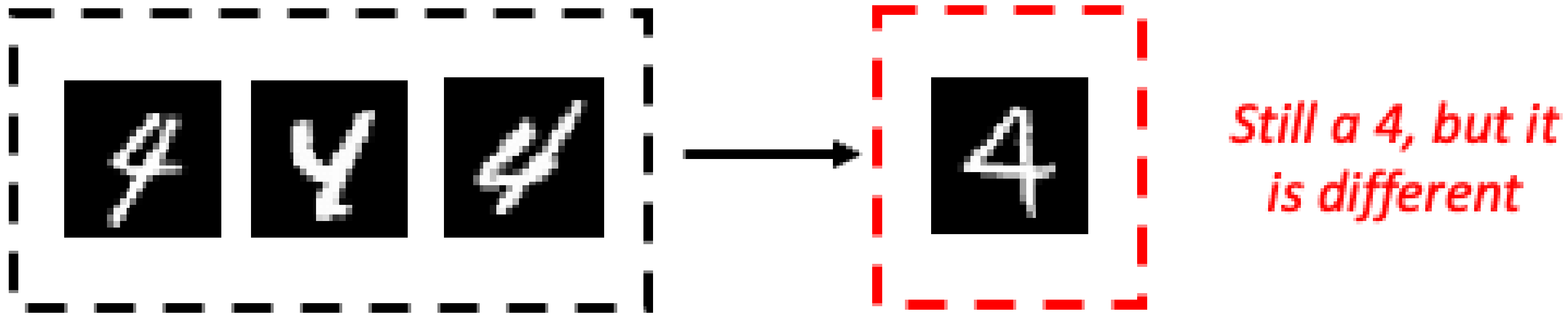


Some Examples

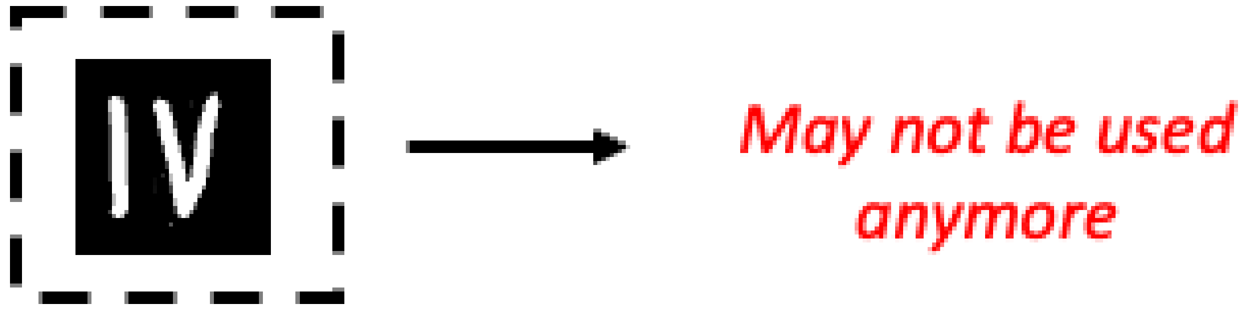
Learn to classify **new classes**



Update model to accommodate for **changes within existing classes**



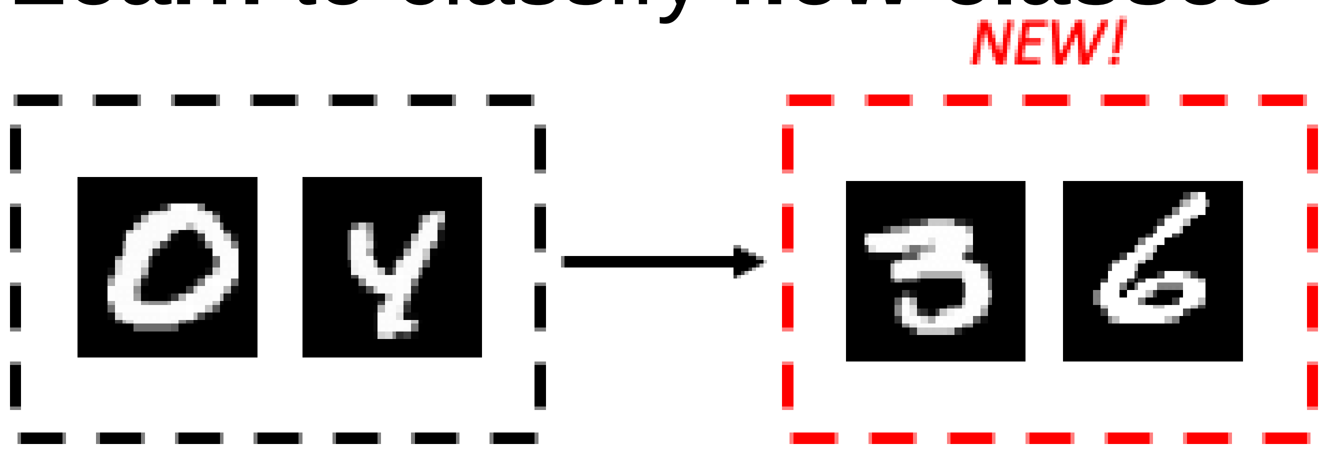
Forget that which is **no longer needed**



Some Examples

Related Research Areas / Jargon

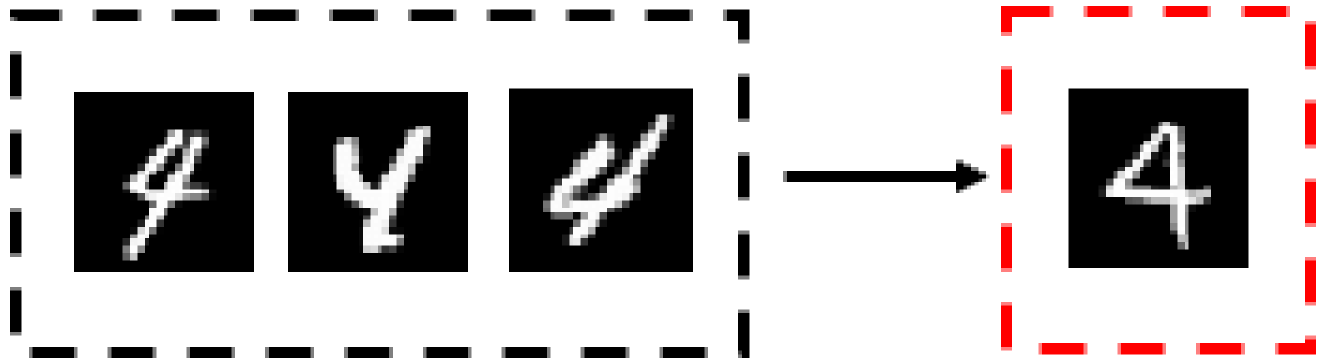
Learn to classify **new classes**



Class Evolution (Stream Learning)

Class Incremental (Continual Learning)

