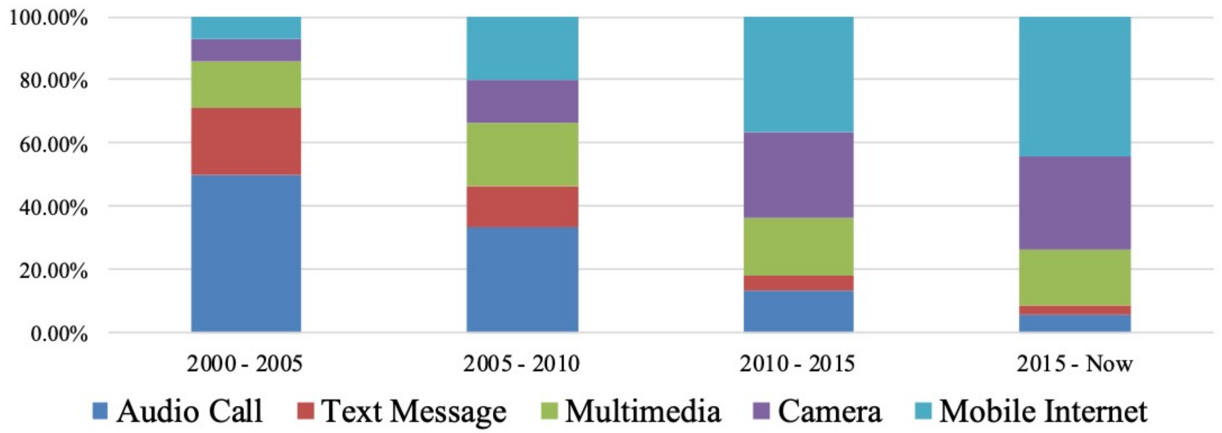
Concept drift

Streams Processing

What is concept drift?

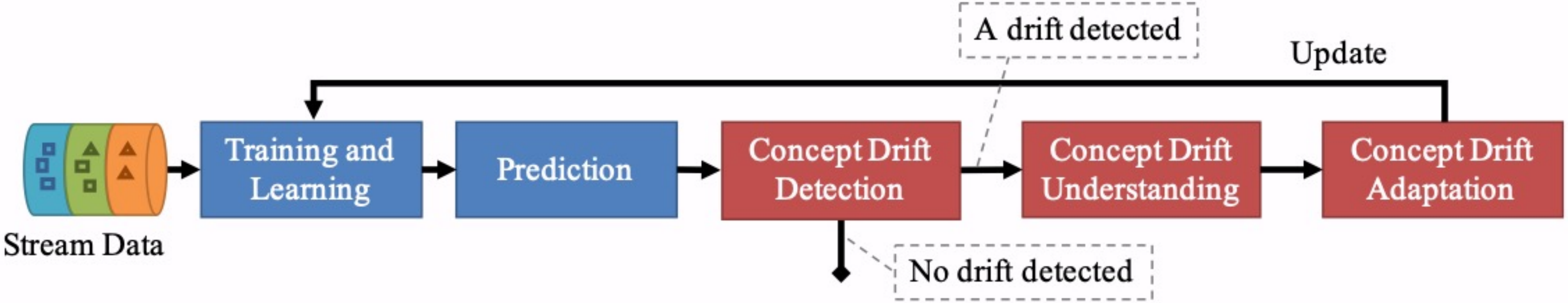
- Example: Mobile phone usage
- Challenges:
    - how to accurately detect concept drift in unstructured and noisy datasets
    - how to quantitatively understand concept drift in an explainable way
    - how to effectively react to drift by adapting related knowledge
- highly prominent and significant issue in the context of the big data era because the uncertainty of data types and data distribution is an inherent nature of big data.



