Streaming Machine Learning (SML)

Alessio Bernardo & Emanuele Della Valle
05-07-2021

Part II

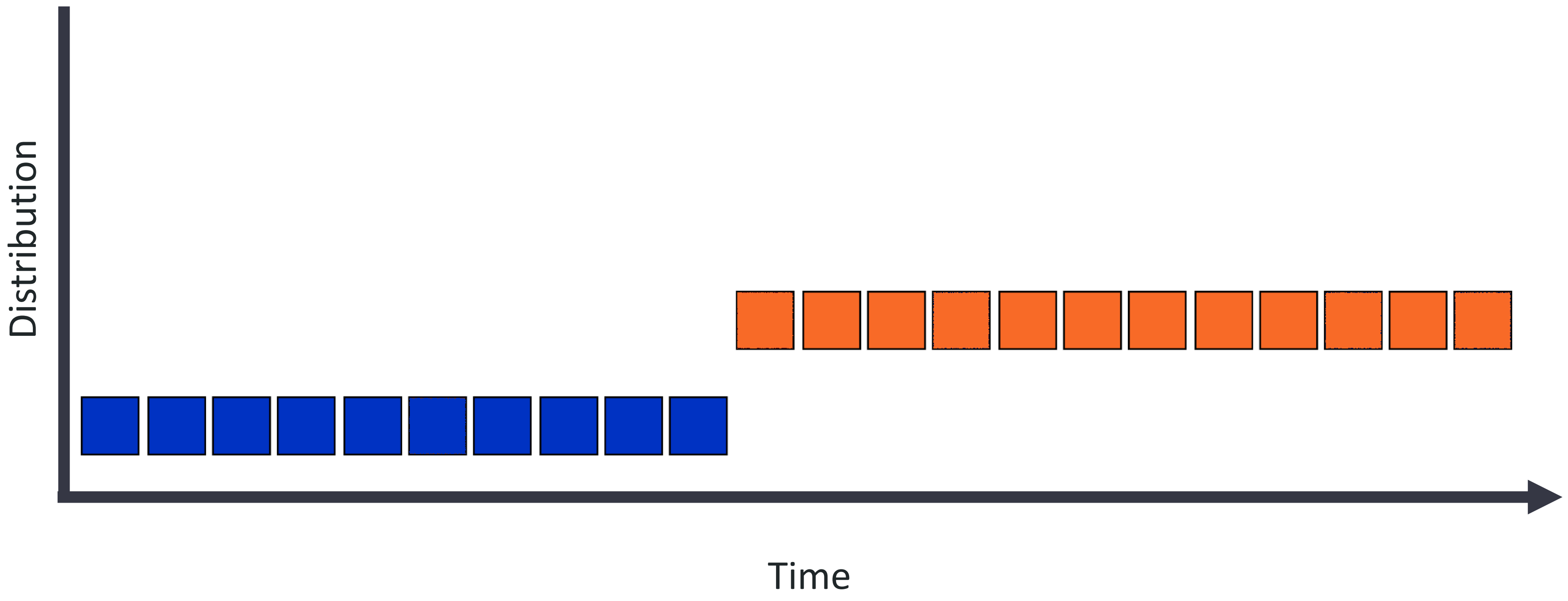
Concept Drift

Credits

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