Update model to accommodate for **changes within existing classes**

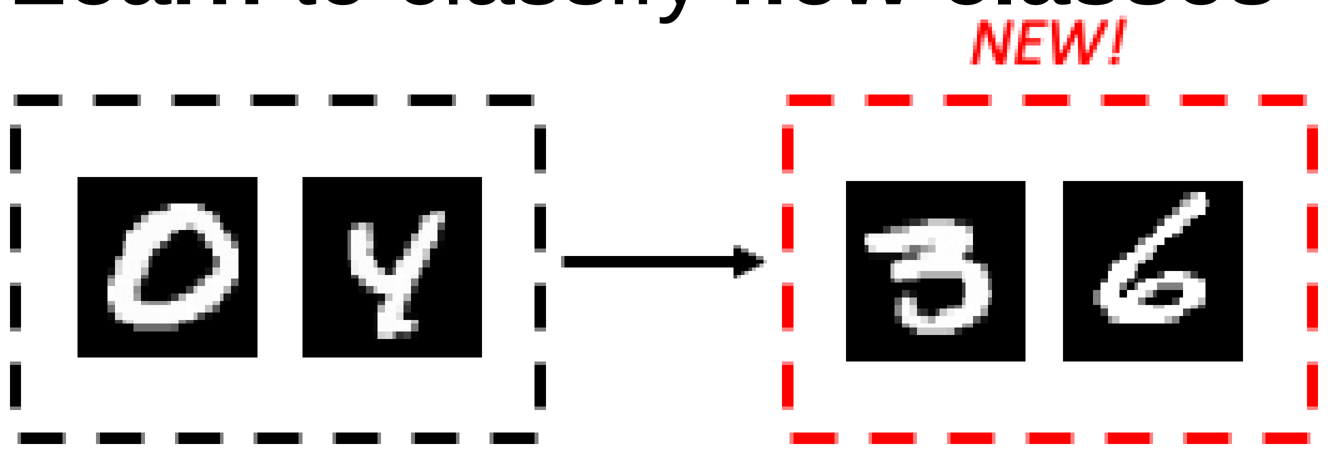


Forget that which is **no longer needed**

Some Examples

Related Research Areas / Jargon

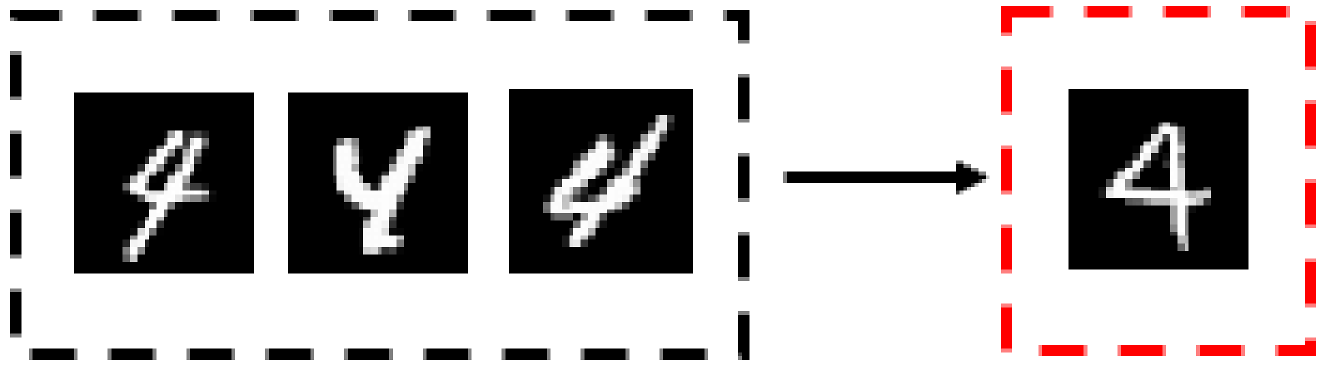
Learn to classify **new classes**



Class Evolution (Stream Learning)

Class Incremental (Continual Learning)

Update model to accommodate for **changes within existing classes**



Concept Drift (Stream Learning)

Domain Incremental (Continual Learning)

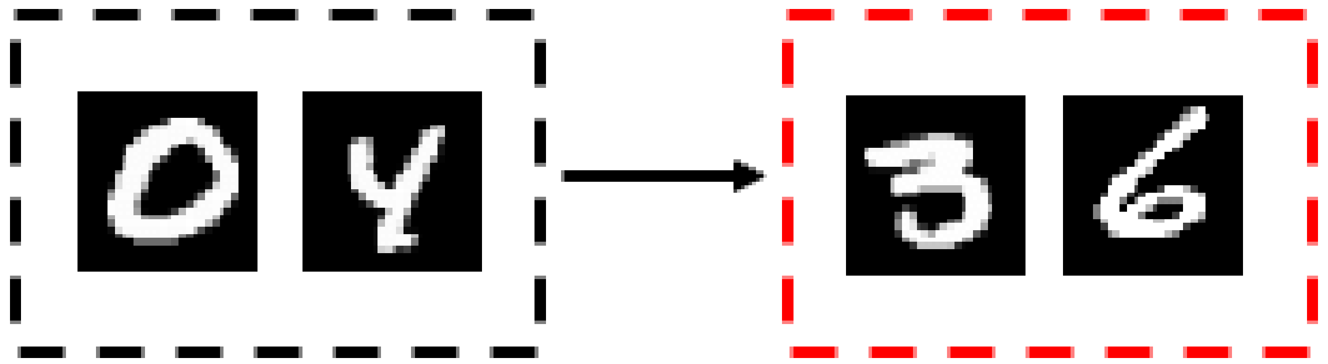
Forget that which is **no longer needed**

Some Examples

Related Research Areas / Jargon

Learn to classify **new classes**

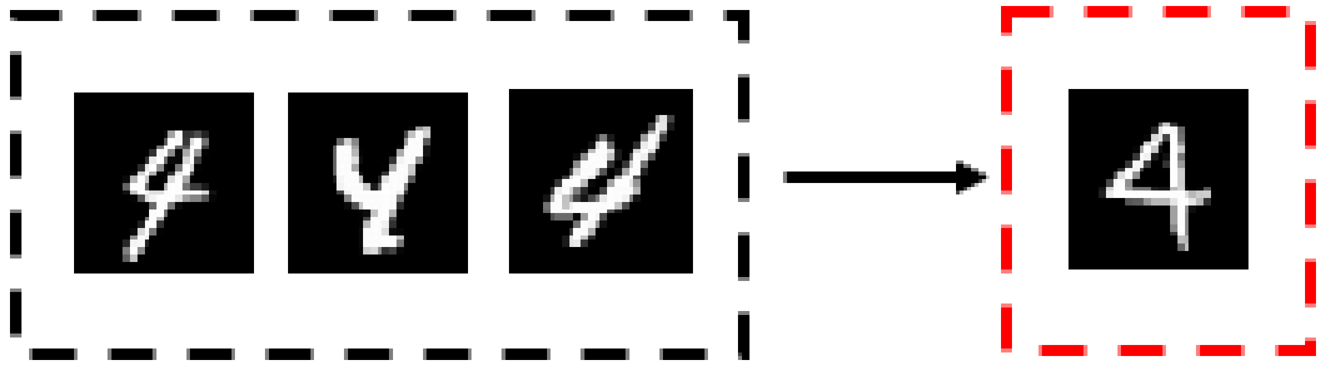
NEW!



Class Evolution (Stream Learning)

Class Incremental (Continual Learning)

Update model to accommodate for **changes within existing classes**



Concept Drift (Stream Learning)

Domain Incremental (Continual Learning)

Forget that which is **no longer needed**

May not be used anymore

Class Evolution (Stream Learning)

Assumptions

Assumptions

Independent and identically distributed (iid)

