

# Streaming Machine Learning (SML)

---

Alessio Bernardo & Emanuele Della Valle

14-03-2022

# Part II

---

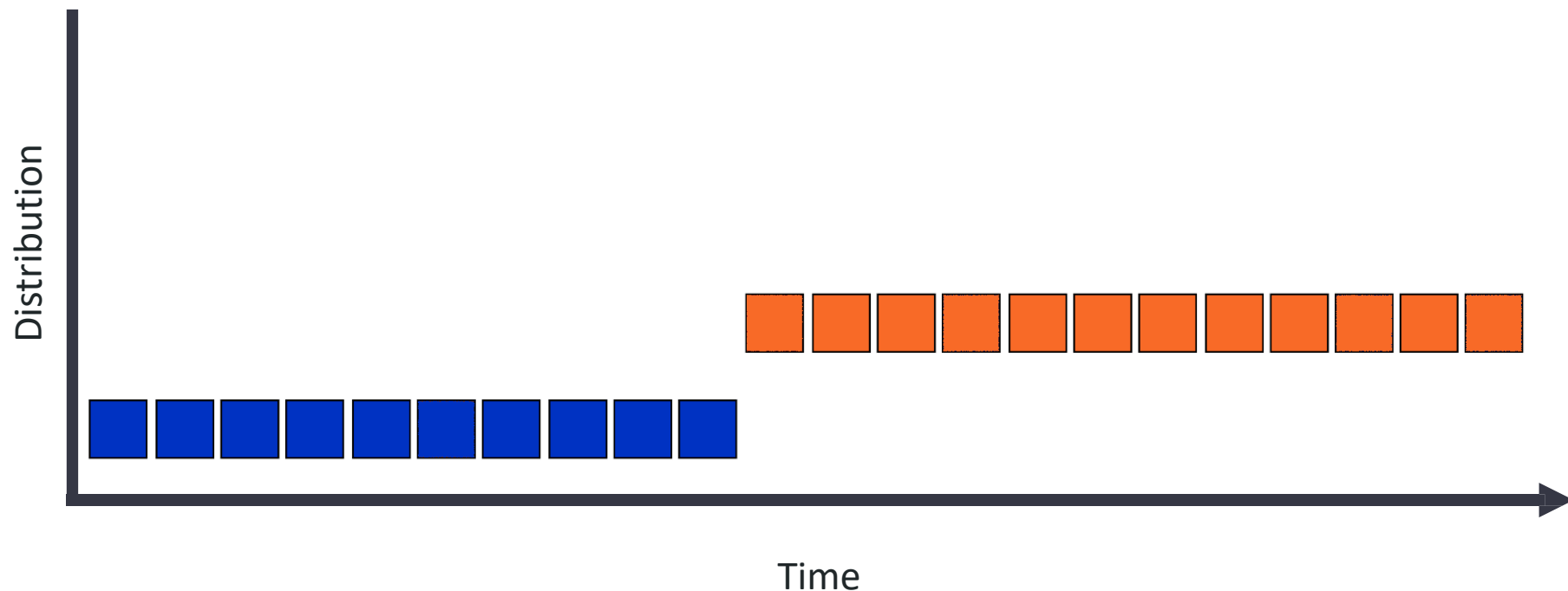
## Concept Drift

# Credits

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

# What is Concept Drift?

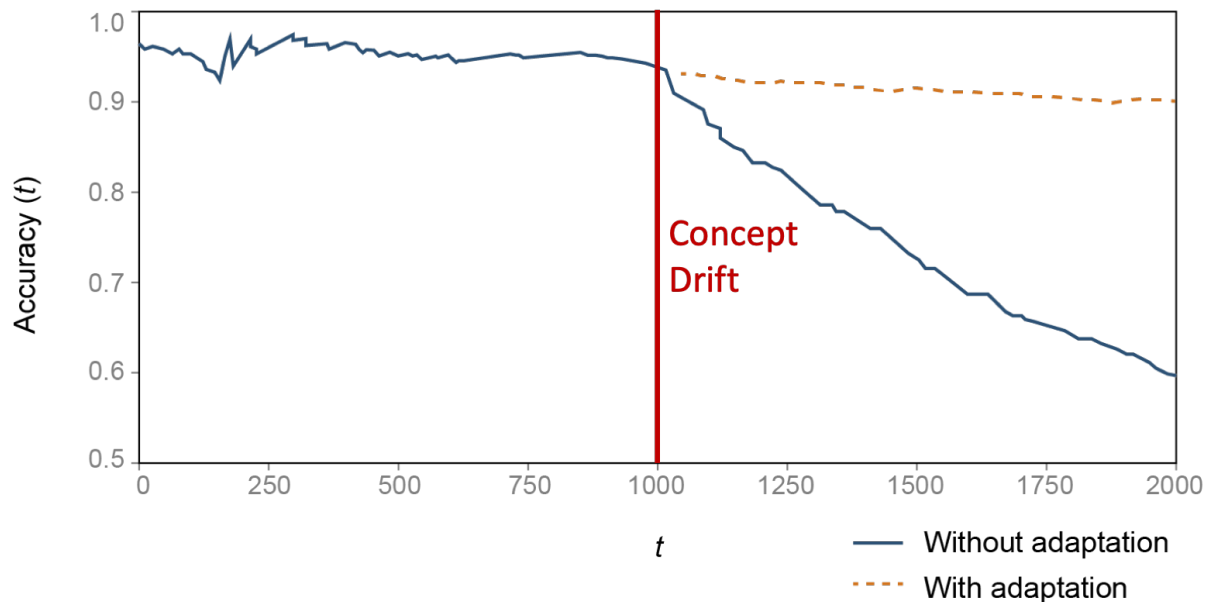
# What is Concept Drift?



Source: 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 Concept Drift?