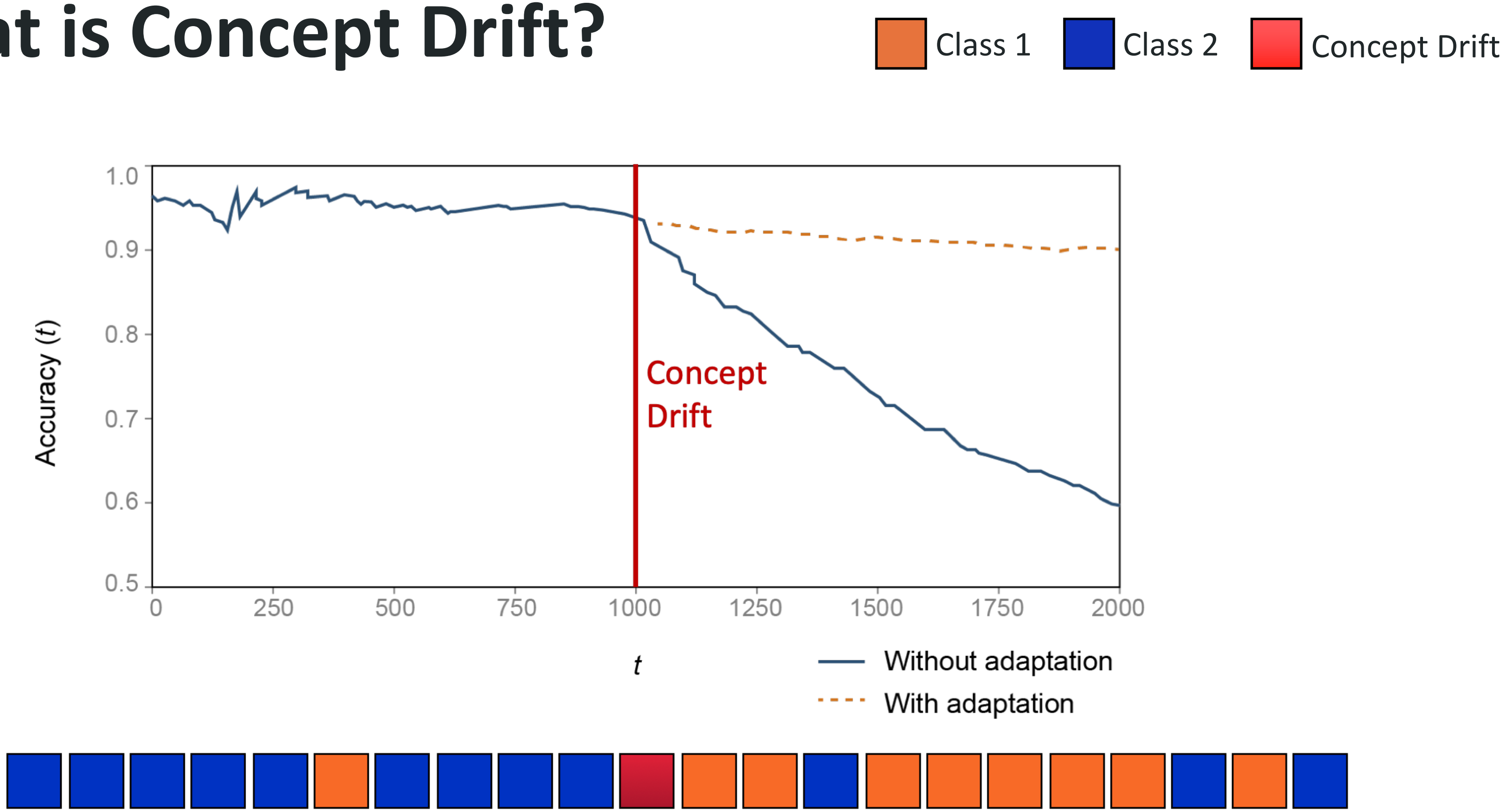
What is Concept Drift?

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?



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?

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:

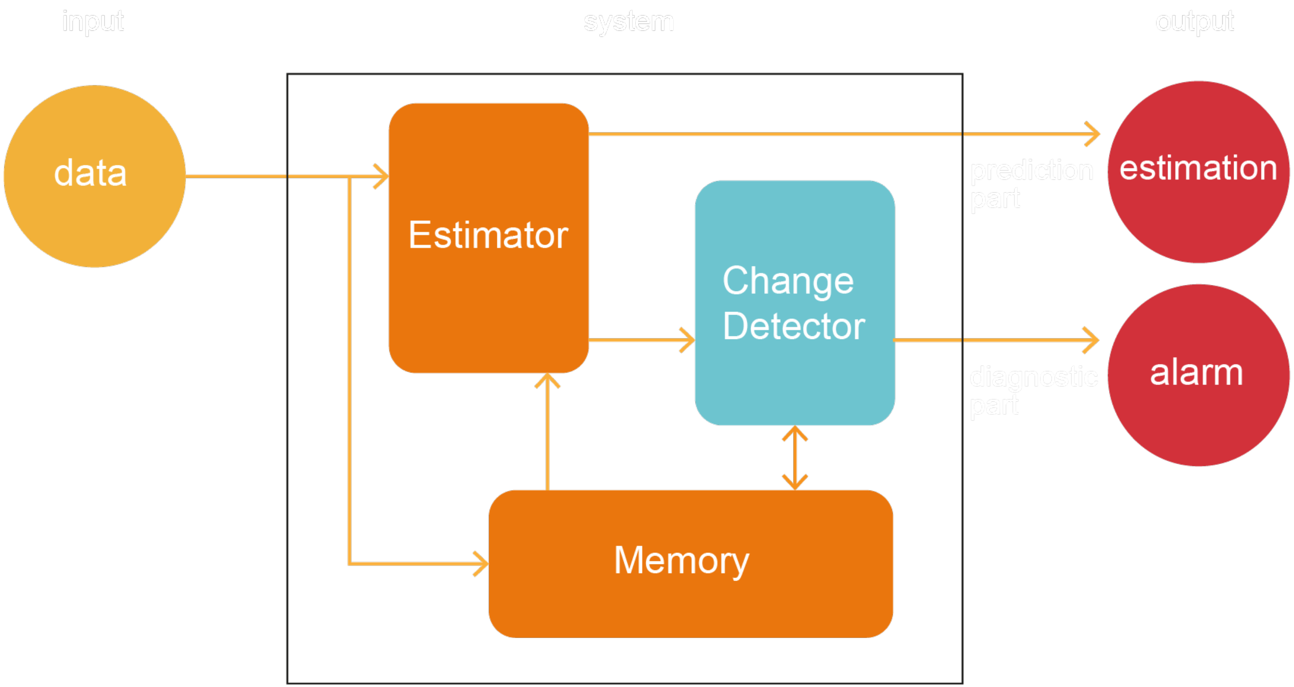
$$\left|\hat{X}_{t+1} - X_{t+1}\right|$$

Outputs

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

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.

Concept Drift Characteristics

Concept Drift Characteristics

Given an input sequence $X_1, X_2, \ldots, X_t$ 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:

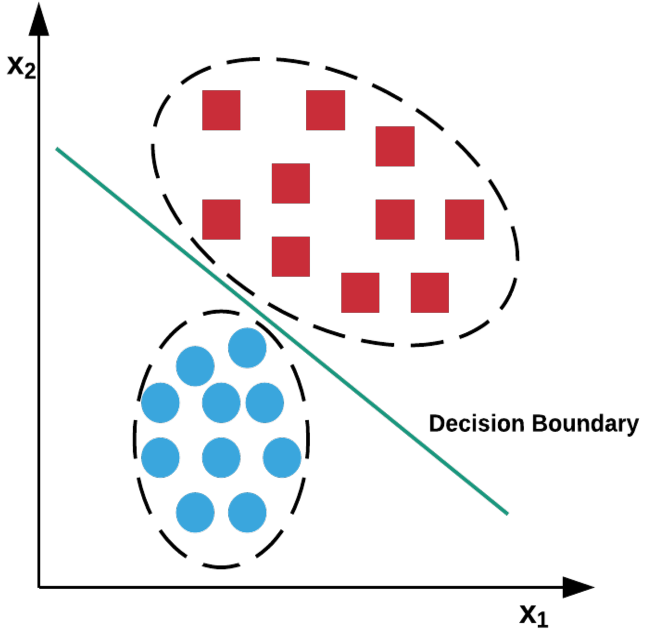
$$p(y|X_t) = \frac{p(y) * p(X_t|y)}{p(X_t)}$$

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.