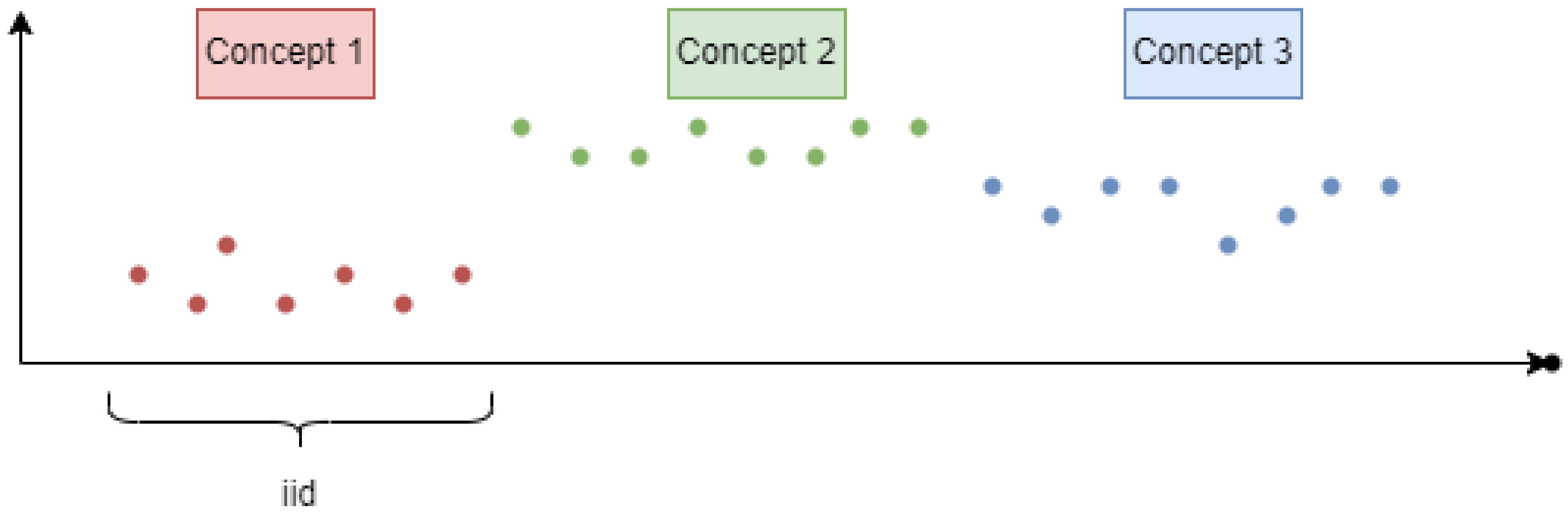
Each data point in the stream **comes from the same probability distribution** &
The values of one data point **does not provide any information** about the values of another data point

Assumptions

Independent and identically distributed (iid)

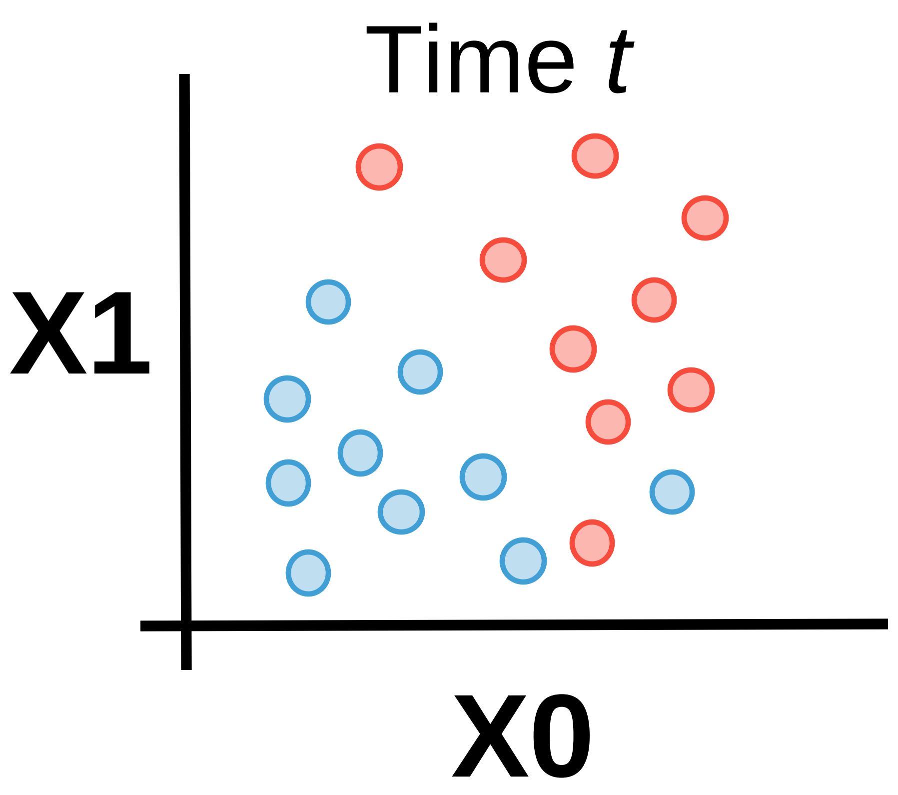
The presence of **Concept Drift (CD)** violates the **identically distributed** assumption

Different sub-populations (concepts) exist at different time intervals



Concept Drift Example

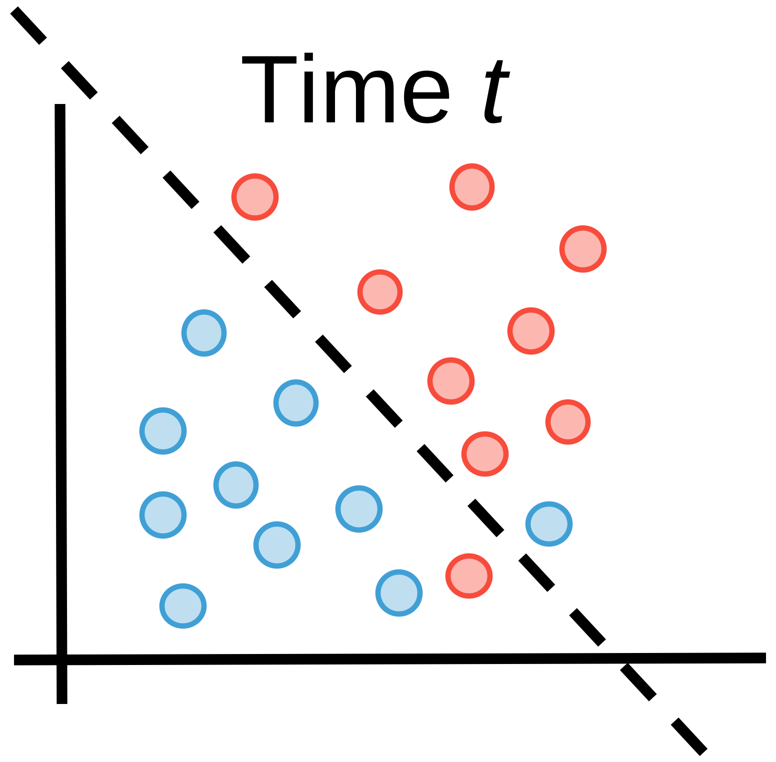
Concept drift example



Assume a simple classification problem

- Two classes
- Two features (X0 and X1)

Concept drift

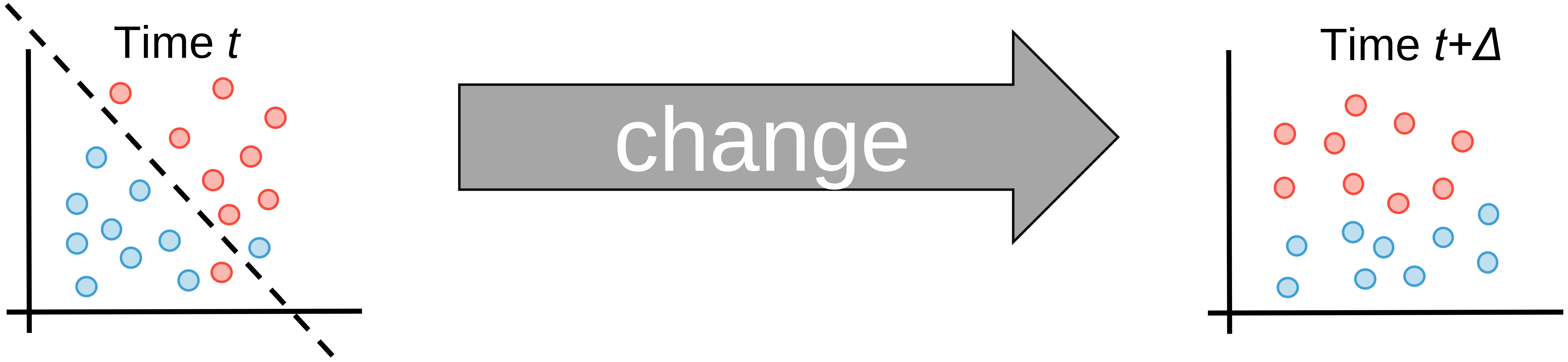


An accurate model!