Class 1 Class 2 Concept Drift



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

# What is Concept Drift?

## Problem

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

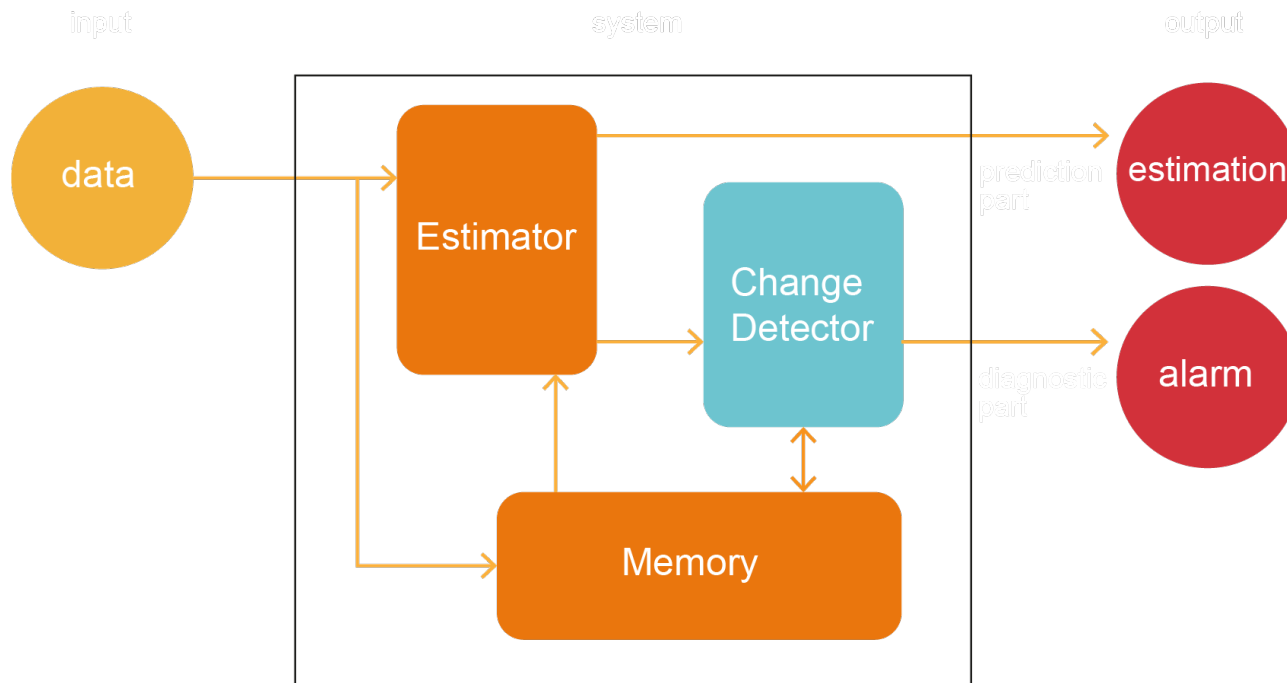
$$|\hat{X}_{t+1} - X_{t+1}|$$

## Outputs

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

Source: 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 Concept Drift?