Differences regarding conventional ML

- Conventional machine learning has two main components:
  - training/learning
  - Prediction

- Learning under concept drift presents three new components:
  - drift detection (whether or not drift occurs)
  - drift understanding (when, how, where it occurs)
  - drift adaptation (reaction to the existence of drift)

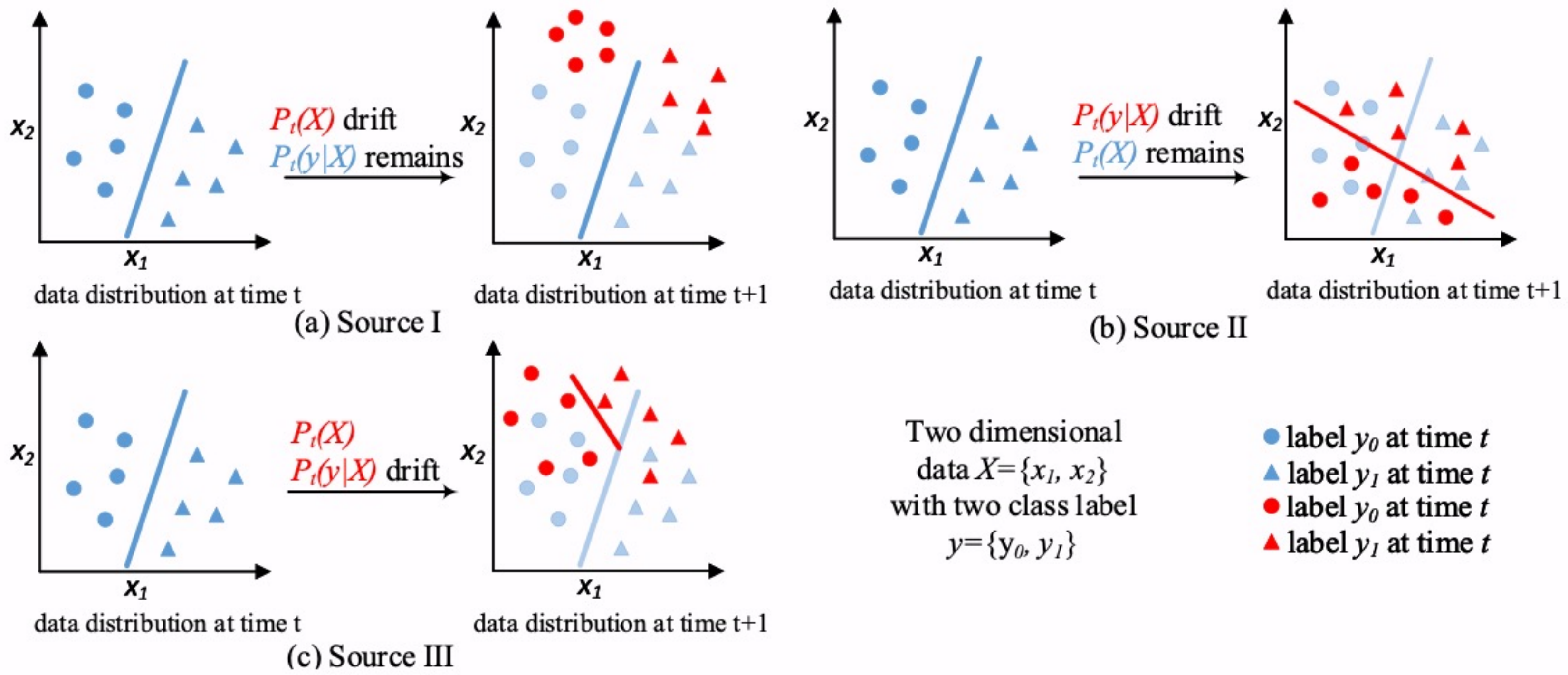


How to learn under concept drift?