We can build a very simple **linear model** to separate the two classes!

Concept drift

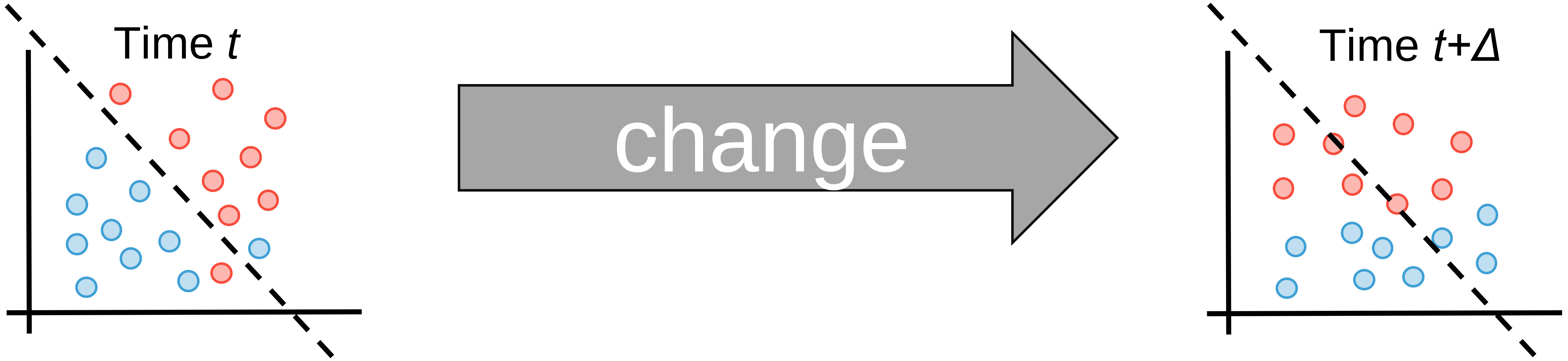
What if the data distribution **changes**?



An accurate model!

Concept drift

What if the data distribution **changes**?



An accurate model!

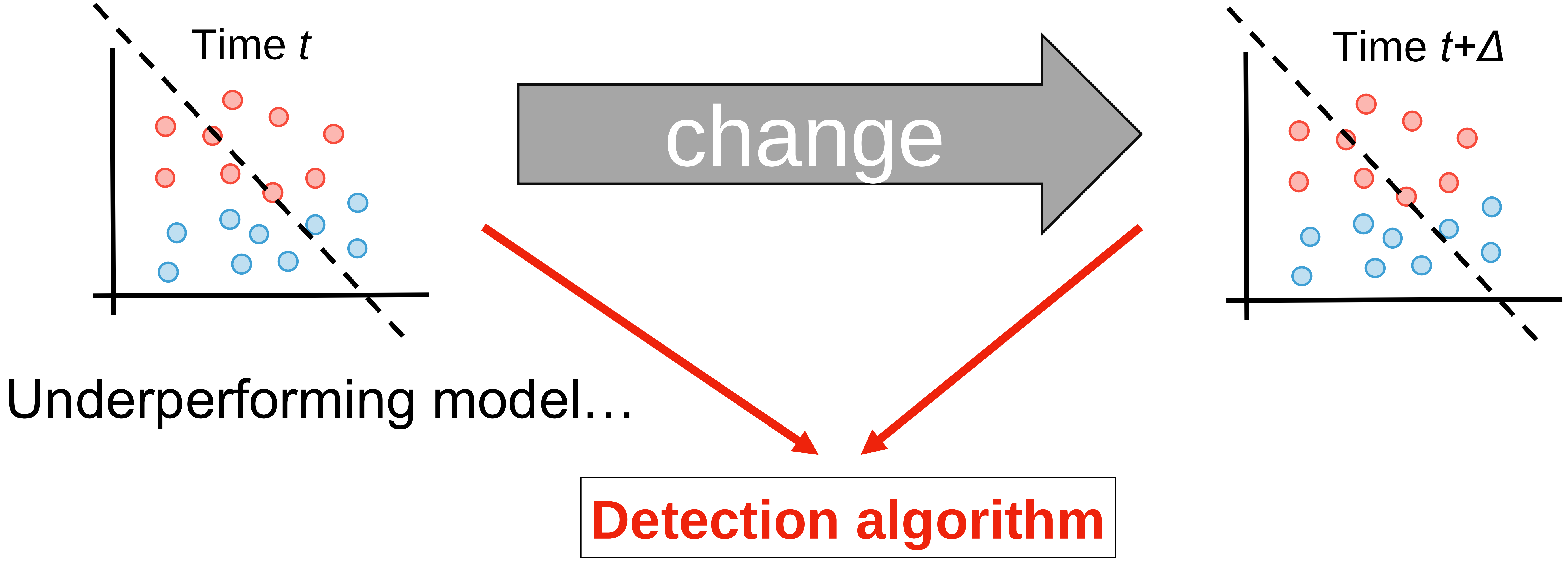
Not accurate anymore

What can we do about CD?

Detect & **Adapt** (update the model)