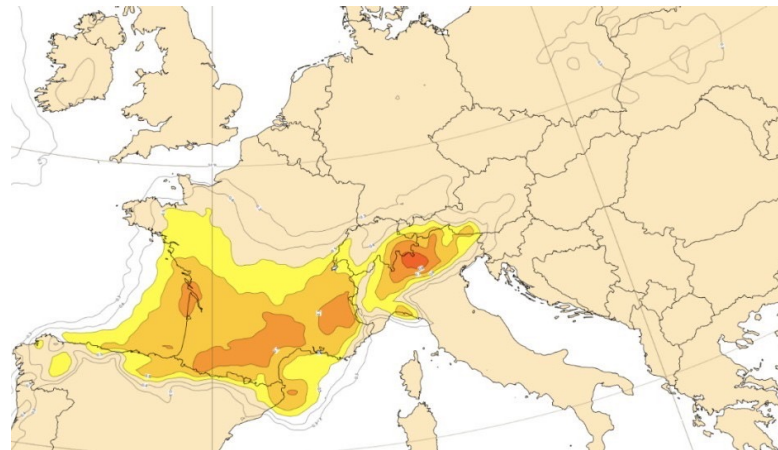
Bifet, A. and Gavaldá, R. **Adaptive Learning from evolving data streams**. In International Symposium on Intelligent Data Analysis (pp.249-260).Springer 2009, August.



# What is 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



<https://www.meteosvizzera.admin.ch/home/attualita/meteosvizzera-blog.subpage.html/it/data/blogs/2021/8/efi---extreme-forecast-index.html>

# 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:

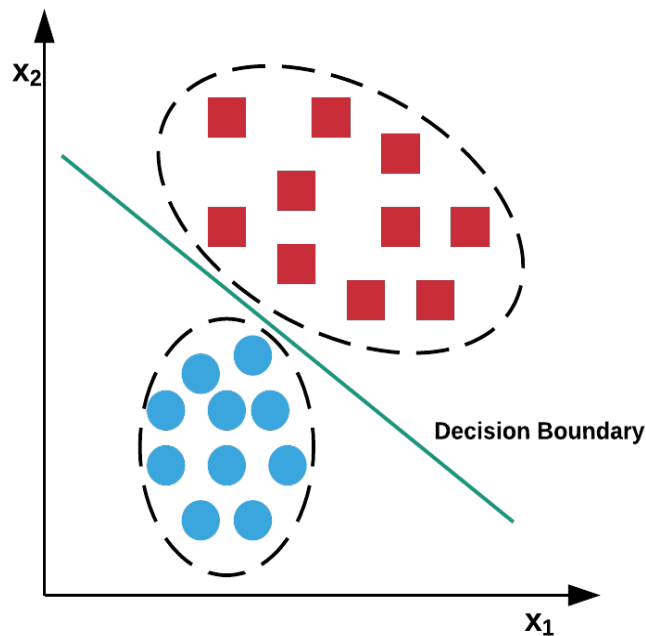
$$p(y|X_t) = \frac{p(y) * p(X_t|y)}{p(X_t)}$$

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.

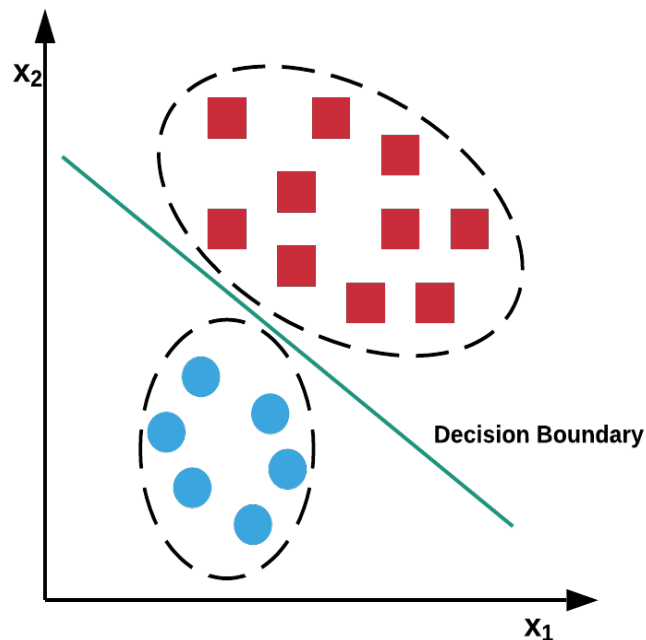
Tsymbol, A. (2004). **The problem of concept drift: definitions and related work**. Computer Science Department, Trinity College Dublin, 106(2), 58.