Rigorous definition of CD

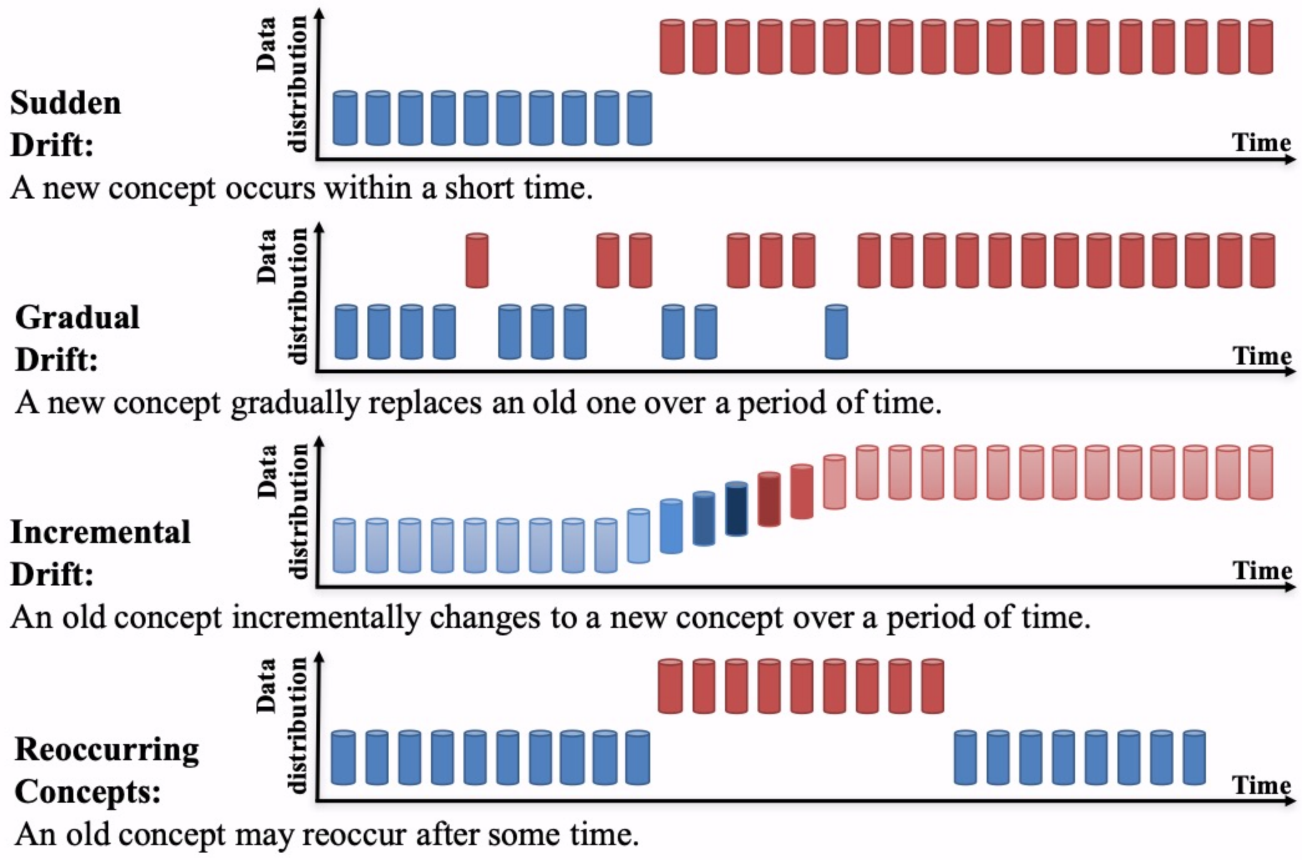
- Given a time period $[0, t]$
- A set of samples $S_{0,t} = \{d_0, \ldots, d_t\}$
- Where a data point is $d_i = (X_i, y_i)$
- Distribution of the set $F_{0,t}(X, y)$
- Concept drift at t+1 is defined by $F_{0,t}(X, y) \neq F_{t+1,\infty}(X, y)$

