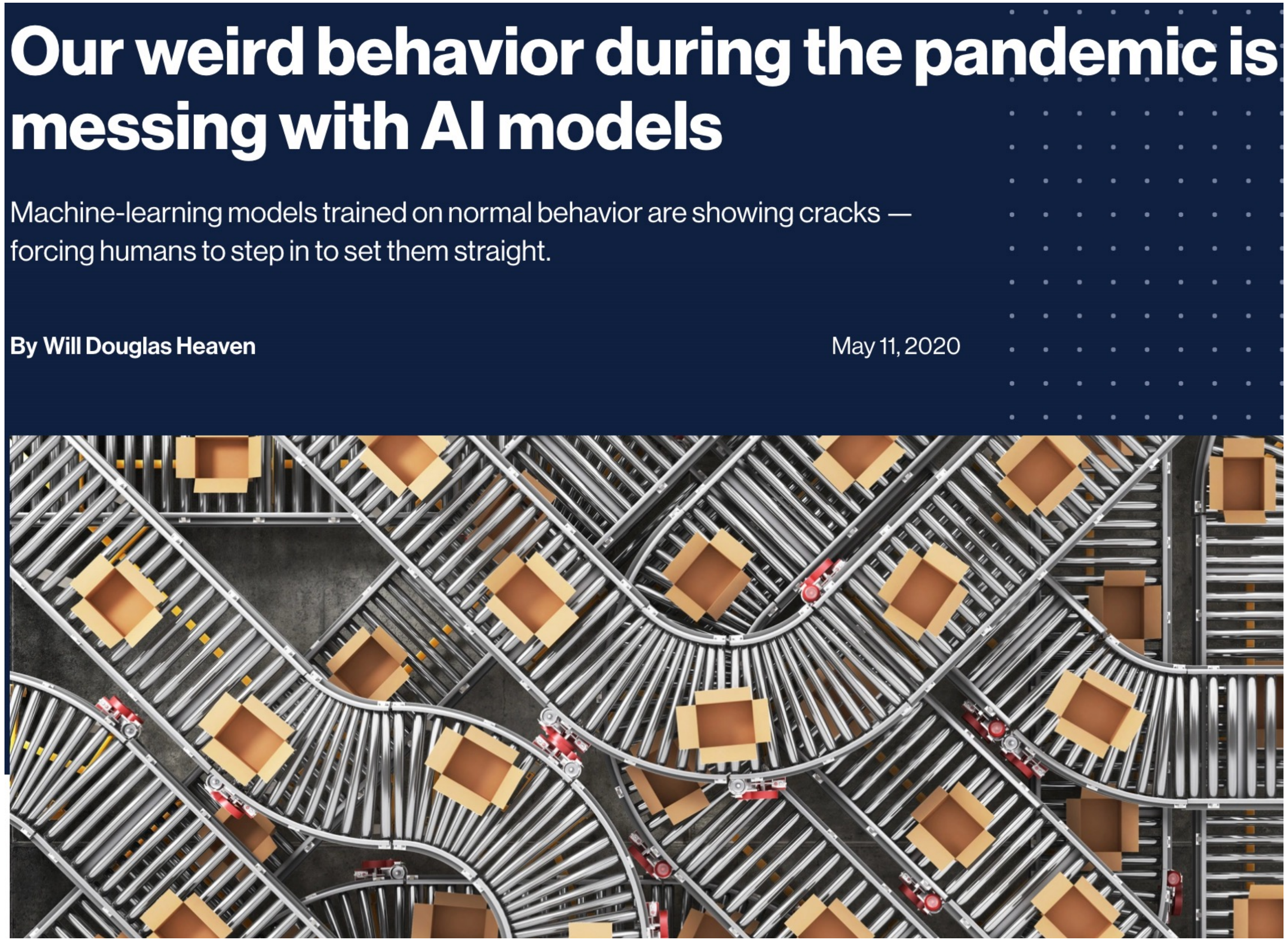
Streaming Machine Learning
Taming Concept Drift

Alessio Bernardo

Post-doc @ Politecnico di Milano

CTO & Co-founder @ Motus ml

POLITECNICO
MILANO 1863



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

Heaven, W. D. **Our weird behavior during the pandemic is messing with AI models**. MIT Technology Review, 2020

12 Data and Analytics Trends to Keep on Your Radar

April 05, 2022 Contributor: Laurence Goasduff

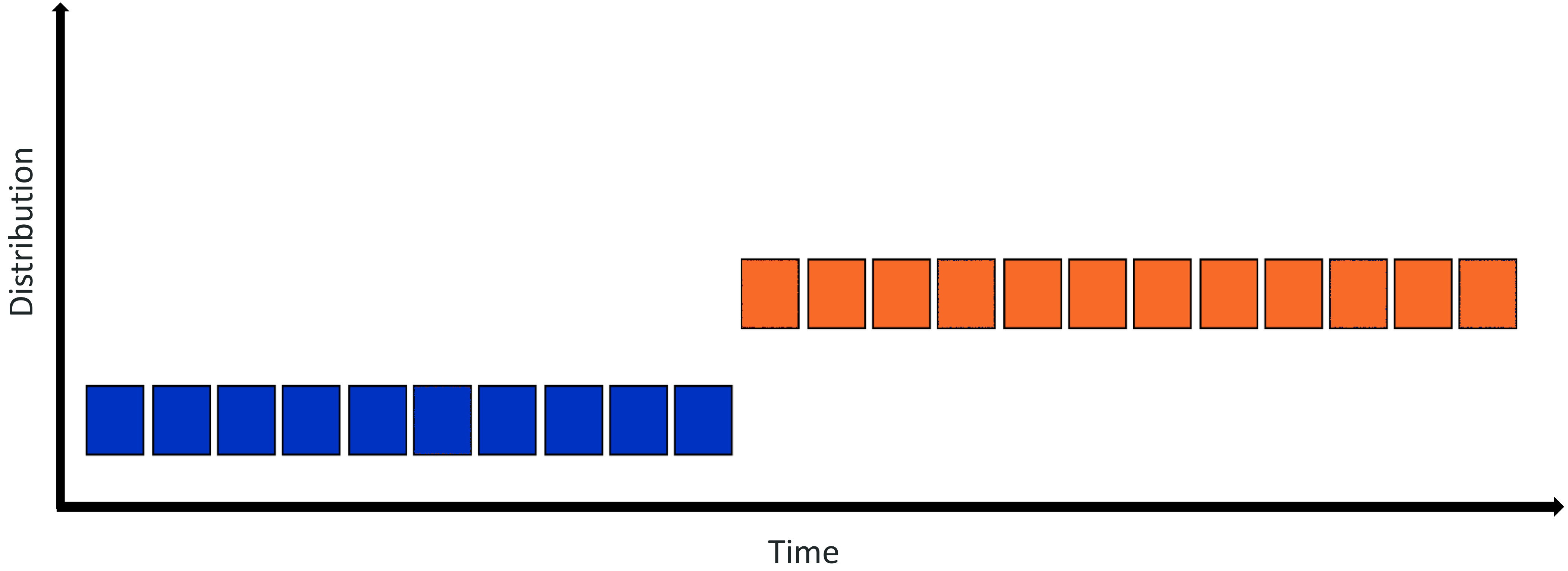
Adaptive artificial intelligence (AI) systems, data sharing and data fabrics are among the trends that data and analytics leaders need to build on to drive new growth, resilience and innovation.

https://www.gartner.com/en/articles/12-data-and-analytics-trends-to-keep-on-your-radar

What is Concept Drift?

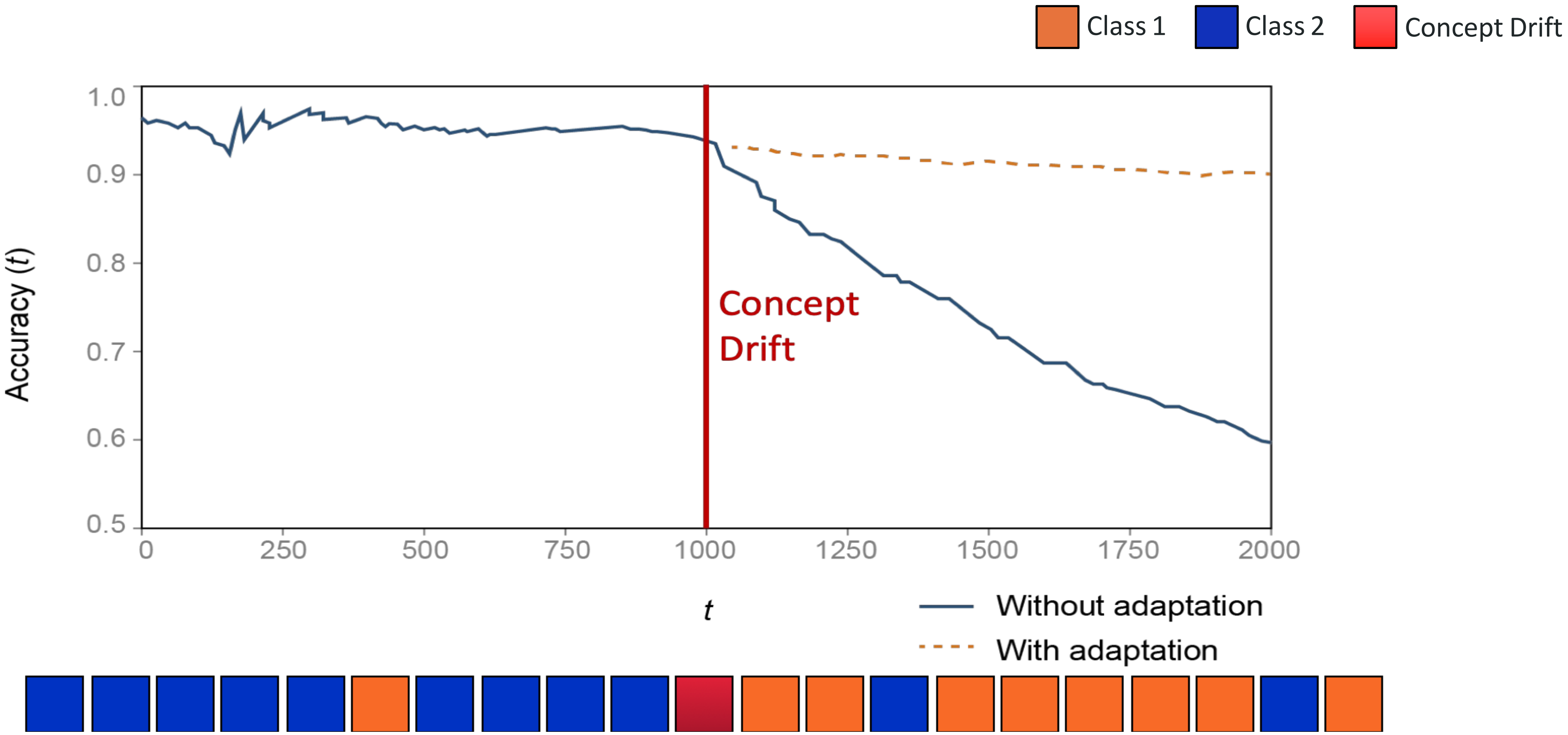
What is a concept drift?