# Concept Drift Characteristics

Original distribution



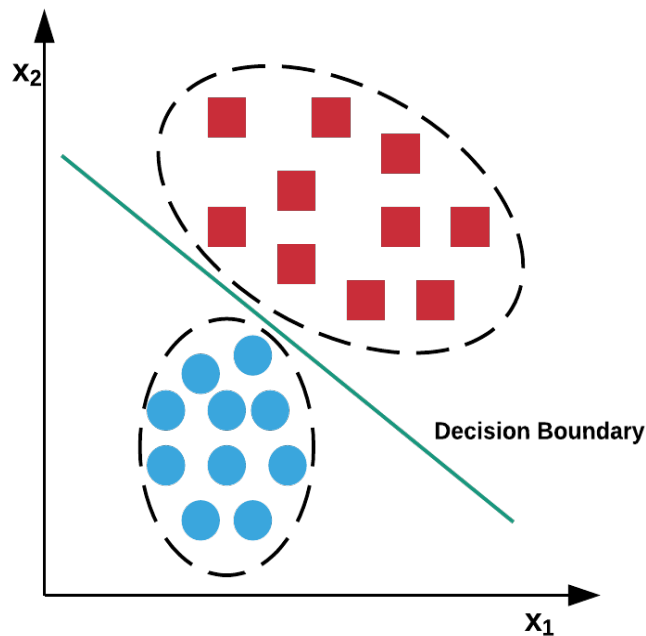
$p(y)$  changes



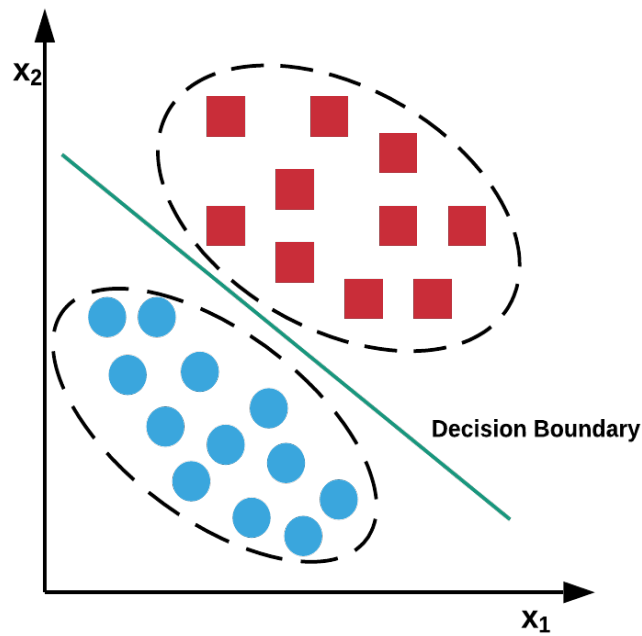
Tsymbol, A. (2004). **The problem of concept drift: definitions and related work.** Computer Science Department, Trinity College Dublin, 106(2), 58.