Sources of CD



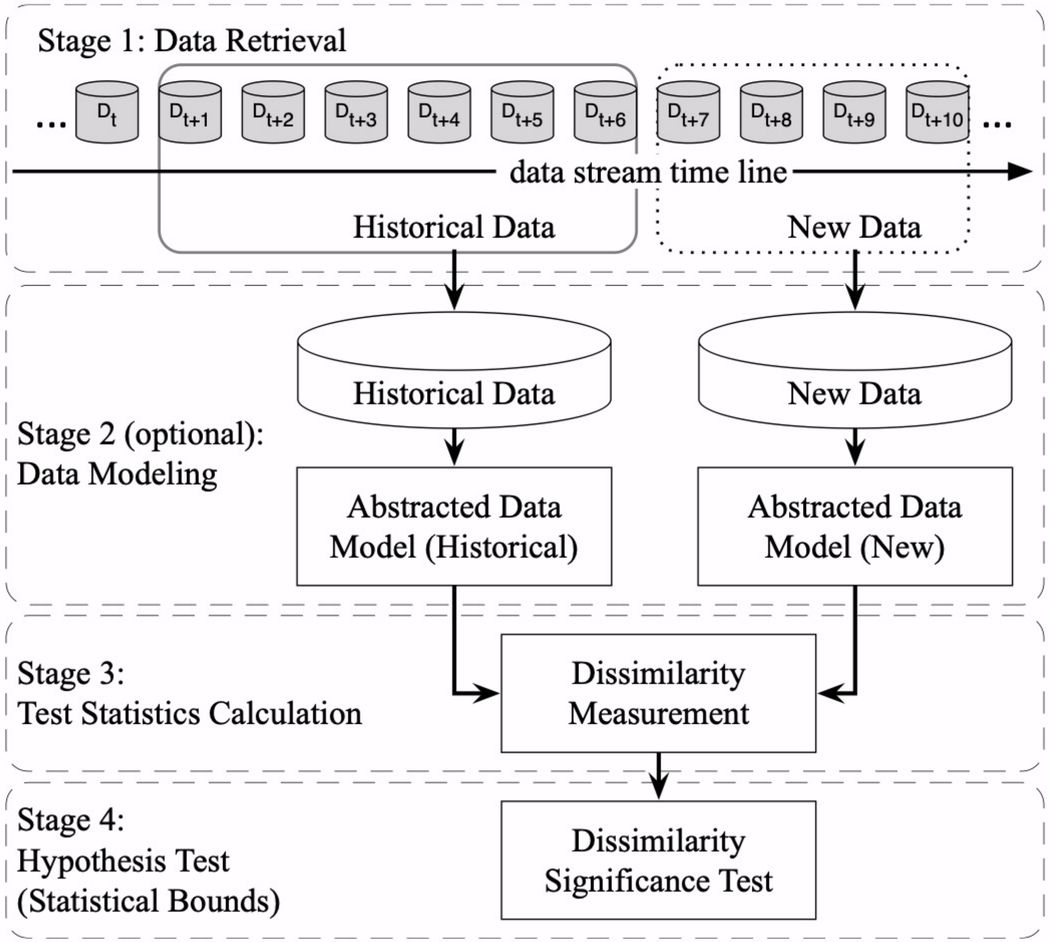
- Source I: $P_t(X) \neq P_{t+1}(X) \quad P_t(y|X) = P_{t+1}(y|X)$
- Source II: $P_t(X) = P_{t+1}(X) \quad P_t(y|X) \neq P_{t+1}(y|X)$
- Source III: $P_t(X) \neq P_{t+1}(X) \quad P_t(y|X) \neq P_{t+1}(y|X)$