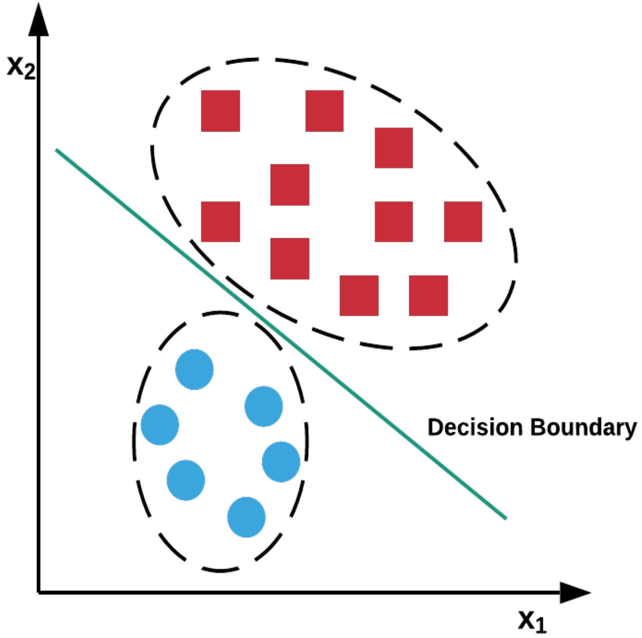
Tsymbal, A. (2004). **The problem of concept drift: definitions and related work.** Computer Science Department, Trinity College Dublin, 106(2), 58.

Concept Drift Characteristics

Original distribution



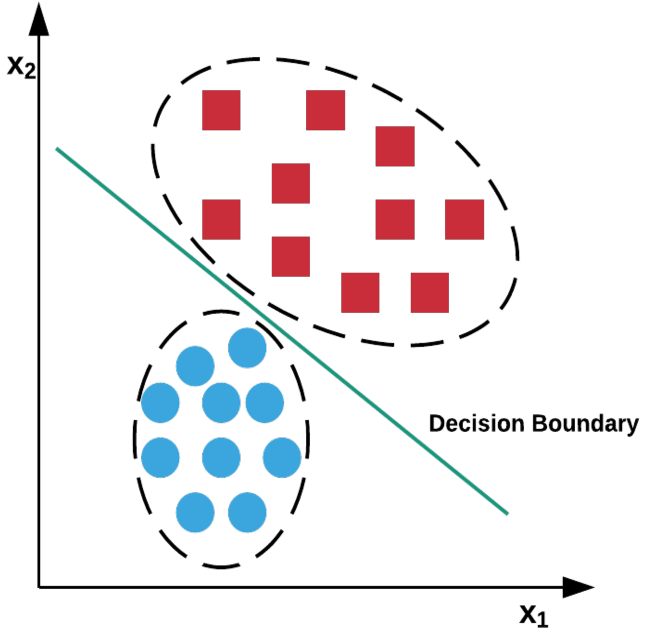
p(y) changes



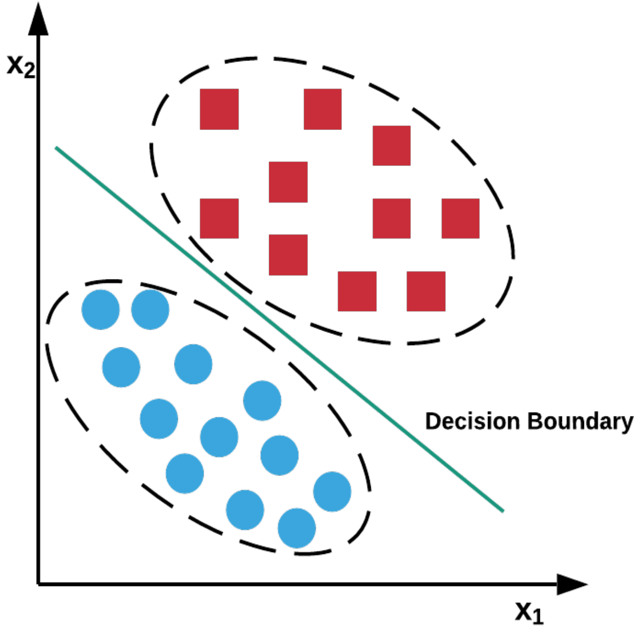
Tsymbal, A. (2004). **The problem of concept drift: definitions and related work.** Computer Science Department, Trinity College Dublin, 106(2), 58.