# Concept Drift Characteristics

Original distribution



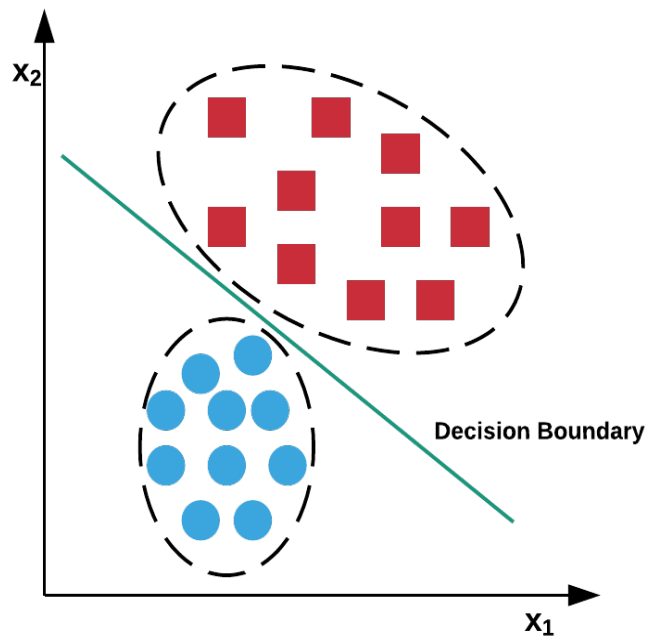
$p(X_t|y)$  changes



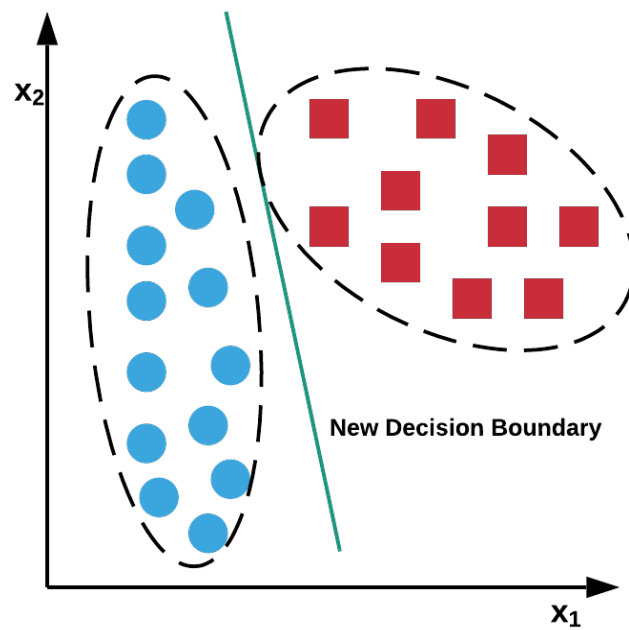
Tsymbol, A. (2004). **The problem of concept drift: definitions and related work.** Computer Science Department, Trinity College Dublin, 106(2), 58.

# Concept Drift Characteristics

Original distribution



$p(y | X_t)$  changes



Tsymbol, A. (2004). **The problem of concept drift: definitions and related work.** Computer Science Department, Trinity College Dublin, 106(2), 58.

# Concept Drift Characteristics

## Example

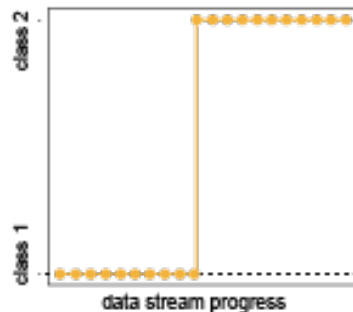
- Consider the case of predicting extreme weather phenomena occurrences based on atmospheric pressure and temperature. Usually, extreme weather phenomena occur in the case of low atmospheric pressure and high temperature.

# Concept Drift Characteristics

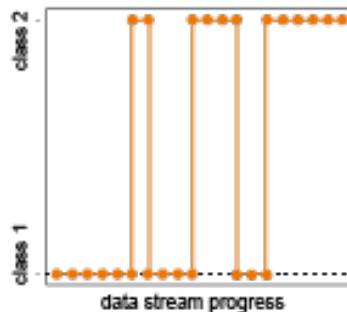
- **$p(y)$  concept drift:** in the XX century, the distribution of atmospheric pressure and temperature did not change, but the extreme weather phenomena were more frequent.
- **$p(X|y)$  concept drift:** in the first two decades of XXI century, the atmospheric pressure and air temperature conditions, in which these phenomena occur, also started to change, but not so drastically to move the decision boundary we use for predicting them.
- **$p(y|X)$  concept drift:** due to the on going climate change, these phenomena start occurring more frequently with higher atmospheric pressure and lower temperature. As a consequence, we have to update the decision boundary to keep an high predictive performance.



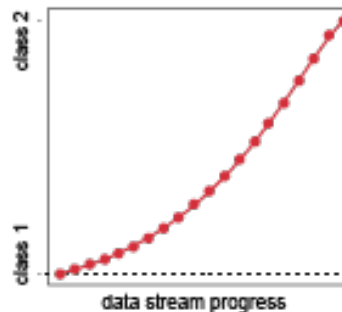
# Concept Drift Characteristics



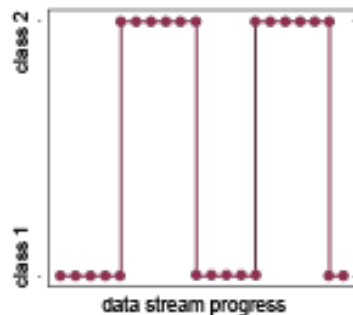
(a) Sudden.



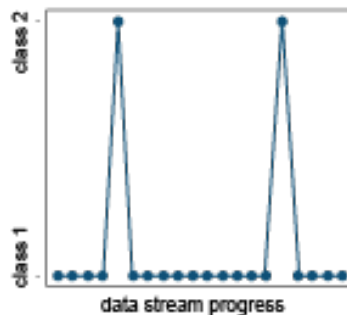
(b) Gradual.