Non identically distributed data



A. Tsymbal **The problem of concept drift: definitions and related work** Trinity College Dublin, 2004

What is a concept drift?



A. Tsymbal **The problem of concept drift: definitions and related work** Trinity College Dublin, 2004

What is a concept drift?

POLITECNICO MILANO 1863

Problem

Given an input sequence $X_1,X_2, \ldots ,X_t$ we want to output at instant t an alarm signal if there is a distribution change and a prediction $\hat{X}_{t+1}$ minimizing the prediction error:

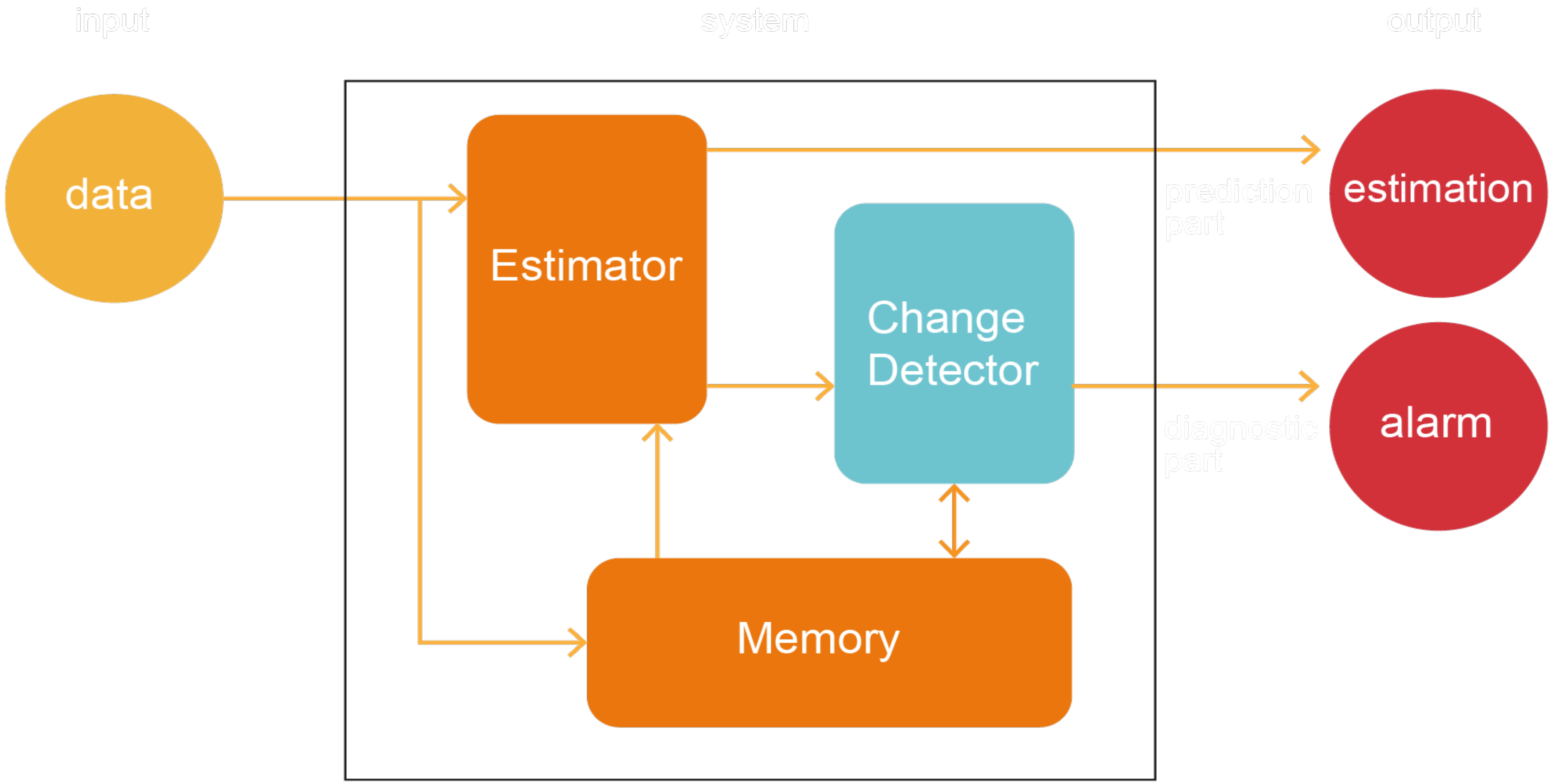
$$\left|\hat{\boldsymbol{X}}_{t+1} - \boldsymbol{X}_{t+1}\right|$$

Outputs

- an estimation of some important parameters of the input distribution, and
- a signal alarm indicating that distribution changes has recently occurred

A. Bifet, R. Gavaldà, G. Holmes, B. Pfahringer **Machine Learning for Data Streams: with Practical Examples in MOA**. The MIT Press (March 2, 2018)

What is a concept drift?

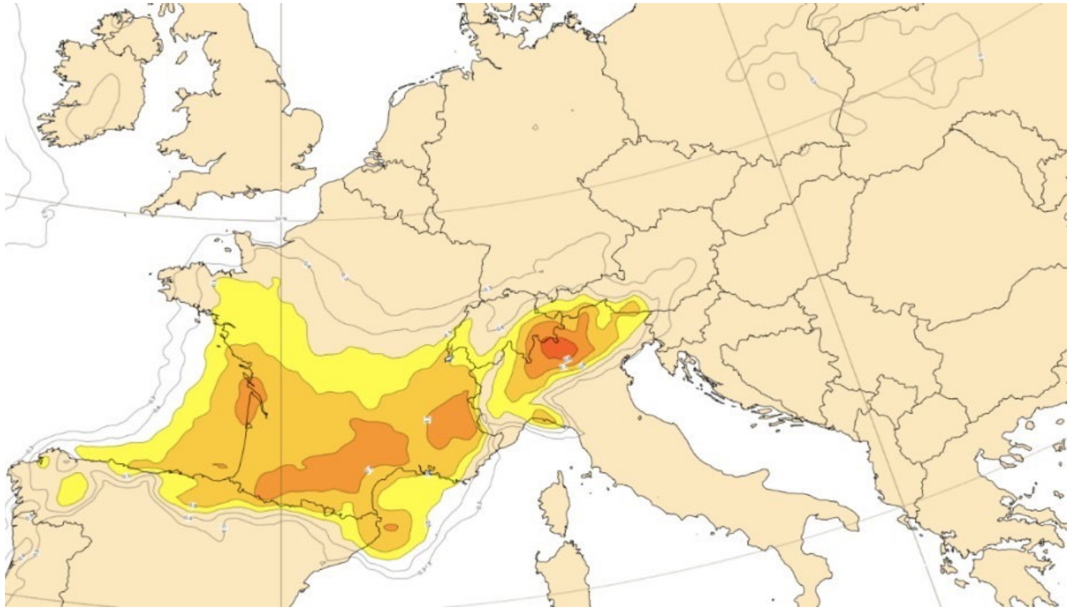


Bifet, A. and Gavaldá, R. **Adaptive Learning from evolving data streams.** In International Symposium on Intelligent Data Analysis (pp.249-260).Springer 2009, August.

What is a concept drift?

Example - Weather forecast

- The chaotic nature of the atmosphere leads to continuous and sudden weather changes (concept drifts)
- Weather forecast models must detect these changes and adapt to them, without be retrained from scratch



Concept Drift Characteristics

Concept Drift Characteristics

Given an input sequence $X_1, X_2, \dots, X_t$ to classify X_t we need to know the prior probability of observing each class, ***p(y)***, and the conditional probability of observing X_t given each class, ***p(X_t|y)***. Using the Bayes' theorem:

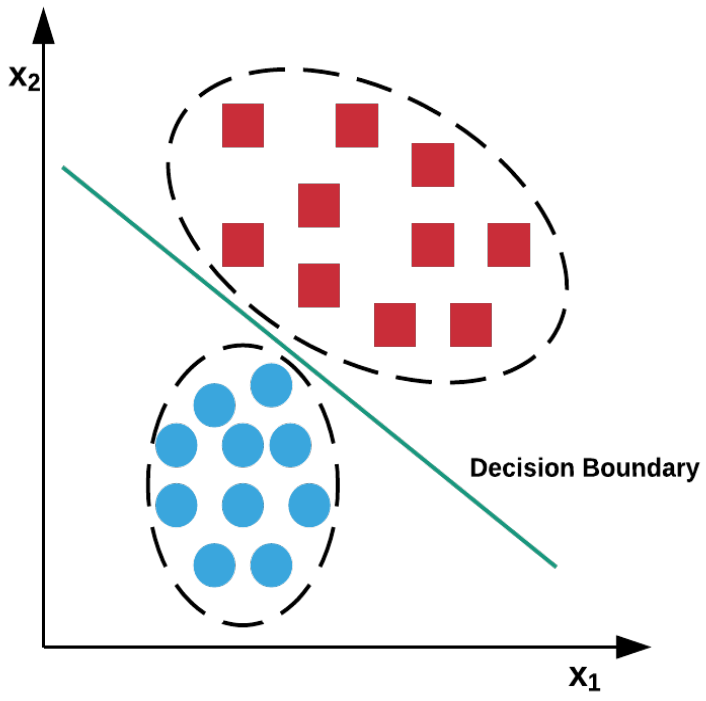
$$p(y|X_t) = \frac{p(y) * p(X_t|y)}{p(X_t)}$$

it is possible to compute the probability that X_t is an instance of class *y*, where ***p(X_t)*** is the probability of observing X_t. Since the latter is constant for all the classes *y*, it can be ignored.