Concept Drift Characteristics

Original distribution



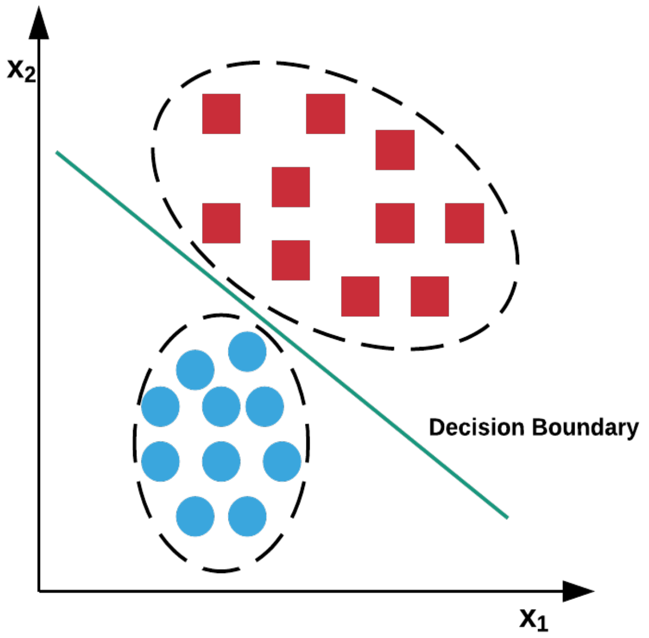
$p(X_t|y)$ changes



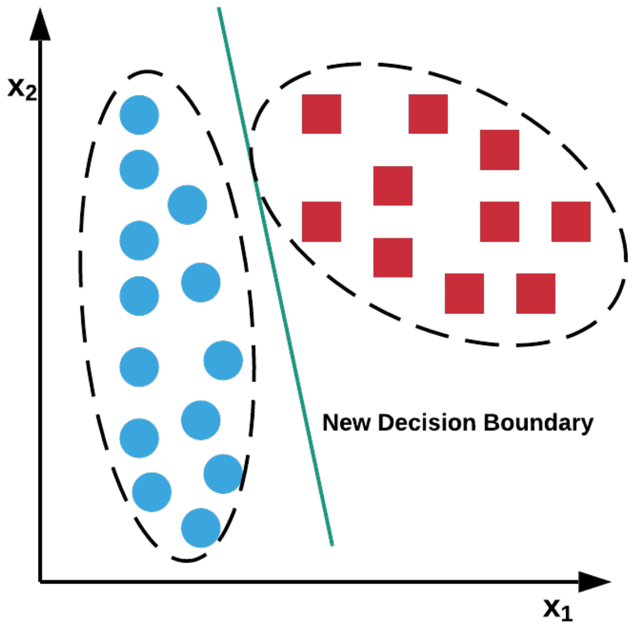
Tsymbal, A. (2004). **The problem of concept drift: definitions and related work.** Computer Science Department, Trinity College Dublin, 106(2), 58.