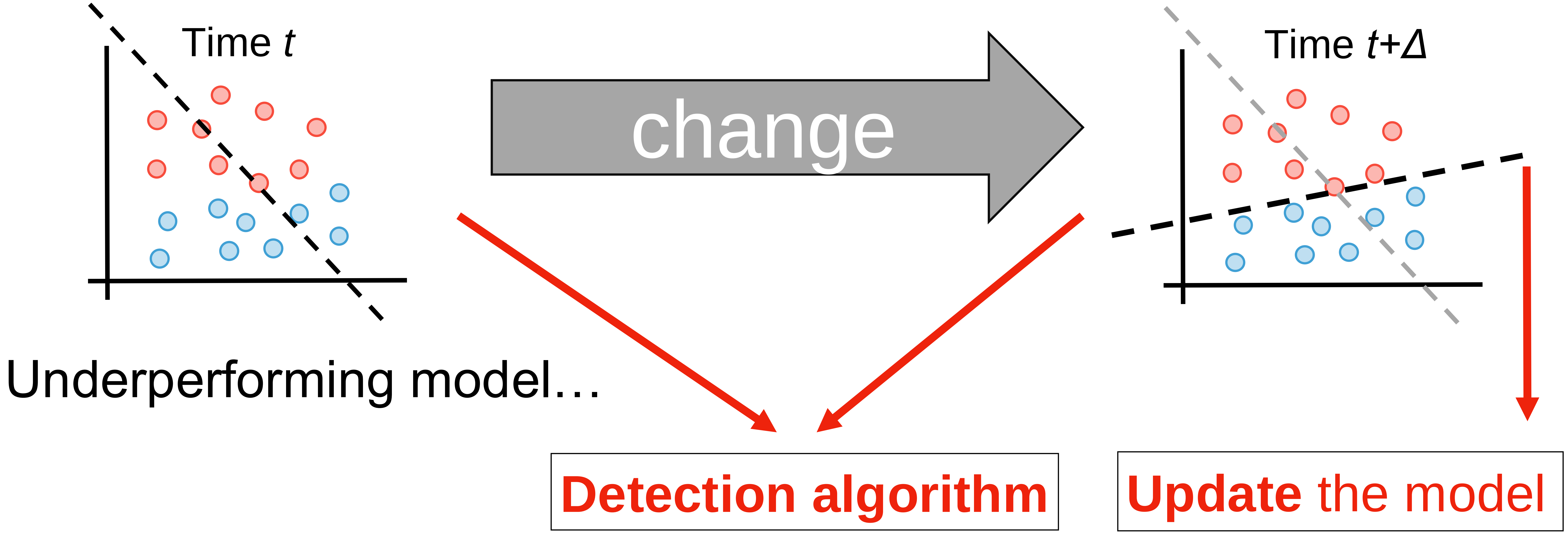
Concept drift

The data distribution may change overtime



Concept drift

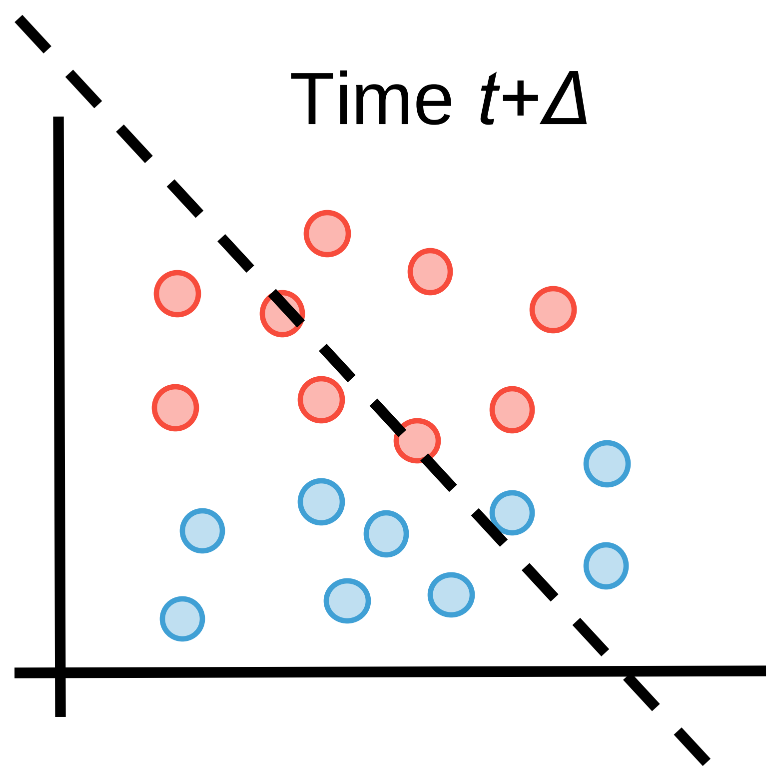
The data distribution may change overtime



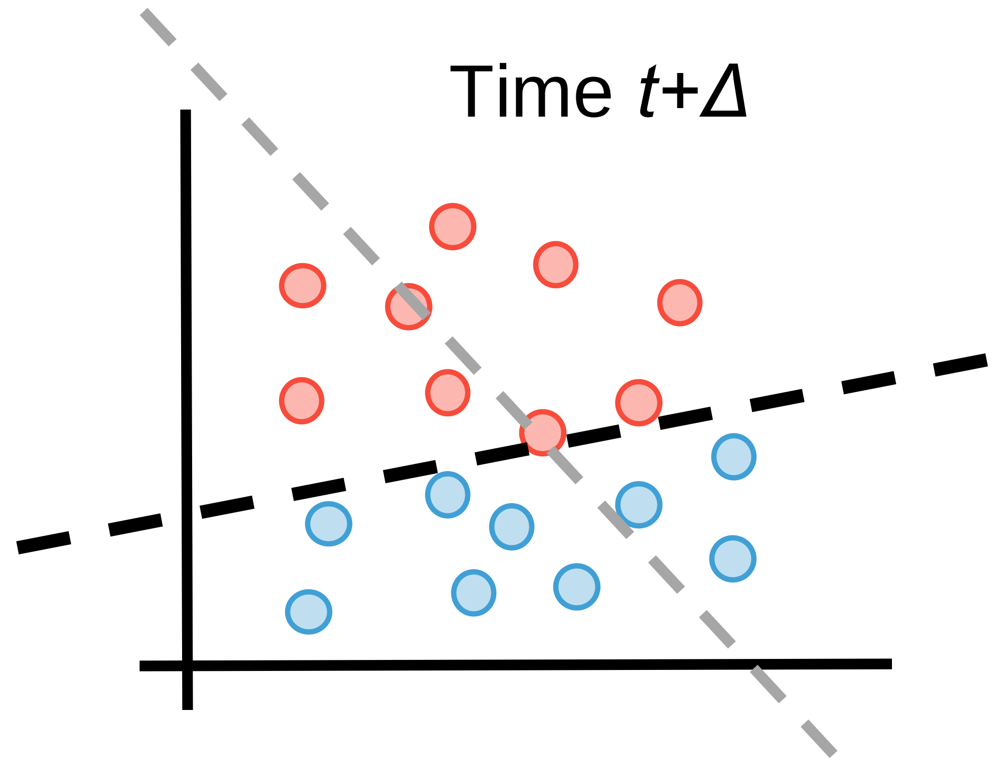
Concept drift

Some questions:

- What **data** should we use to train the updated model?
- How do we **detect** changes?
- What can the detection algorithm observe?