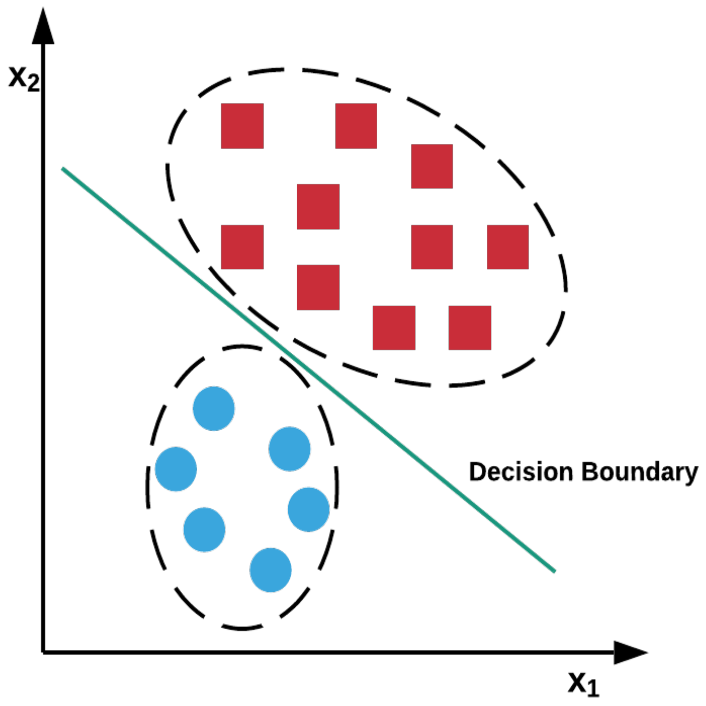
A. Tsymbal **The problem of concept drift: definitions and related work** Trinity College Dublin, 2004

Concept Drift Characteristics

Original distribution



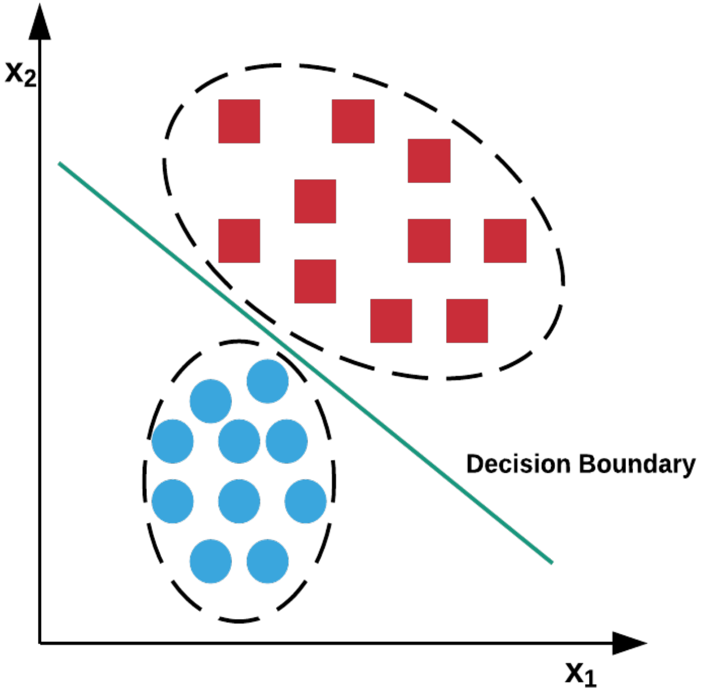
p(y) changes



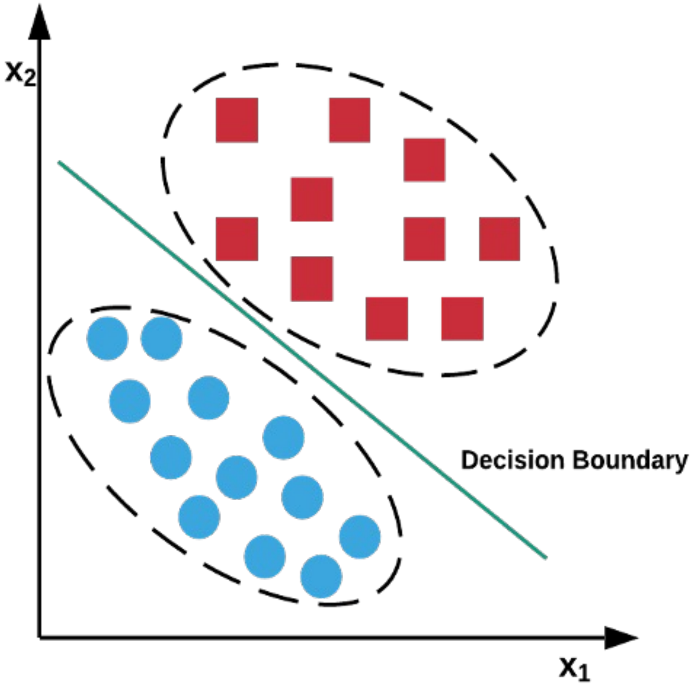
A. Tsymbal **The problem of concept drift: definitions and related work** Trinity College Dublin, 2004

Concept Drift Characteristics

Original distribution



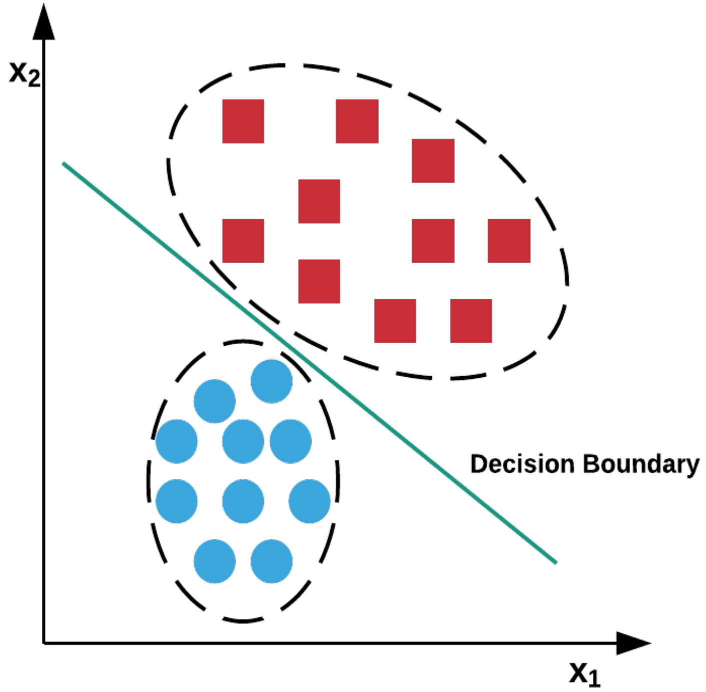
$p(X_t|y)$ changes



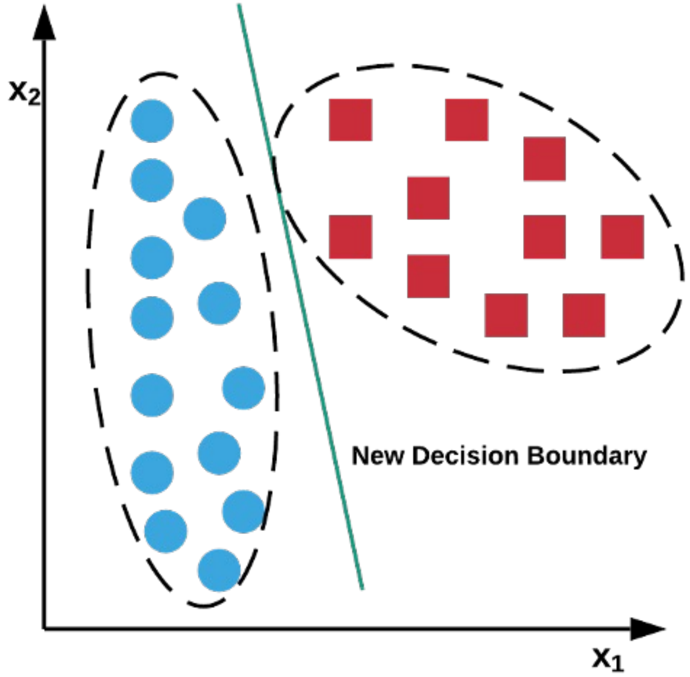
A. Tsymbal **The problem of concept drift: definitions and related work** Trinity College Dublin, 2004

Concept Drift Characteristics

Original distribution



$p(y|X_t)$ **changes**



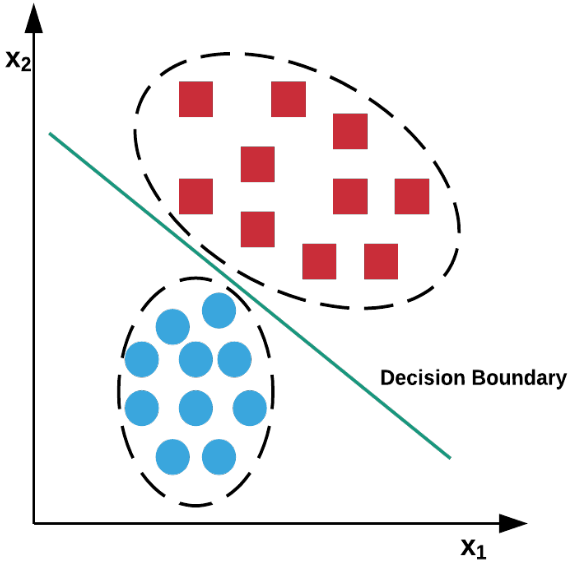
A. Tsymbal **The problem of concept drift: definitions and related work** Trinity College Dublin, 2004