Concept Drift Characteristics

Original distribution

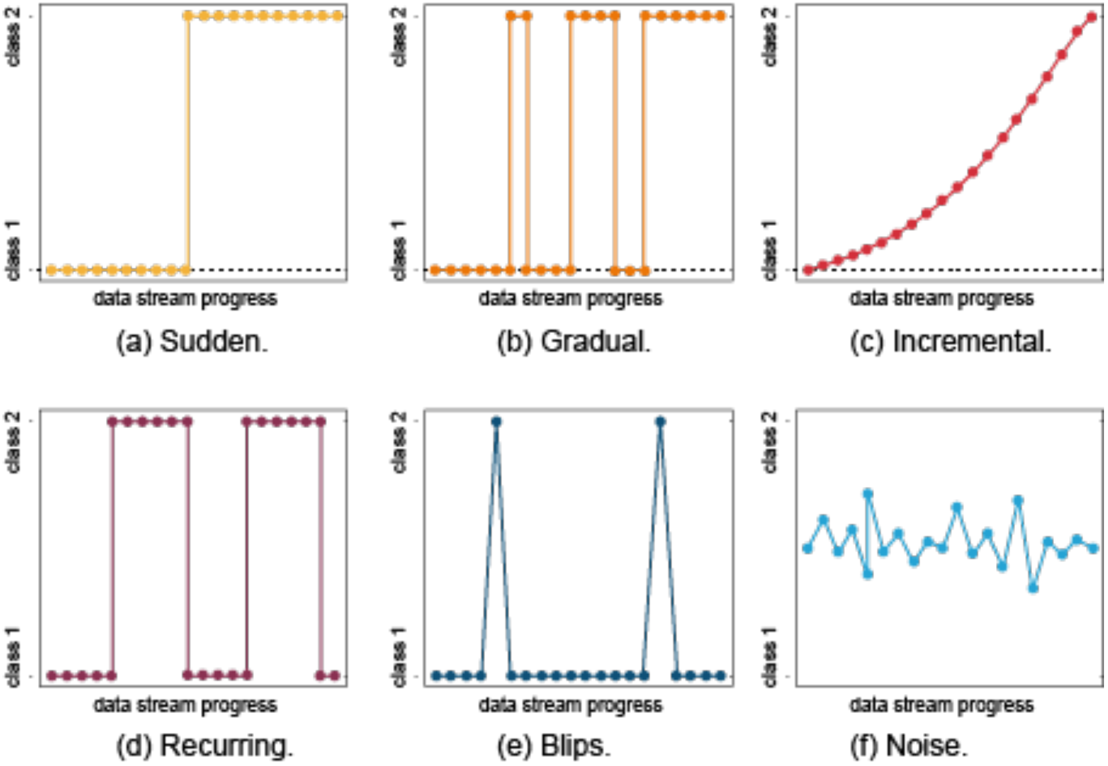


p(y|X$_t$) changes



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

Concept Drift Characteristics



Tsymbal, 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?”

Tsymbal, 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.

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

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.

$$g = [\,]$$
$$g[t] = g[t-1] + (x_t - \hat{x}) - v$$
$$G_t = \min(g)$$

$$If\ g[t] - G_t > h \text{ then Alarm}$$

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:

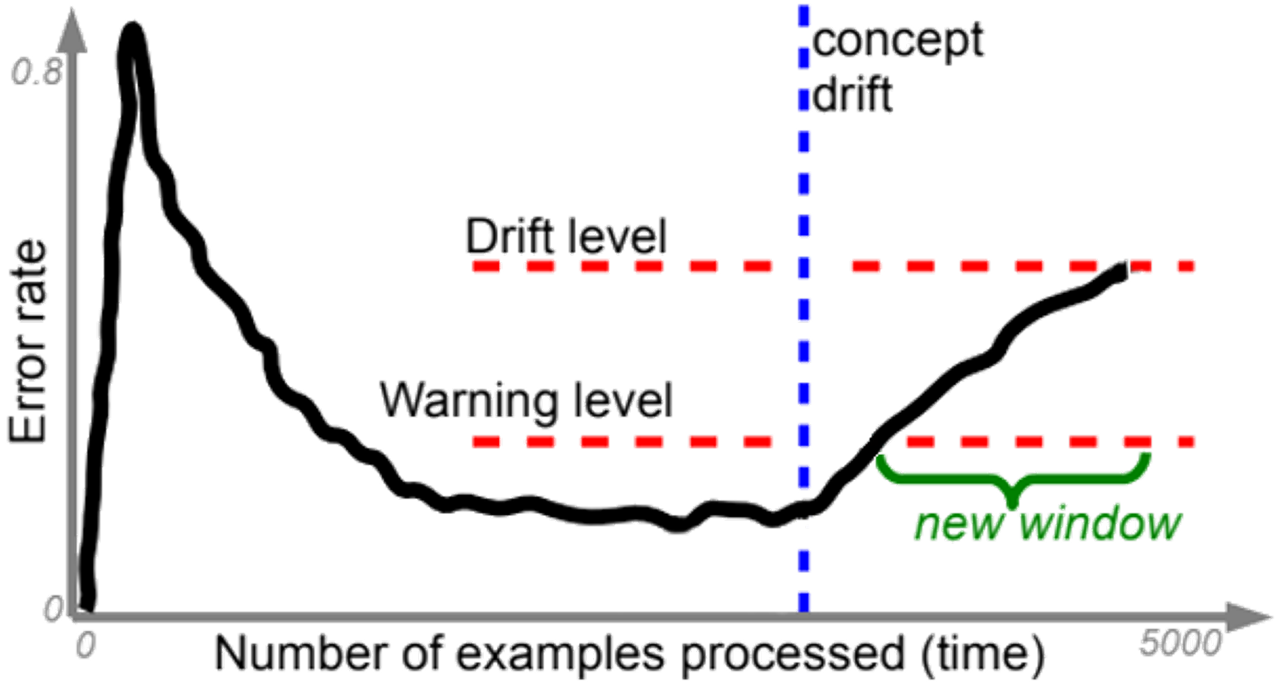
- the most straightforward figure of merit to monitor
- changes in p_t prompt adaptation only when performances are affected