Types of Concept Drift



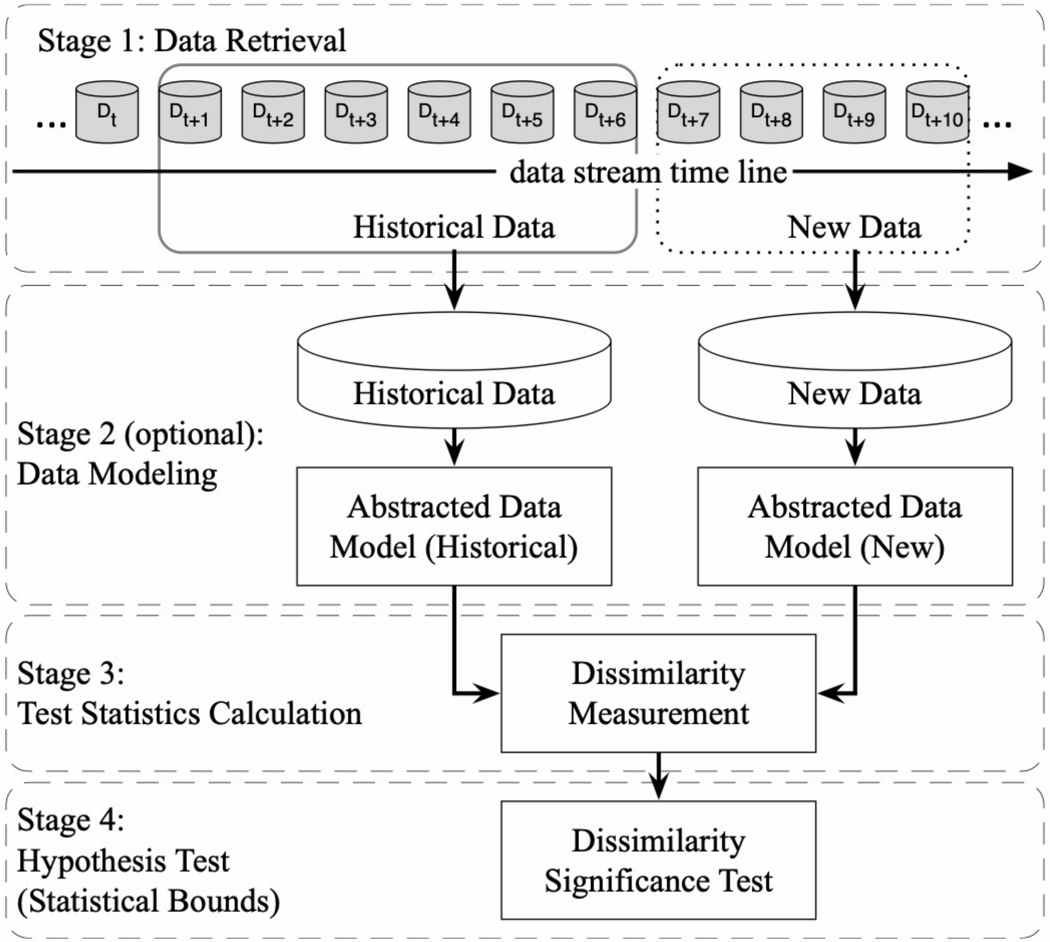
CD detection

- Stage 2: aims to abstract the retrieved data and extract the key features containing sensitive information, that is, the features of the data that most impact a system if they drift. It mainly concerns dimensionality reduction, or sample size reduction, to meet storage and online speed requirements