Concept Drift Characteristics

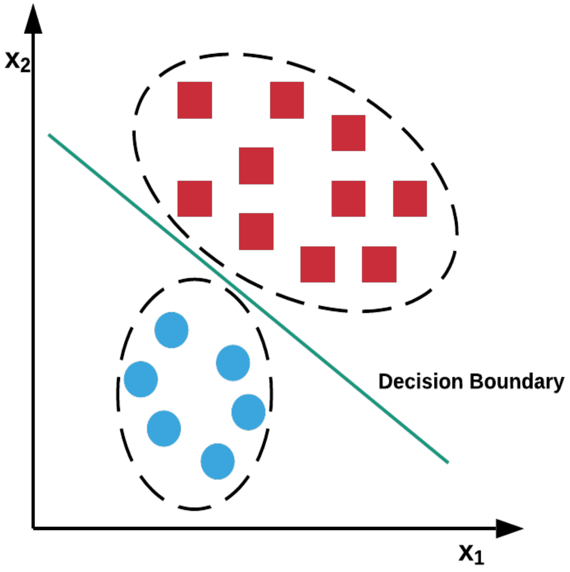
POLITECNICO MILANO 1863

Virtual/Data drift: cases in which only the input distribution changes

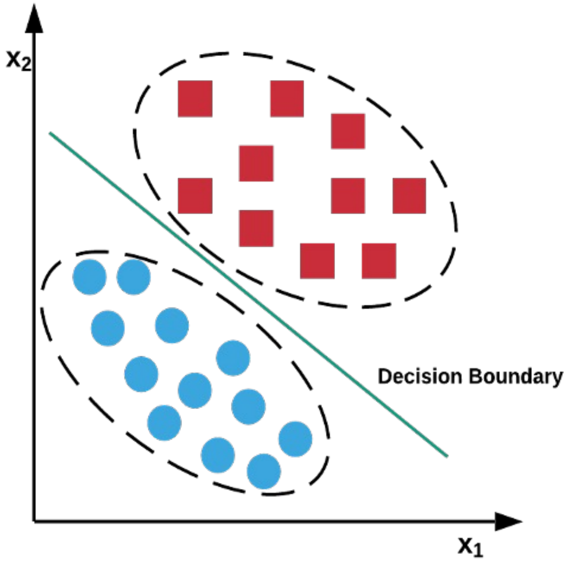
Original distribution



p(y) changes



p(X_t|y) changes

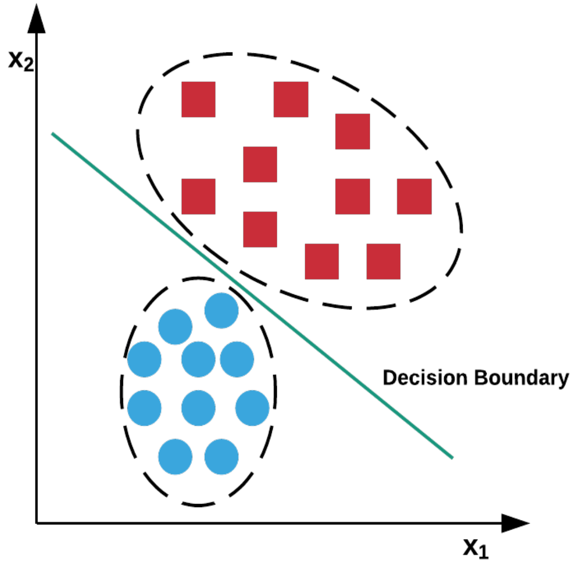


A. Tsymbal **The problem of concept drift: definitions and related work** Trinity College Dublin, 2004

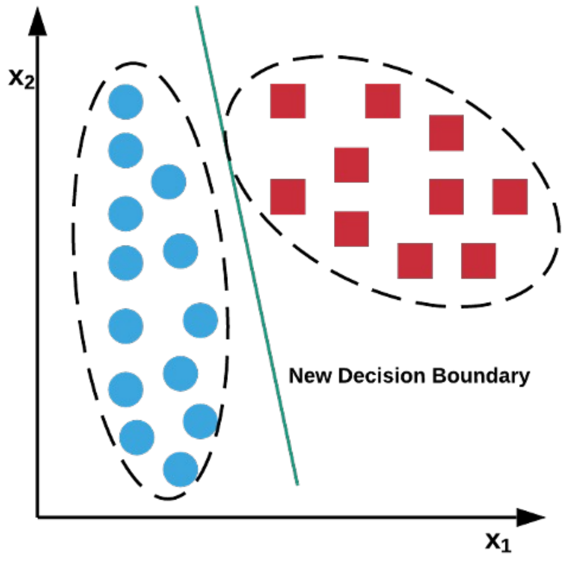
Concept Drift Characteristics

Real/Concept drift: cases in which the change of input distribution causes a boundary shift

Original distribution



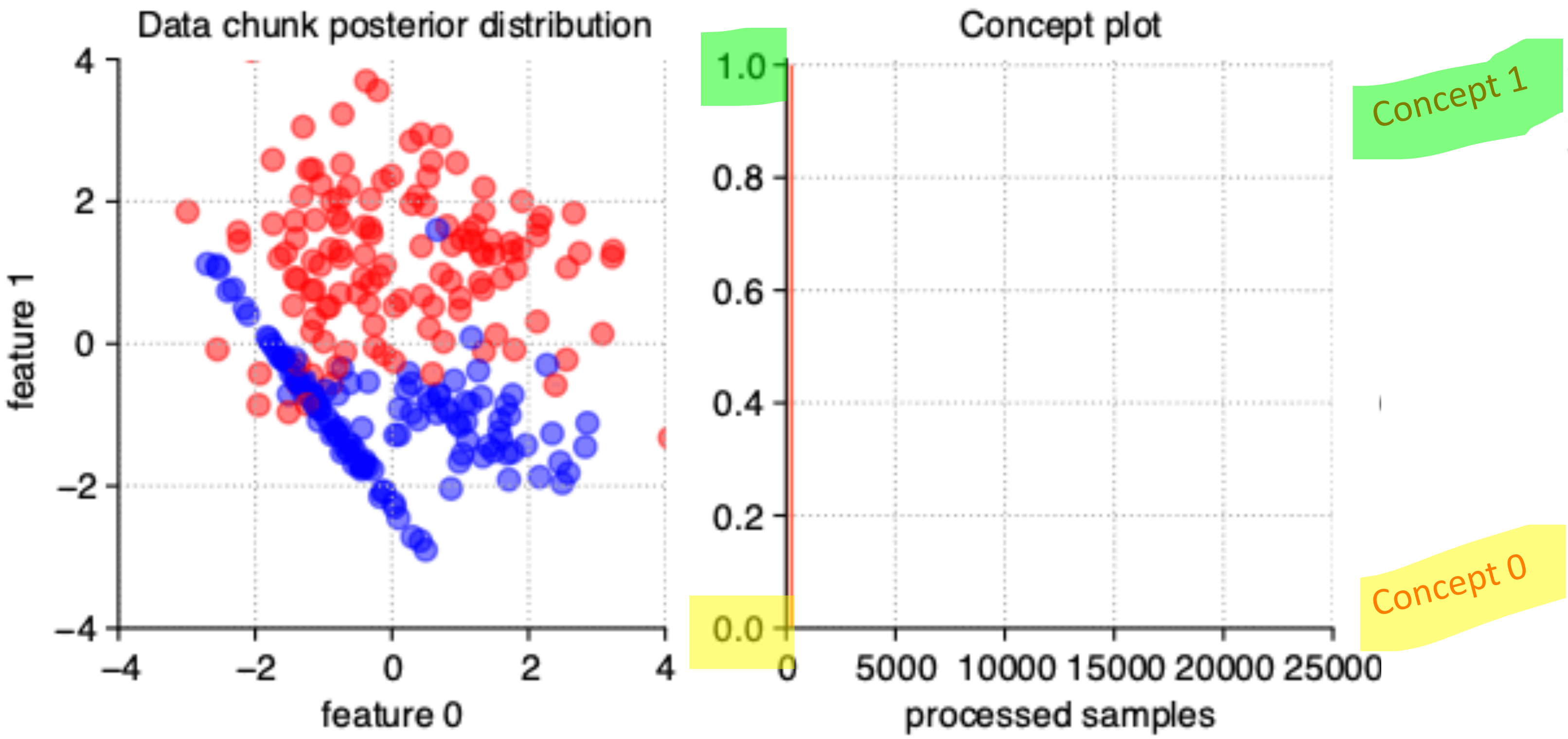
$p(y|X_t)$ changes



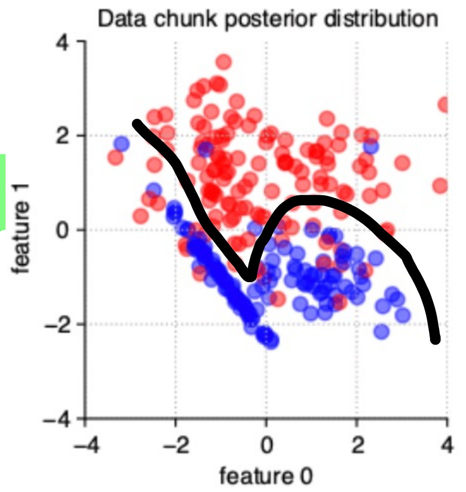
A. Tsymbal **The problem of concept drift: definitions and related work** Trinity College Dublin, 2004

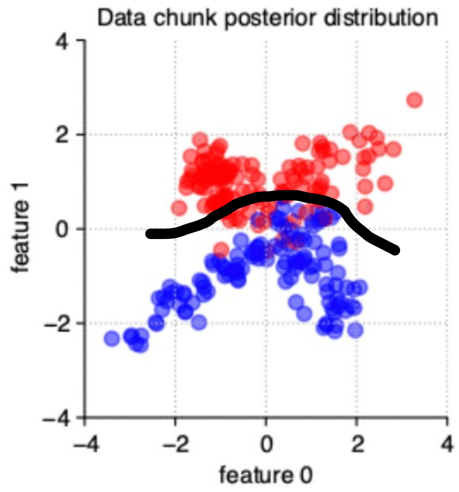
Concept Drift Characteristics

Real/Concept drift: cases in which the change of input distribution causes a boundary shift



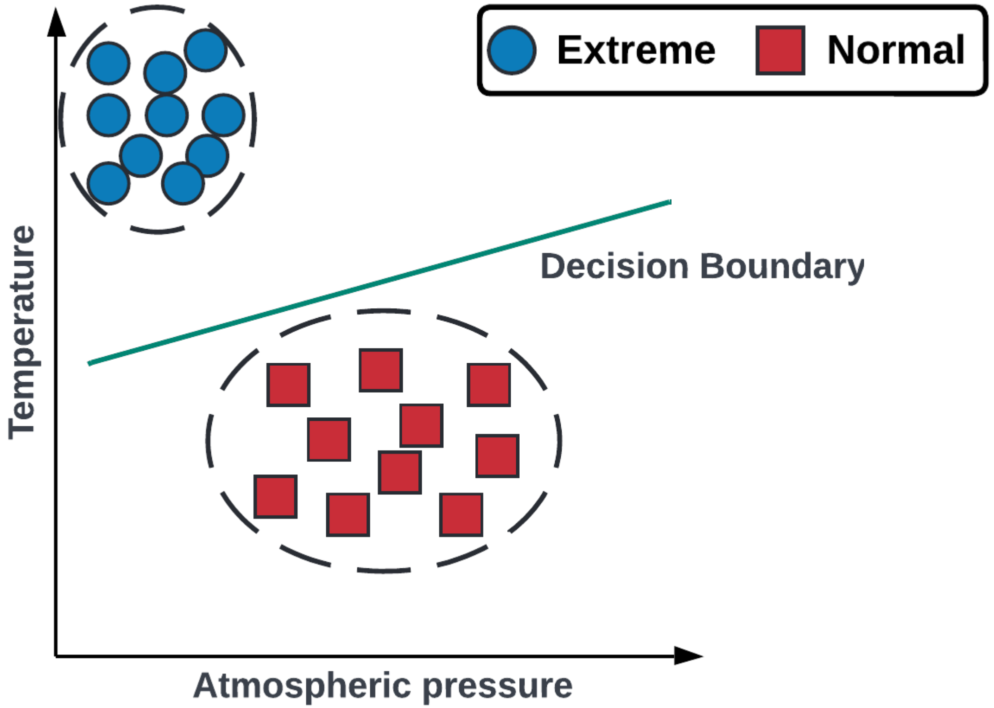
src: https://stream-learn.readthedocs.io/en/latest/_images/incremental.gif