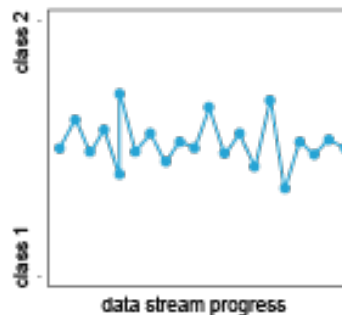
(c) Incremental.



(d) Recurring.



(e) Blips.



(f) Noise.

Tsymbol, A. (2004). **The problem of concept drift: definitions and related work.** Computer Science Department, Trinity College Dublin, 106(2), 58.

# Concept Drift Characteristics

## 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?"

Tsymbol, A. (2004). **The problem of concept drift: definitions and related work.** Computer Science Department, Trinity College Dublin, 106(2), 58.

# Concept Drift Detectors

# Concept Drift Detectors

## 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

Gama, João, et al. **A survey on concept drift adaptation**. ACM computing surveys (CSUR) 46.4 (2014): 1-37.

# 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. (2003). **The cusum test for parameter change in time series models**. Scandinavian Journal of Statistics, 30(4), 781-796.

# Concept Drift Detectors

## 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  $h$ .

$$g_0 = 0$$

$$g_t = g_{t-1} + (x_t - \hat{x}) - v$$

$$G_t = \min(g_t, G_{t-1}) *$$

$$\text{if } g_t - G_t > h \text{ then Alarm} *$$

\* When the signal is decreasing, we should use:

$$G_t = \max(g_t, G_{t-1})$$

$$\text{if } G_t - g_t > h \text{ then Alarm}$$

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:

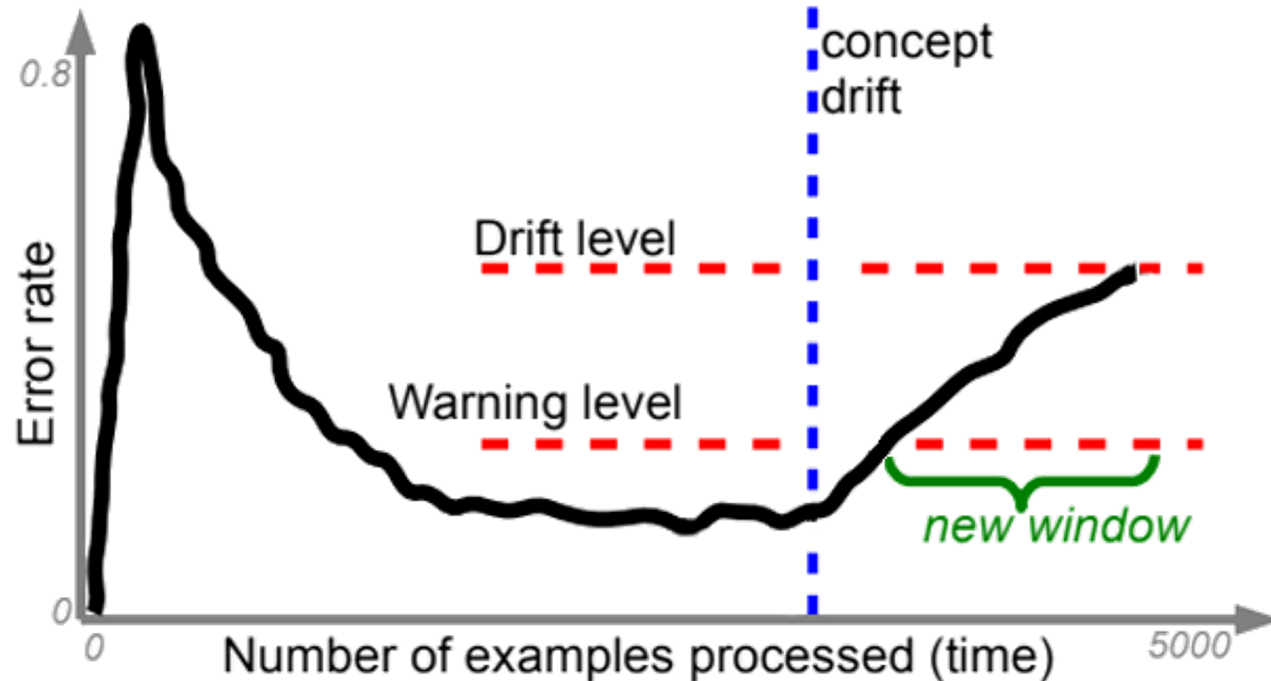
- the most straightforward figure of merit to monitor
- changes in  $p_t$  prompt adaptation only when performances are affected

### Cons:

- Concept drift detection from supervised samples only

Gama, João, et al. **A survey on concept drift adaptation**. ACM computing surveys (CSUR) 46.4 (2014): 1-37.