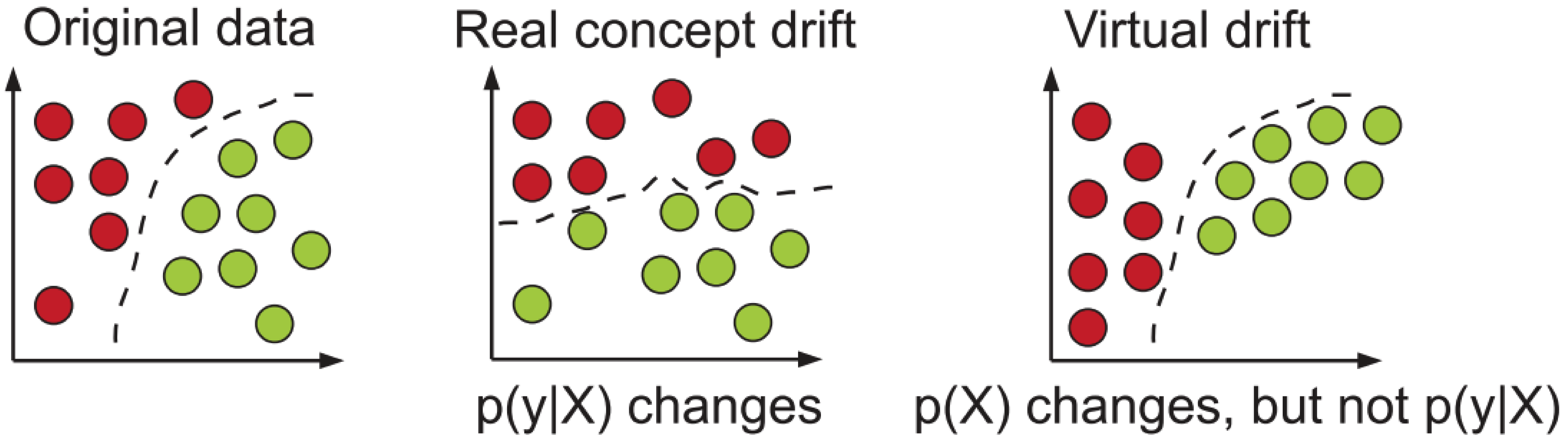


Underperforming model



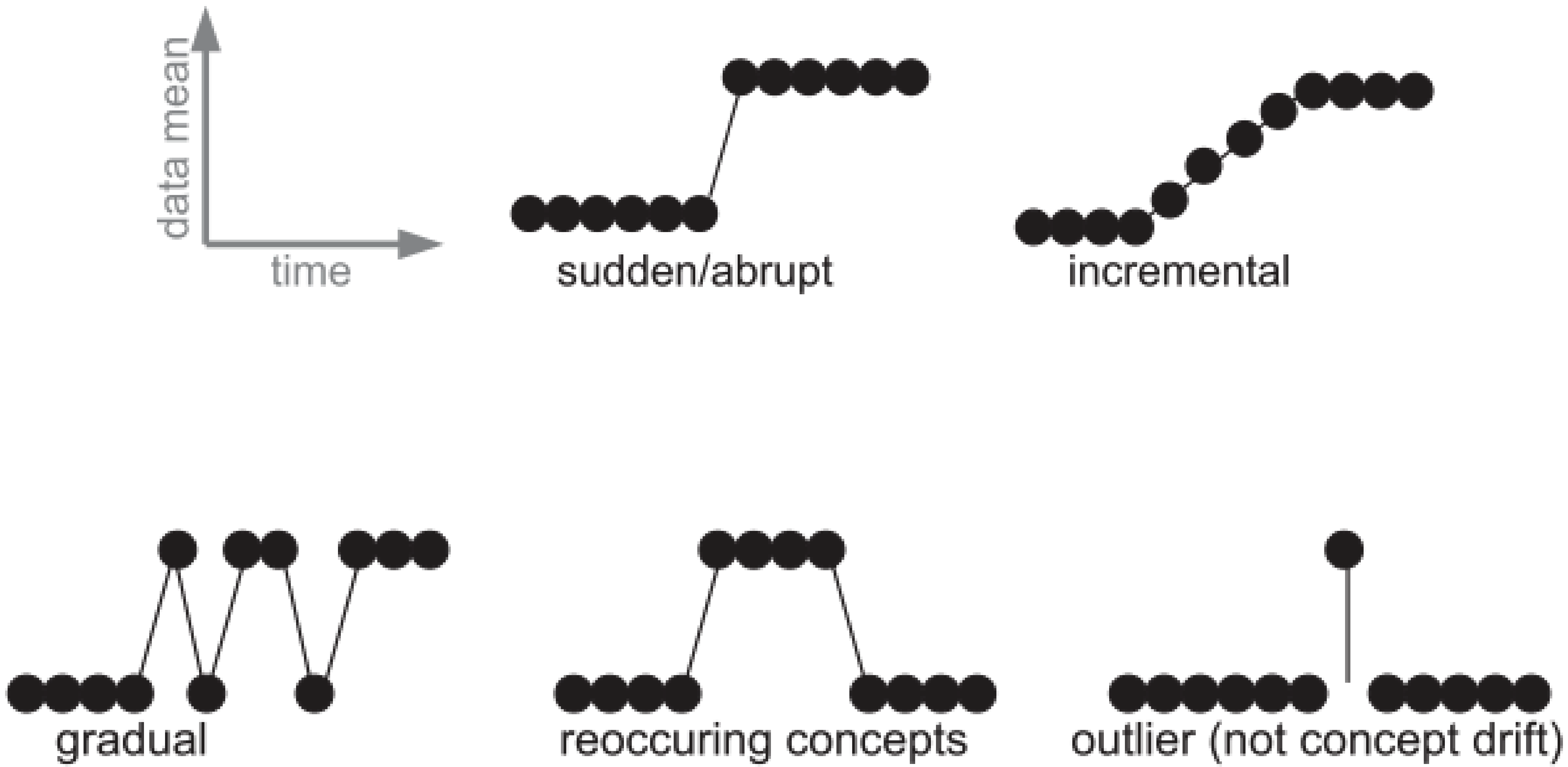
Updated model

Real x Virtual



Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. *ACM computing surveys (CSUR)*

Rate of change



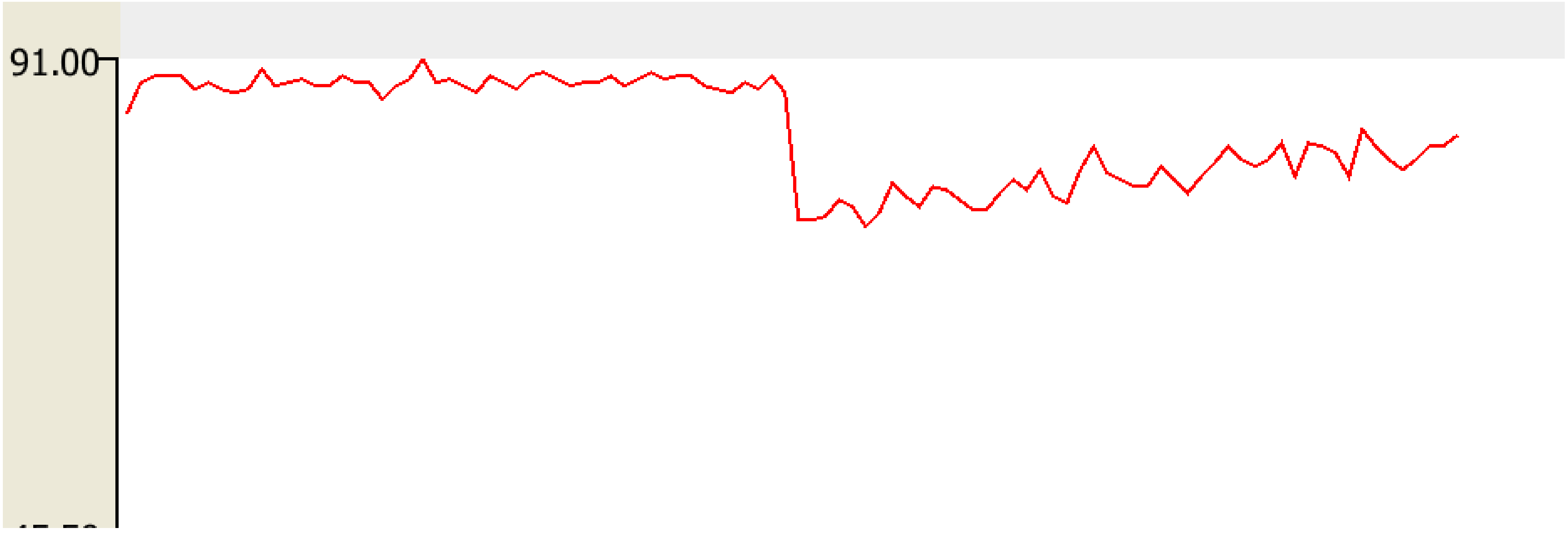
Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. *ACM computing surveys (CSUR)*

Supervised vs. Unsupervised

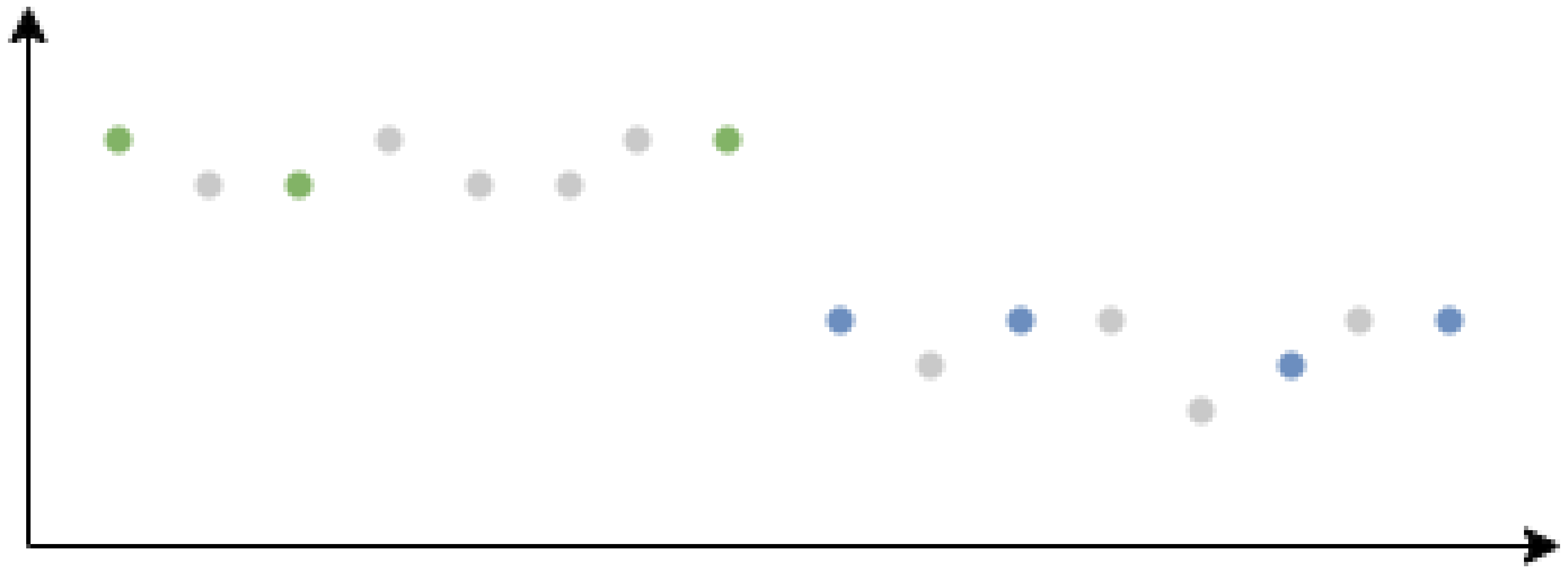
- If labels are **available**, we can monitor metrics such as accuracy
- If accuracy decreases, update model