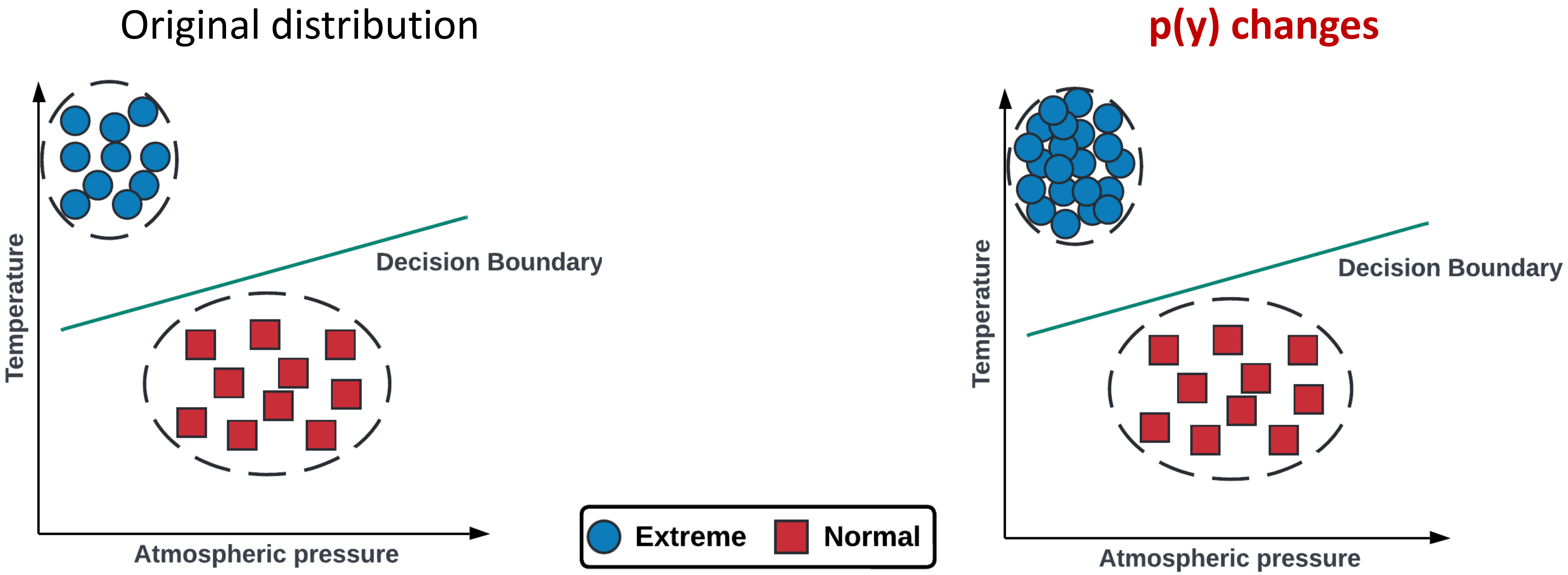
Concept Drift Characteristics

Example: consider the case of predicting extreme weather phenomena occurrences based on atmospheric pressure and temperature. Usually, extreme weather phenomena occur in the case of low atmospheric pressure and high temperature.



Concept Drift Characteristics

p(y) concept drift: in the XX century, the distribution of atmospheric pressure and temperature did not change, but the extreme weather phenomena were more frequent.

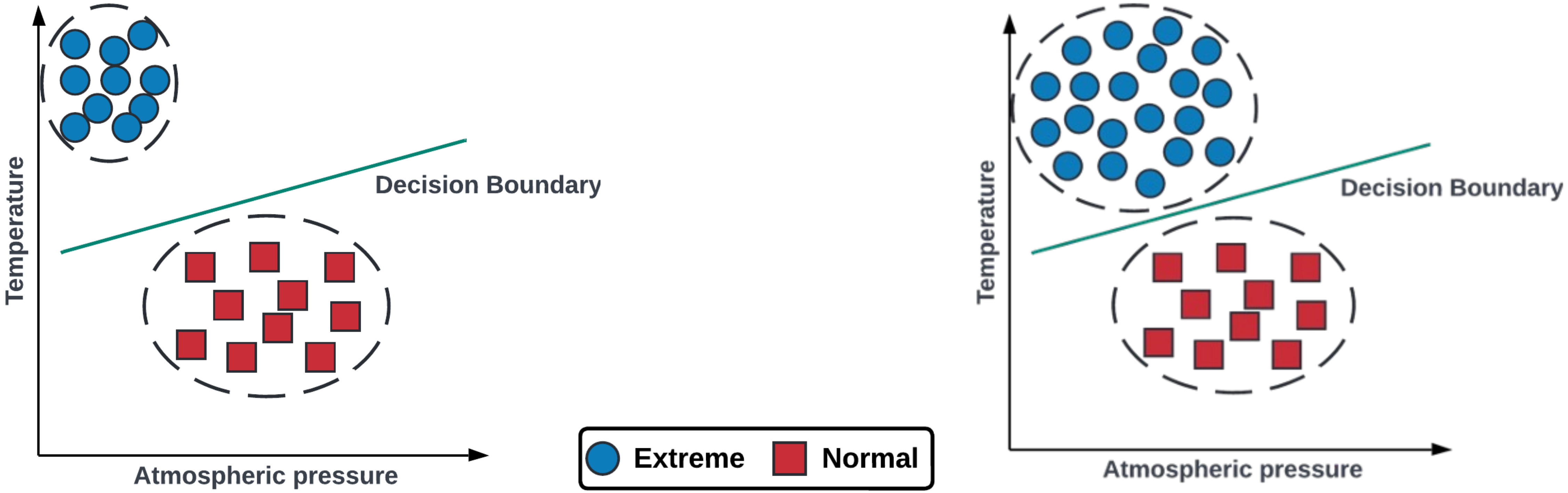


Concept Drift Characteristics

$p(X_t|y)$ concept drift: in the first two decades of XXI century, the atmospheric pressure and air temperature conditions, in which these phenomena occur, also started to change, but not so drastically to move the decision boundary we use for predicting them.

Original distribution

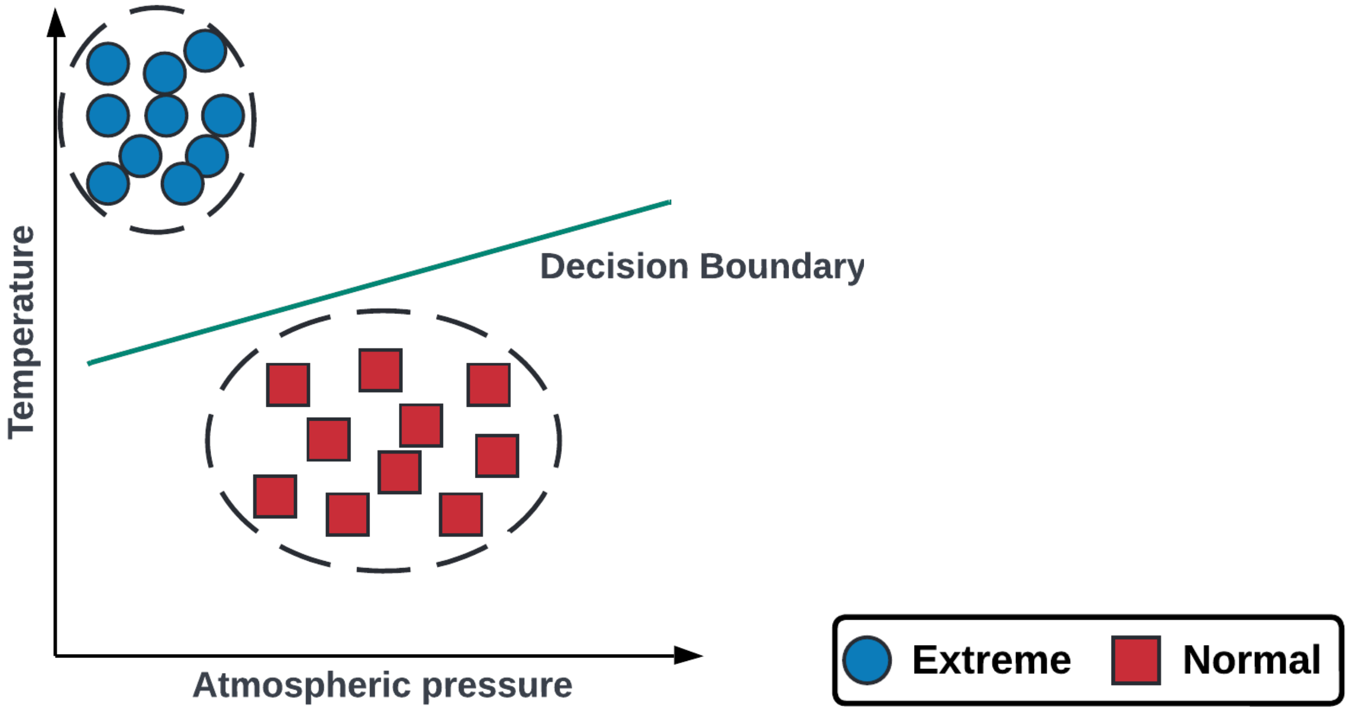
$p(X_t|y)$ changes



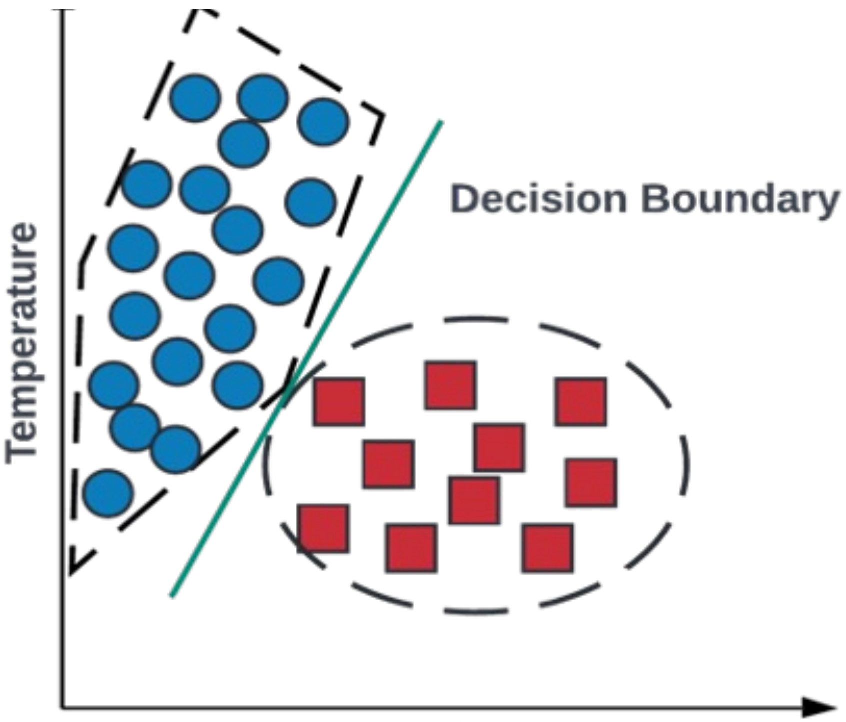
Concept Drift Characteristics

$p(y|X_t)$ concept drift: due to the on-going climate change, these phenomena start occurring more frequently with higher atmospheric pressure and lower temperature. Therefore, we must update the decision boundary to keep a high predictive performance.