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.

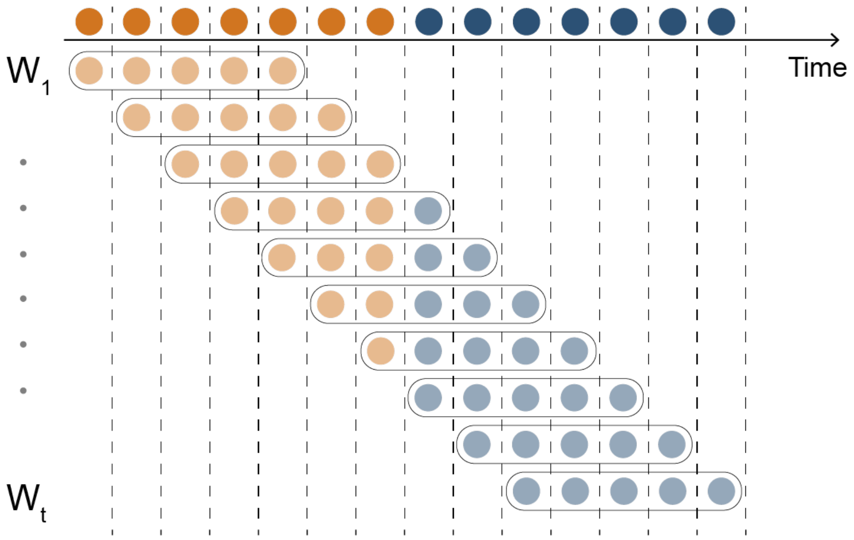
Concept Drift Detectors



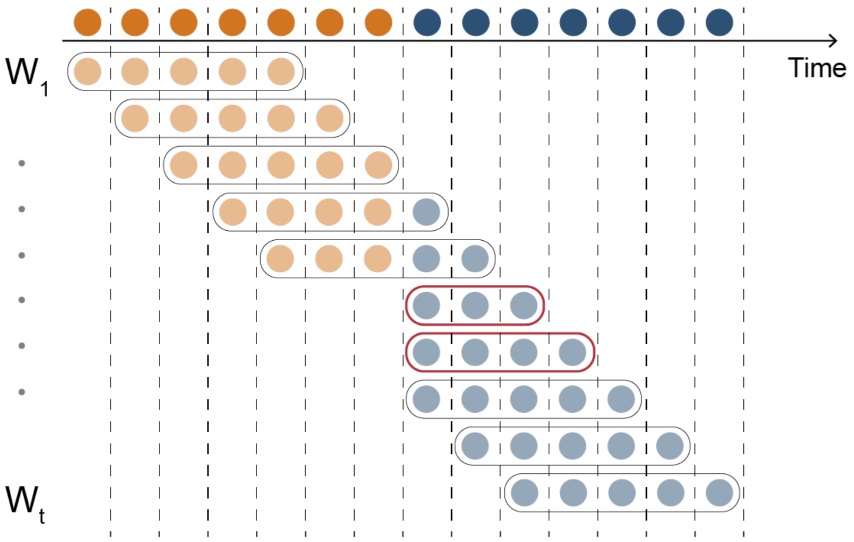
Gama, et. al, **Learning with Drift Detection**, SBIA 2004, Springer.