- If labels are **unavailable**, detecting accuracy changes takes a long time
- Use unsupervised concept drift detector to detect drifts in P(X)

Accuracy



Time



Time

ADaptive WINdow (ADWIN)

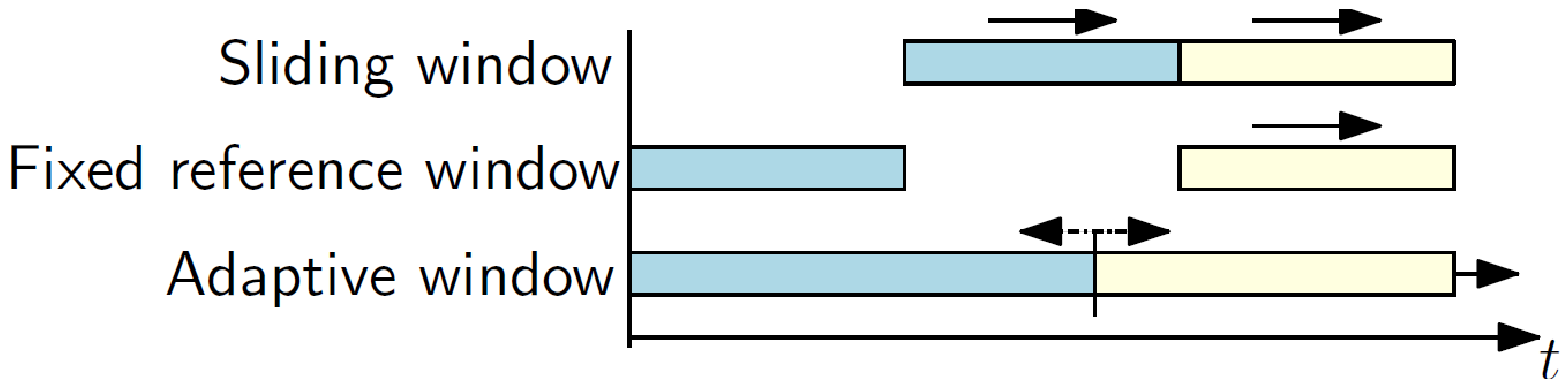
- Univariate
- Window based methods rely on a window that sums up past data and a sliding window summarising recent data
- Statistical tests are used to compare the distribution over the two windows
  - Null hypothesis: the distributions are equal
  - A rejection of the null hypothesis indicates a significant difference between the distributions of these windows (i.e. signals a change has happened)

Bifet, A., & Gavalda, R. (2007). Learning from time-changing data with adaptive windowing. *SIAM international conference on data mining*

ADaptive WINdow (ADWIN)

- Uses **adaptive windows of variable size** that are recalculated online according to the rate of observed change of data in the windows

- Window is **increased** when there is no change, and **decreased** when a change has been detected



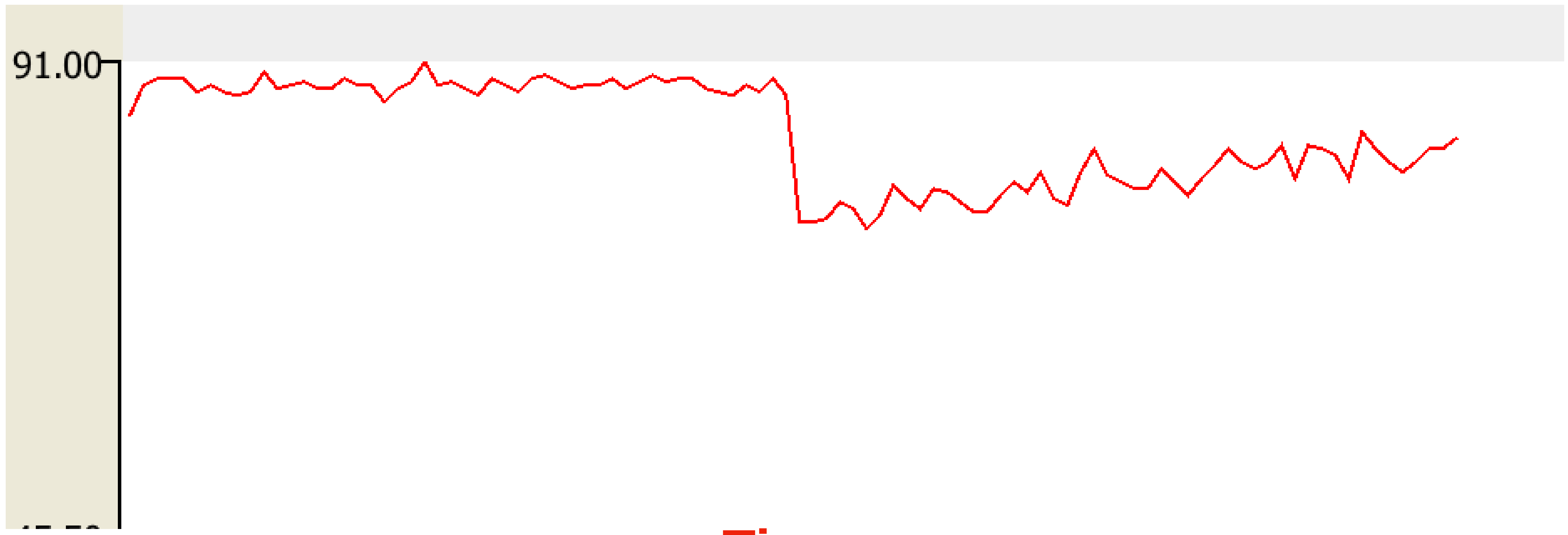
Bifet, A., & Gavalda, R. (2007). Learning from time-changing data with adaptive windowing. *SIAM international conference on data mining*

Evaluation

Evaluating CD Detection

Common approach (proxy): **“Attach the method to a classifier, if the accuracy goes up, then the detector works”**

Accuracy



Time

Not necessarily the detector is successful in detecting changes, maybe it is just randomly resetting the classifier!