Original distribution

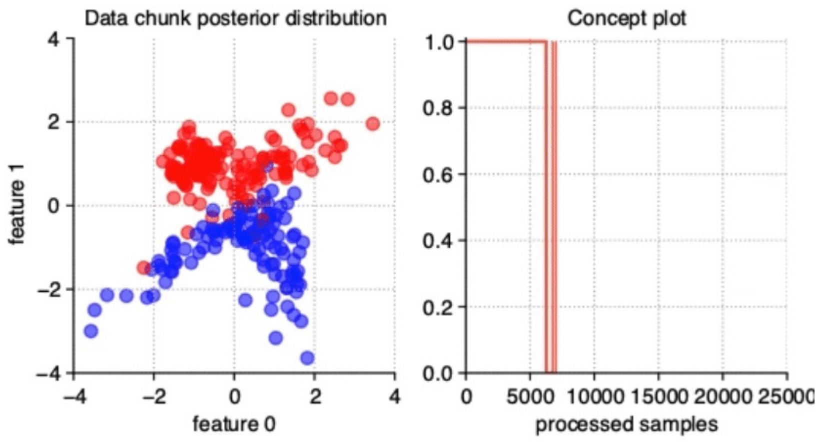


$p(y|X_t)$ changes



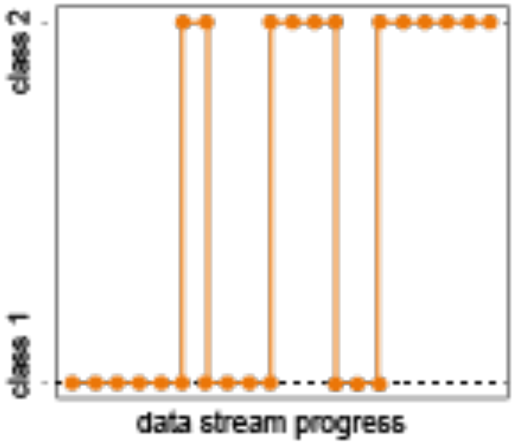
Concept Drift Characteristics

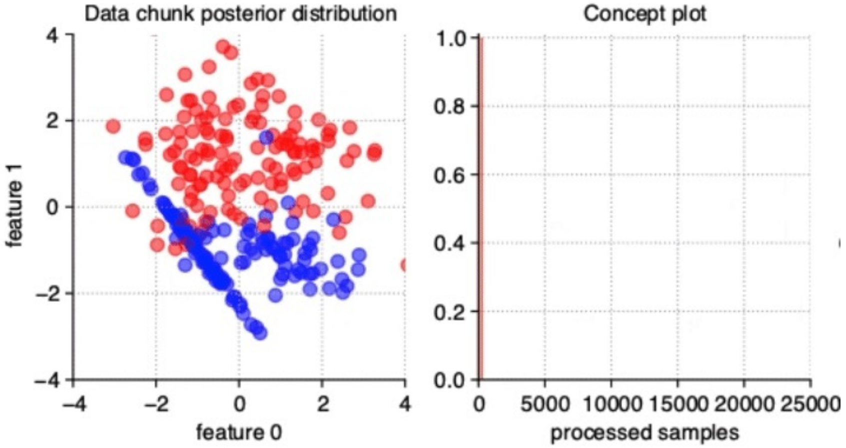
- **Abrupt/Sudden drift**



src: https://stream-learn.readthedocs.io/en/latest/_images/sudden.gif

- **Gradual drift**

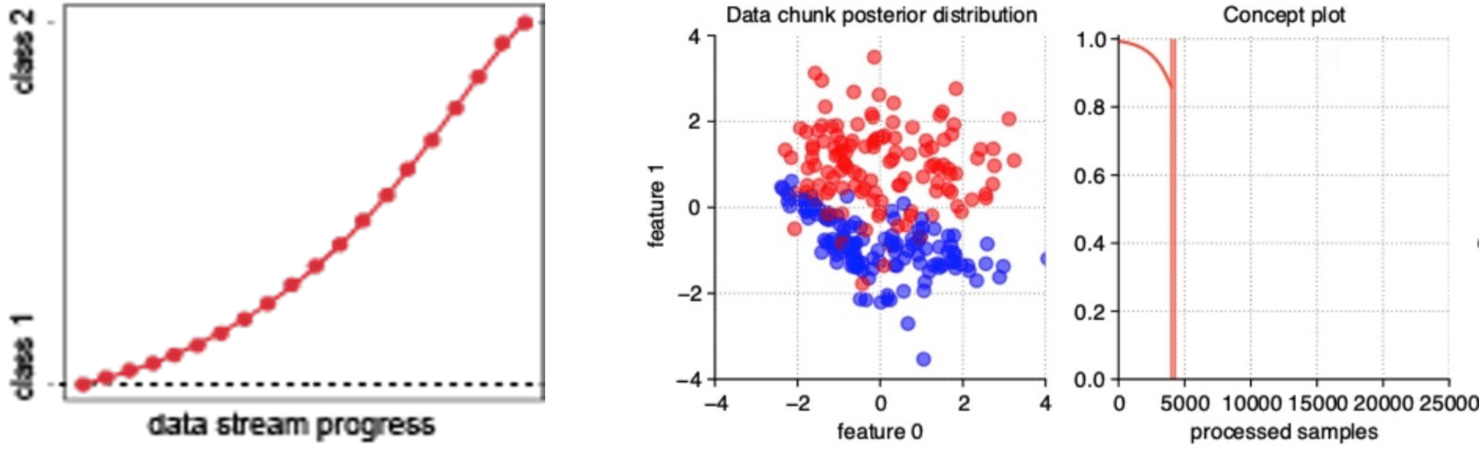




src: https://stream-learn.readthedocs.io/en/latest/_images/gradual.gif

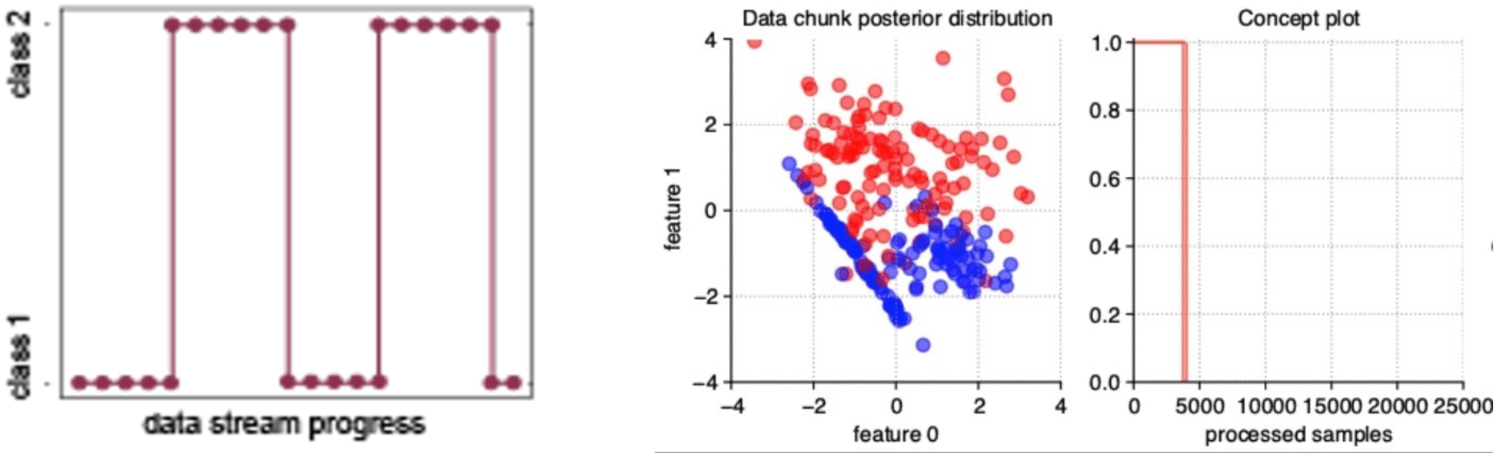
Concept Drift Characteristics

- **Incremental drift**



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

- **Recurring drift**



src: https://stream-learn.readthedocs.io/en/latest/_images/recurring.gif

Concept Drift Characteristics

- **Feature drift:** it occurs when a subset of features are not relevant anymore to learn the underline concept. Consequently, if a model strongly relies on that subset to predict the new samples, the prediction can be wrong
- **Feature evolution:** it occurs when new features appear/disappear over time or the set of all possible values of a feature changes. If a new feature becomes available and deemed relevant, the model should be able to incorporate it into its learning process. Similarly, if a feature becomes unavailable, the learning model should ignore its existence
- **Concept evolution:** it occurs when new class labels appear/disappear. This is natural in some domains, such as Intrusion Detection Systems, where new threats continuously appear as attackers always introduce new strategies

H. M. Gomes, et al. **Machine learning for streaming data: state of the art, challenges, and opportunities.** SIGKDD Explor, 2019

Concept Drift vs Anomaly Detection

Concept Drift question: "Is yesterday's model capable of explaining today's data?"

Anomaly detection question: "Do these samples conform the normal ones?"