# Concept Drift Detectors

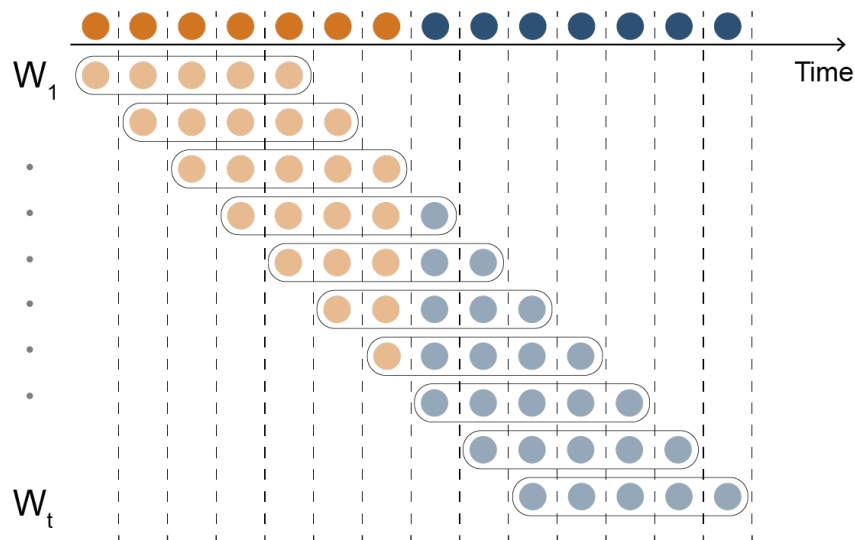


Gama, et. al, **Learning with Drift Detection**, SBIA 2004, Springer.