We must use specific metrics to evaluate a detector

Evaluating CD Detection

Important: we need the ground-truth of drift location for some of these

Some Metrics:

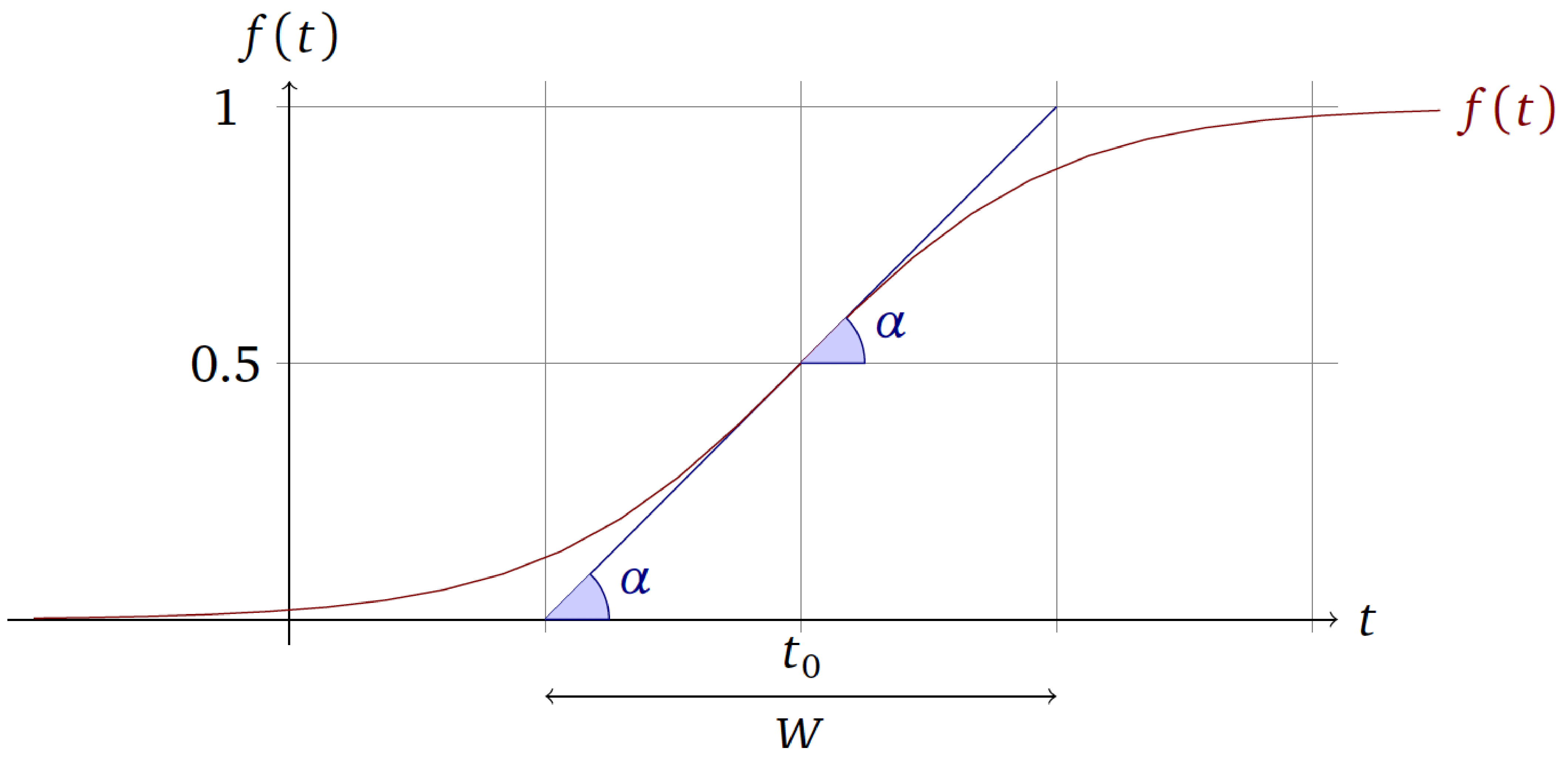
- Mean Time between False Alarms (MTFA)
- Mean Time to Detection (MTD)
- And others: MDR, ARL, MTR, …

Bifet, A. (2017). Classifier concept drift detection and the illusion of progress. In *Artificial Intelligence and Soft Computing ICAISC, 2017*

Simulating CD

Univariate Drift

“Model a concept drift event as a **weighted combination of two pure distribution** that characterizes the target concepts before and after the drift.” [Bifet et al, 2011]



A sigmoid function $f(t) = 1/(1 + e^{-s(t-t_0)})$.

[Bifet et al, 2011] Bifet, A., & Kirkby, R. (2011). Data stream mining a practical approach. Chapter 2.7.1

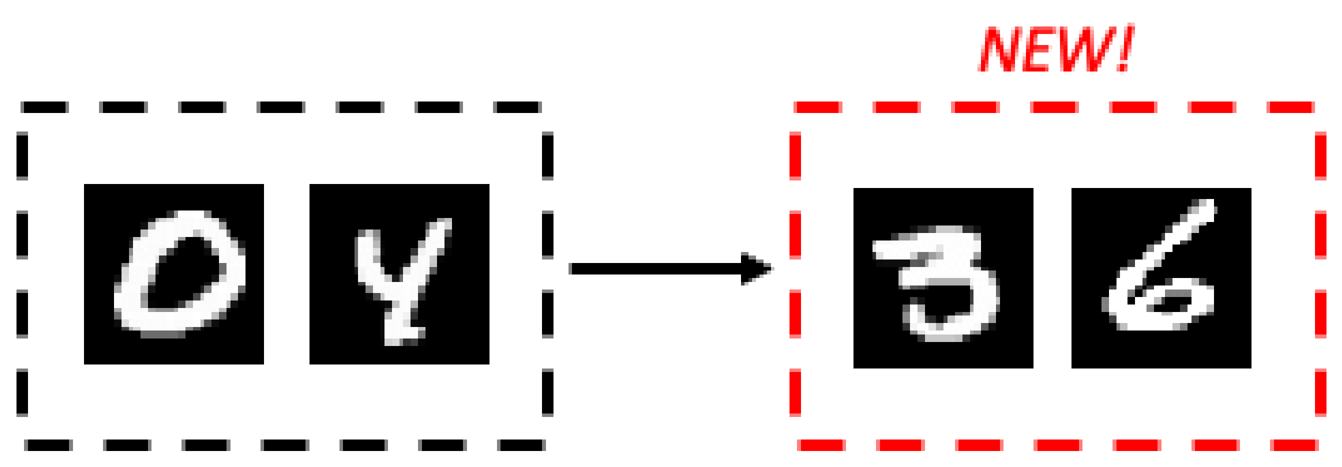
Multivariate Drift

- How can we simulate drift in multivariate data?
- Useful when evaluating unsupervised drift detectors

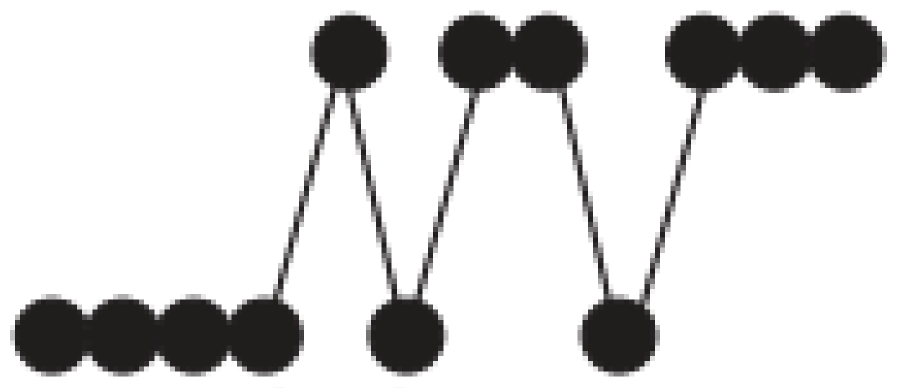
Process:

- Let a subset of classes define each concept
- New concept begins when classes change
- Don’t show the labels to the drift detector!

→ Known change points, drifts of varying difficulty, “complex” drifts



class evolution



gradual

Practical example

03_ECML_2024_drift.ipynb