A. Tsymbal **The problem of concept drift: definitions and related work** Trinity College Dublin, 2004

Concept Drift Detectors

Concept Drift Detectors

POLITECNICO MILANO 1863

Monitoring the input distribution

Pro:

- Does not require supervised samples

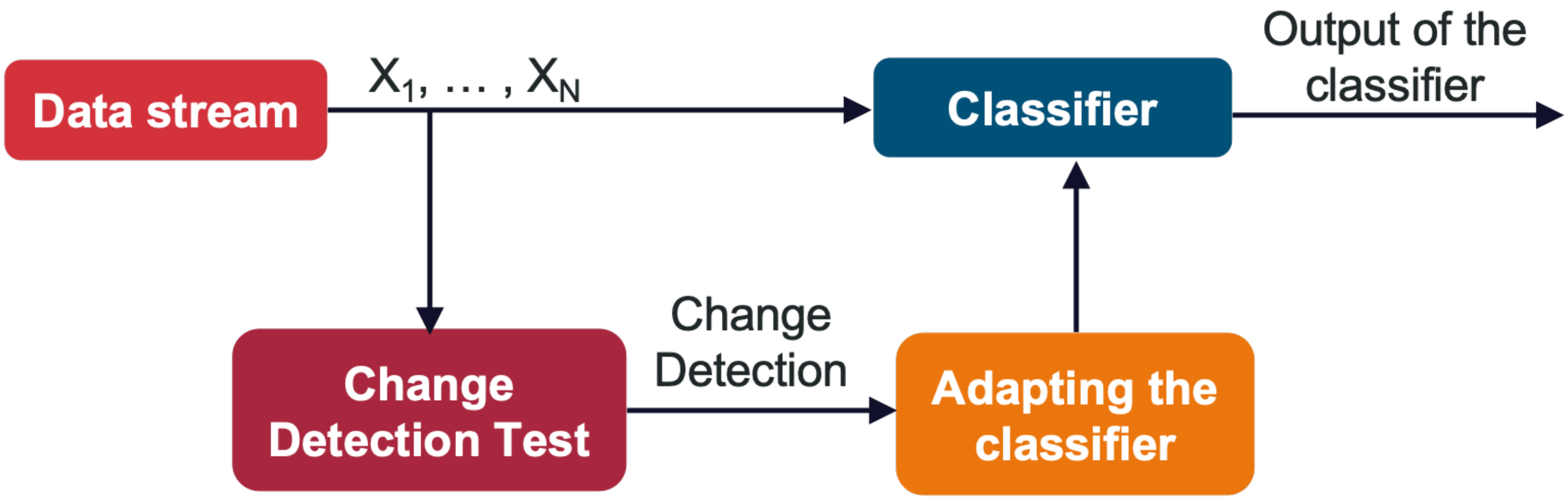
Cons:

- Difficult to design sequential detection tool, i.e., change detection tests when streams are multivariate and distribution unknown
- It does not detect changes that do not affect the distribution of observations

Gama, João, et al. **A survey on concept drift adaptation**. ACM computing surveys, 2014

Concept Drift Detectors

Monitoring the input distribution



Gama, João, et al. **A survey on concept drift adaptation**. ACM computing surveys, 2014

Concept Drift Detectors

Monitoring the input distribution - Cumulative SUM Test (CUSUM)

- It gives an alarm when the mean of the input data is significantly different from zero.
- It is memoryless, and its accuracy depends on the choice of parameters *v* and *h*.
- It is a one-sided test: it assumes that changes can happen only in one direction of the statistics, detecting only increases.

$$g_0 = 0$$
$$g_t = \max(0, g_{t-1} + (x_t - \hat{x}) - v)$$
$if\ g_t > h$ then Alarm

Lee, S., Ha, J., Na, O., & Na, S. **The cusum test for parameter change in time series models**. Scandinavian Journal of Statistics, 2003

Concept Drift Detectors

POLITECNICO MILANO 1863

Monitoring the input distribution - Page Hinkley Test

- It is designed to detect a change in the average of a Gaussian signal and monitors the difference between g_t and G_t.
- Its accuracy depends on the choice of parameters *v* and *h*.

$$g_0 = 0$$
$$g_t = g_{t-1} + (x_t - \hat{x}) - v$$
$$G_t = \min(g_t, G_{t-1}) \ *$$
$$if\ g_t - G_t > h \text{ then Alarm } *$$

* When the signal is decreasing, we should use:

$$G_t = \max(g_t, G_{t-1})$$
$$if\ G_t - g_t > h \text{ then Alarm}$$

Mouss, H., et al. **Test of page-hinckley, an approach for fault detection in an agro-alimentary production system**. IEEE Asian Control Conference, 2004

Concept Drift Detectors

Monitoring the classification error

Pro:

- the most straightforward figure of merit to monitor
- changes in p_t prompt adaptation only when performances are affected

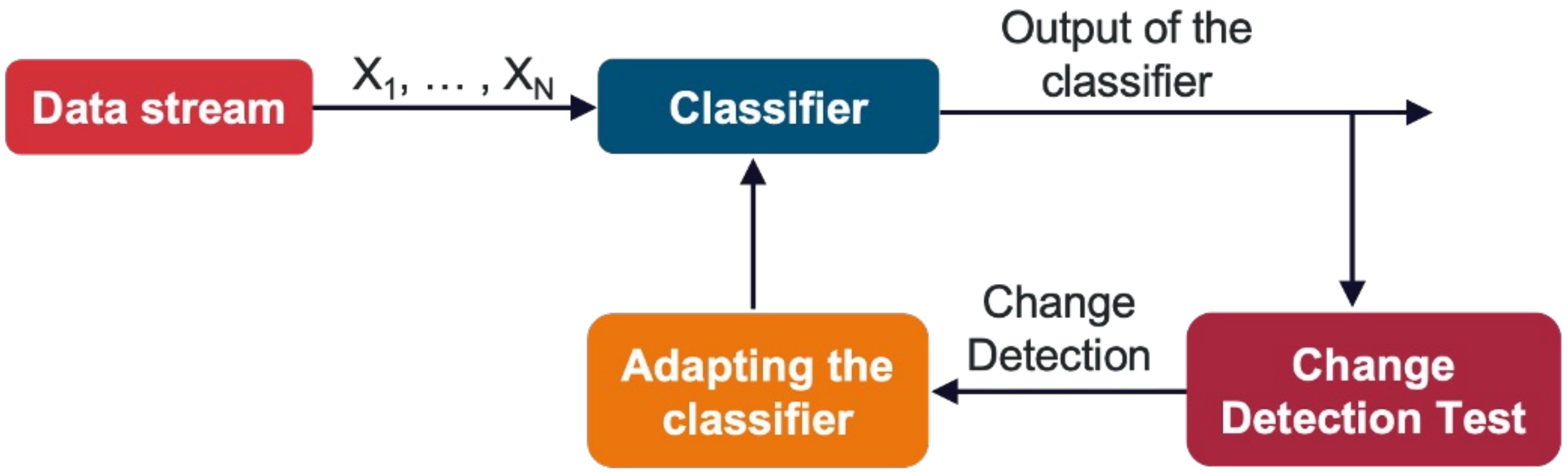
Cons:

- Concept drift detection from supervised samples only

Gama, João, et al. **A survey on concept drift adaptation**. ACM computing surveys, 2014

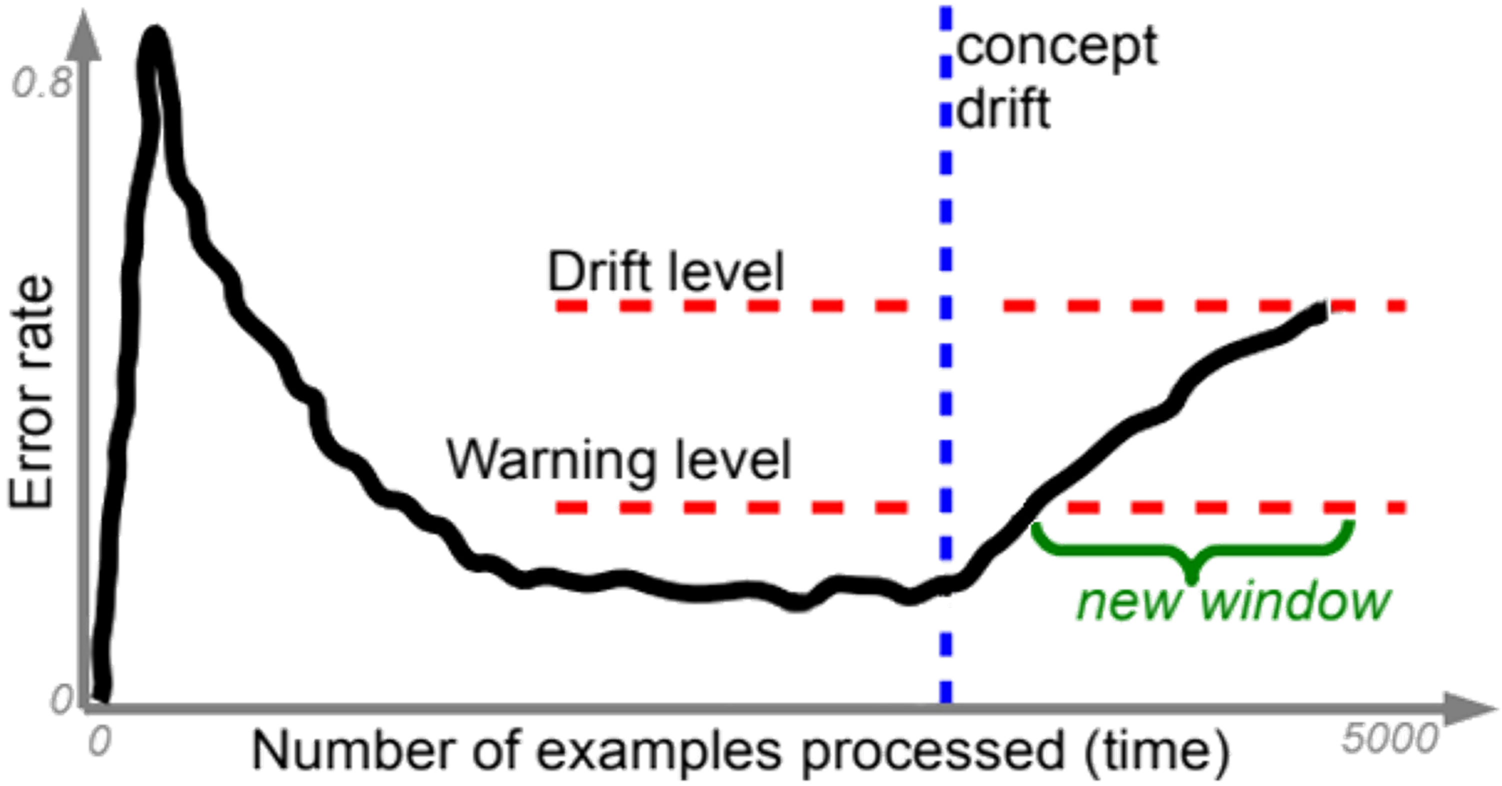
Concept Drift Detectors

Monitoring the classification error



Gama, João, et al. **A survey on concept drift adaptation**. ACM computing surveys, 2014