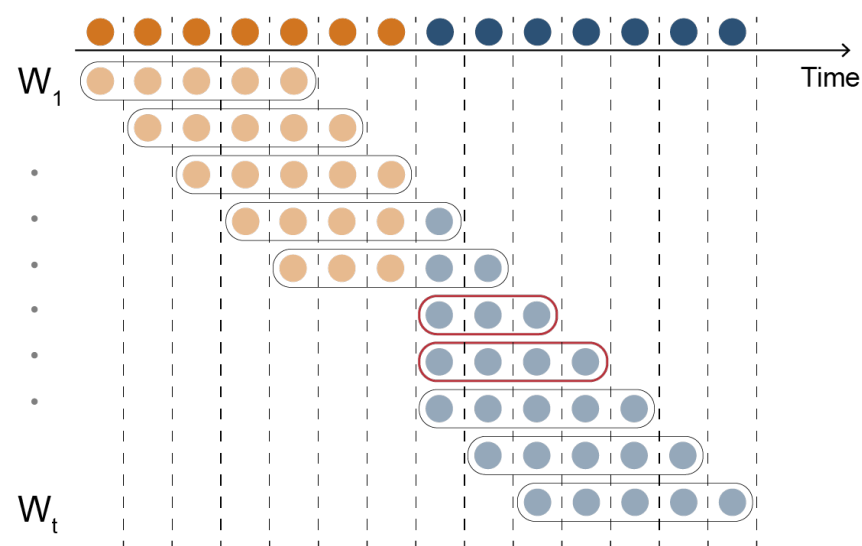


# Concept Drift Detectors

## Passive



## Active

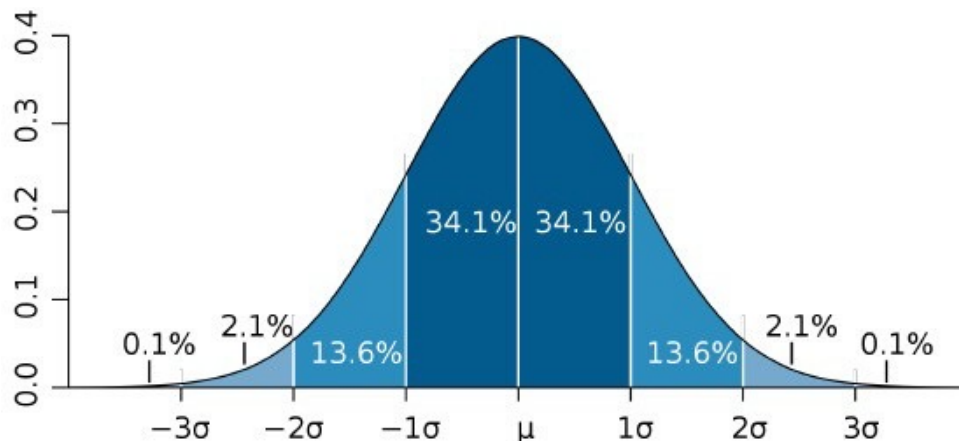


Gama, João, et al. **A survey on concept drift adaptation.** ACM computing surveys (CSUR) 46.4 (2014): 1-37.

# 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.

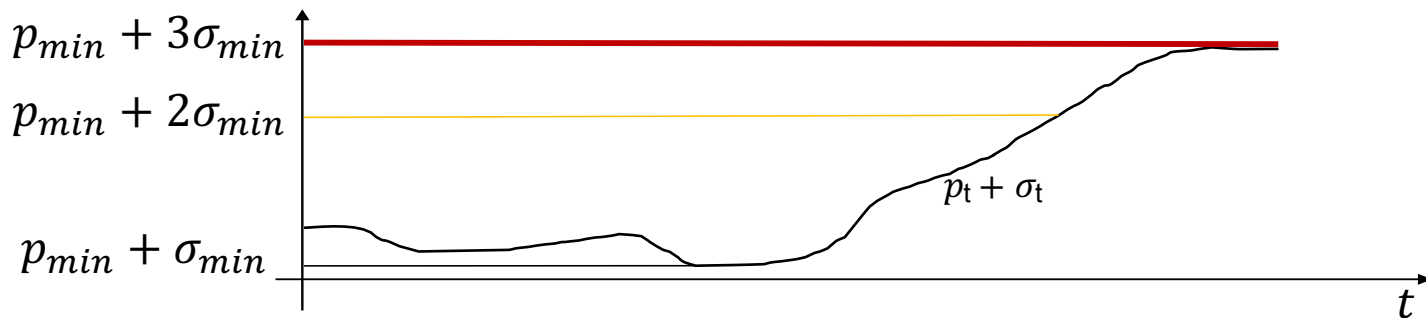


Gama, et. al, **Learning with Drift Detection**, SBIA 2004, Springer.

# 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  $\sigma_t$  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 2004, Springer.

# 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-Garcia, 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 data stream 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

# Concept Drift Detectors

Monitoring the **classification error** – Adaptive data stream sliding **WINDOW (ADWIN)**

$W =$  101010110111111

$W_0 =$  1  $W_1 =$  01010110111111

$W_0 =$  10  $W_1 =$  1010110111111

$W_0 =$  101  $W_1 =$  010110111111

.....

$W_0 =$  101010110  $W_1 =$  111111  $|\hat{\mu}_{W_0} - \hat{\mu}_{W_1}| \geq \epsilon_c$  : **CHANGE DETECTED!**

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

# QUIZ

1. What is concept drifts?
  - a. Sometimes few samples are different from the others
  - b. A change in the model performances
  - c. A change of the new data underlying statistics
2. Why is so important detect concept drift occurrences?
  - a. Because the model risks being trained on data no longer representative and must be adapted
  - b. To be able to raise an alarm, so the user can decide to take some actions or not
  - c. It is not important. A SML model is always able to adapt to concept drifts without any detectors

# QUIZ

1. What is concept drifts?
  - a. Sometimes few samples are different from the others
  - b. A change in the model performances
  - c. **A change of the new data underlying statistics**
2. Why is so important detect concept drift occurrences?
  - a. Because the model risks being trained on data no longer representative and must be adapted
  - b. To be able to raise an alarm, so the user can decide to take some actions or not
  - c. It is not important. A SML model is always able to adapt to concept drifts without any detectors



# QUIZ

1. What is concept drifts?
  - a. Sometimes few samples are different from the others
  - b. A change in the model performances
  - c. **A change of the new data underlying statistics**
2. Why is so important detect concept drift occurrences?
  - a. **Because the model risks being trained on data no longer representative and must be adapted**
  - b. To be able to raise an alarm, so the user can decide to take some actions or not
  - c. It is not important. A SML model is always able to adapt to concept drifts without any detectors

# EXERCISE 2: Concept Drift

## LAB 2: Concept Drift Detectors

---