Concept Drift Detectors

Passive



Active

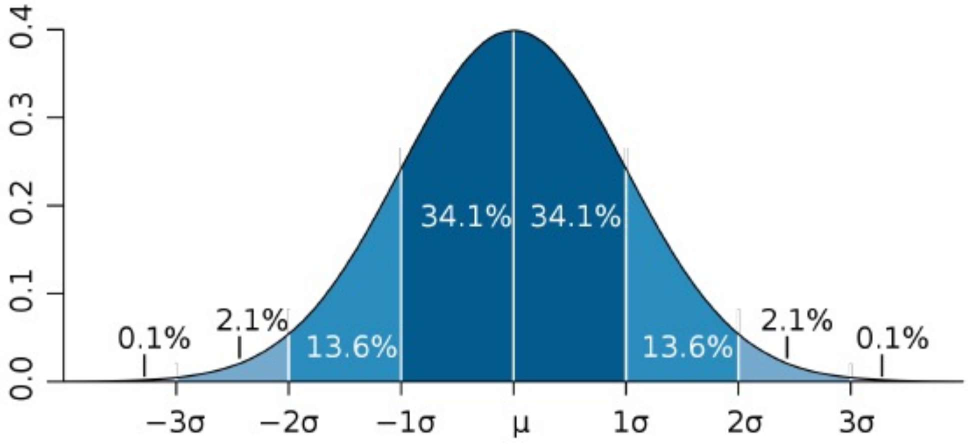


Gama, João, et al. **A survey on concept drift adaptation**. ACM computing surveys (CSUR) 46.4 (2014): 1-37.

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.

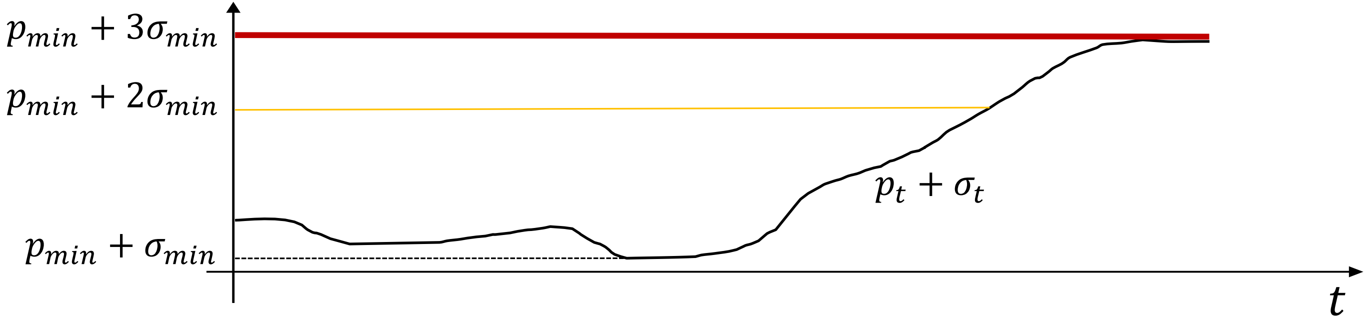


Gama, et. al, **Learning with Drift Detection**, SBIA 2004, Springer.

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 σ_{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 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-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 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)

$$W = \boxed{101010110111111}$$

$$W_0 = \boxed{1} \quad W_1 = \boxed{01010110111111}$$

$$W_0 = \boxed{10} \quad W_1 = \boxed{1010110111111}$$

$$W_0 = \boxed{101} \quad W_1 = \boxed{010110111111}$$

....

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

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