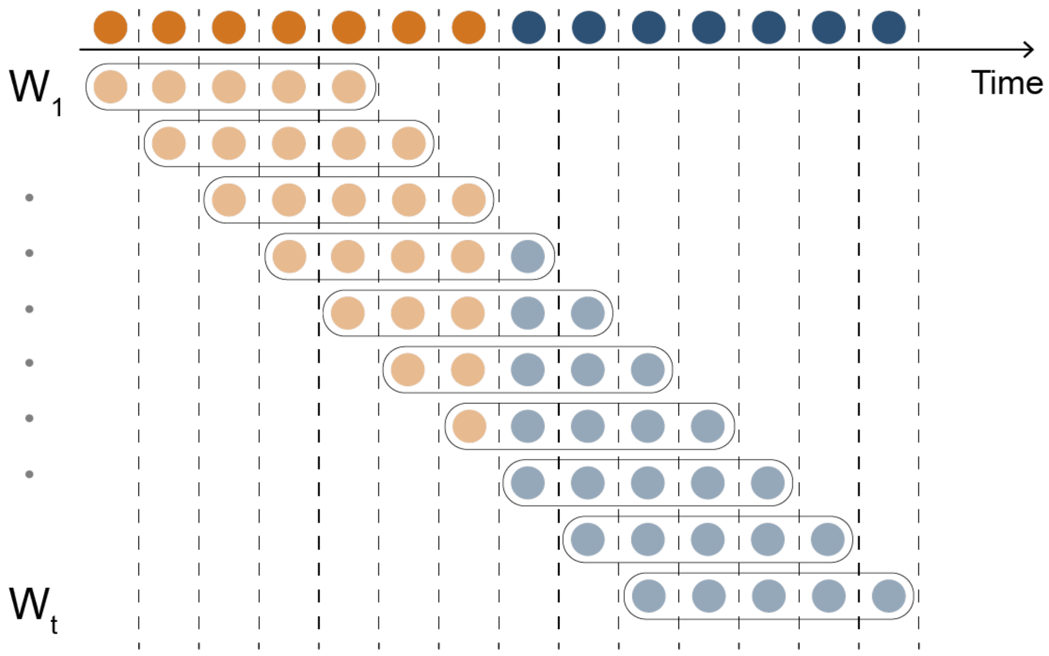
Concept Drift Detectors



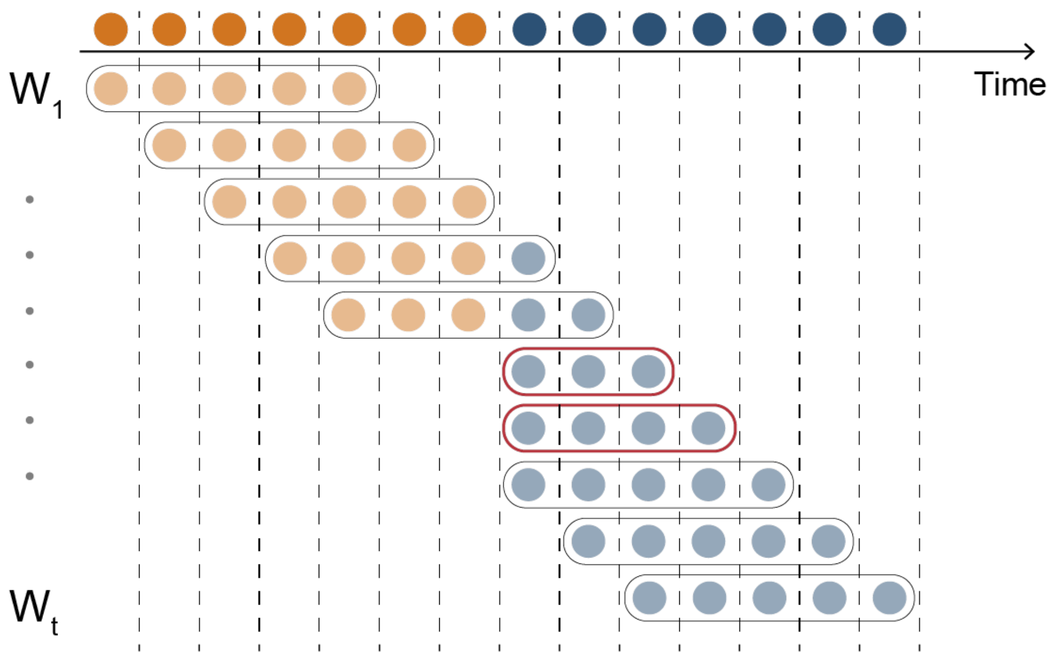
Gama, et. al, **Learning with Drift Detection**, SBIA, Springer, 2004

Concept Drift Detectors

Passive



Active

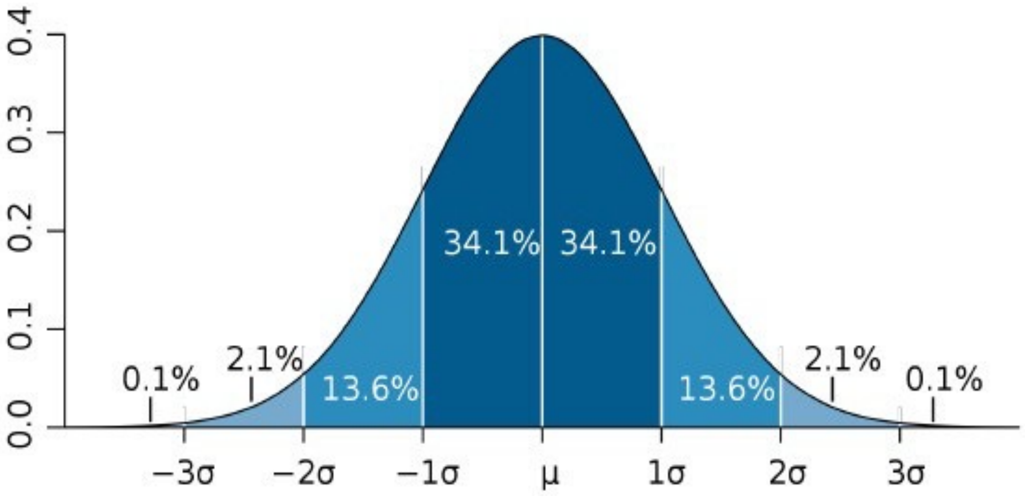


Gama, João, et al. **A survey on concept drift adaptation**. ACM computing surveys, 2014

Concept Drift Detectors

Monitoring the classification error - Drift Detection Method (DDM)

- Detect concept drift as an outlier in the classification error.
- In stationary conditions error decreases, so look for outliers in the tails.

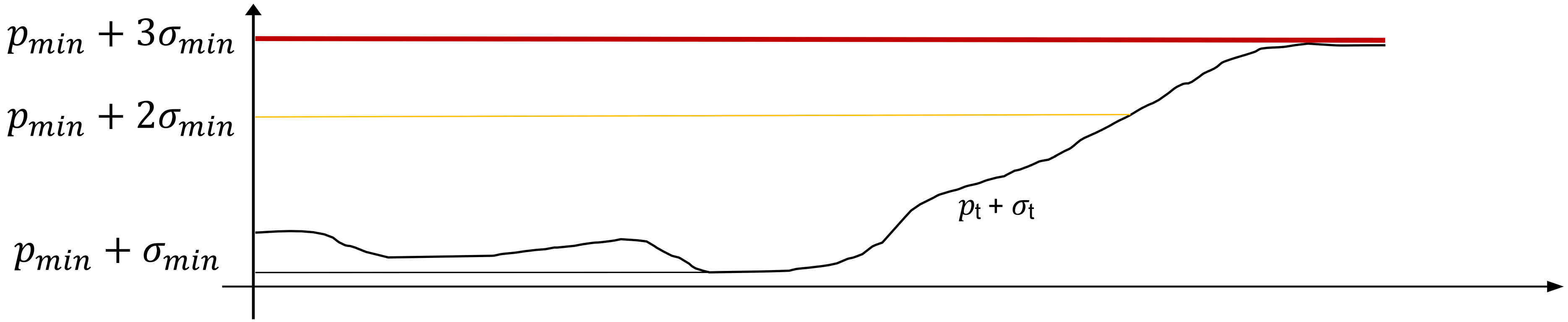


Gama, et. al, **Learning with Drift Detection**, SBIA, Springer, 2004

Concept Drift Detectors

Monitoring the classification error - Drift Detection Method (DDM)

1. Compute the classification error mean p_t and $\sigma_t = \sqrt{\frac{p_t(1-p_t)}{t}}$
2. Let $p_{\min}$ and $\sigma_{\min}$ the minimum p_t and σ_t values seen until now
3. Raise a **warning** when $p_t + \sigma_t > p_{\min} + 2 * \sigma_{\min}$
4. Raise a **change** when $p_t + \sigma_t > p_{\min} + 3 * \sigma_{\min}$



Gama, et. al, **Learning with Drift Detection**, SBIA, Springer, 2004

Concept Drift Detectors

Monitoring the classification error - Early Drift Detection Method (EDDM)

- It considers the distance between two errors classification instead of considering only the number of errors.
- While the learning method is learning, it will improve the predictions and the distance between two errors will increase.
- When a drift occurs, the distance between two errors will decrease.
- Compute the average distance between 2 errors and its std, and look for outliers in the tails.

Baena-Garcıa, M., et al. **Early drift detection method**. In Fourth international workshop on knowledge discovery from data streams 2006.

Concept Drift Detectors

Monitoring the classification error - Adaptive sliding WINdow (ADWIN)

- An adaptive sliding window whose size is recomputed online according to the rate of change observed.
- It does not need parameters

A. Bifet, R Gavalda **Learning from Time-Changing Data with Adaptive Windowing**. SDM, 2007

POLITECNICO
MILANO 1863

Concept Drift Detectors

Monitoring the classification error - Adaptive sliding WINdow (ADWIN)

W = 111111101010110

W_0 = 11111 W_1 = 1101010110

W_0 = 111111 W_1 = 101010110

....

W_0 = 1111111 W_1 = 01010110 $|\hat{\mu}_{W_0} - \hat{\mu}_{W_1}| \geq \epsilon_c$: CHANGE DETECTED!

Min. length W = 10
Min. length W_0, W_1 = 5

A. Bifet, R Gavalda **Learning from Time-Changing Data with Adaptive Windowing**. SDM, 2007

POLITECNICO
MILANO 1863

Exercise 2: Concept Drift Detectors

Credits

- Albert Bifet DATA STREAM MINING 2020-2021 course at Telecom Paris
- Alessio Bernardo & Emanuele Della Valle

Streaming Machine Learning
Taming Concept Drift

Alessio Bernardo

Post-doc @ Politecnico di Milano

CTO & Co-founder @ Motus ml

POLITECNICO
MILANO 1863