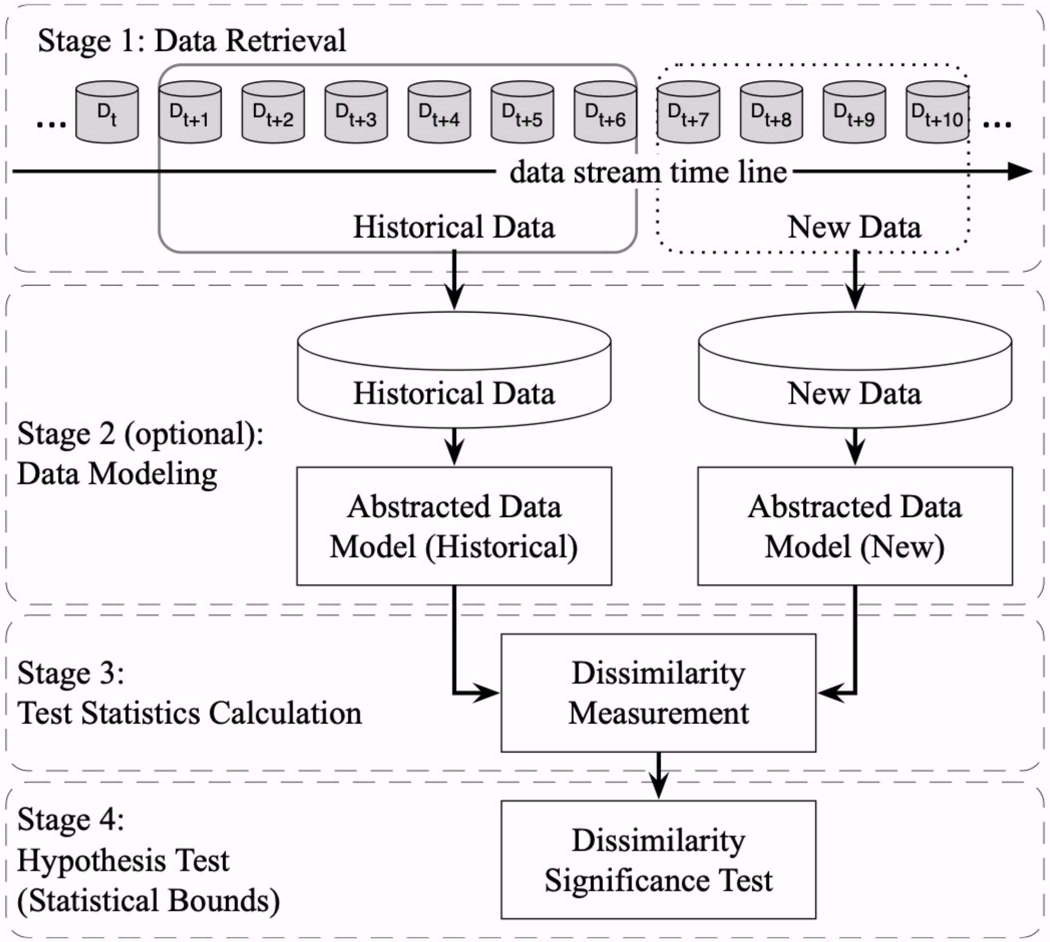
CD detection

- Stage 3: Test Statistics Calculation is the measurement of dissimilarity, or distance estimation. It quantifies the severity of the drift and forms test statistics for the hypothesis test.



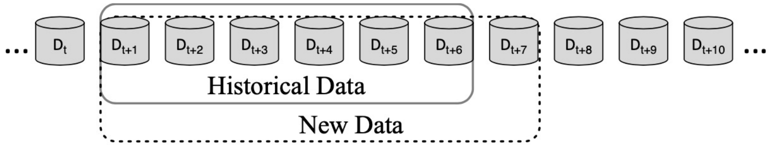
CD detection

- Stage 4: evaluate the statistical significance of the change observed in Stage 3. Used to determine drift detection accuracy by proving the statistical bounds of the test statistics proposed in Stage 3. Allows for answering the question of how likely it is that the change is caused by concept drift and not noise or random sample selection bias.
- The most commonly used hypothesis tests are: estimating the distribution of the test statistics, bootstrapping, the permutation test, and Hoeffding's inequality-based bound identification.



ADaptive WINdowing (ADWIN) Drift Detector

- ADWIN does not require users to define the size of the compared windows in advance;
- it only needs to specify the total size n of a “sufficiently large” window W.
- It then examines all possible cuts of W and computes optimal sub-window sizes n_hist and n_new according to the rate of change between the two sub-windows w_hist and w_new.

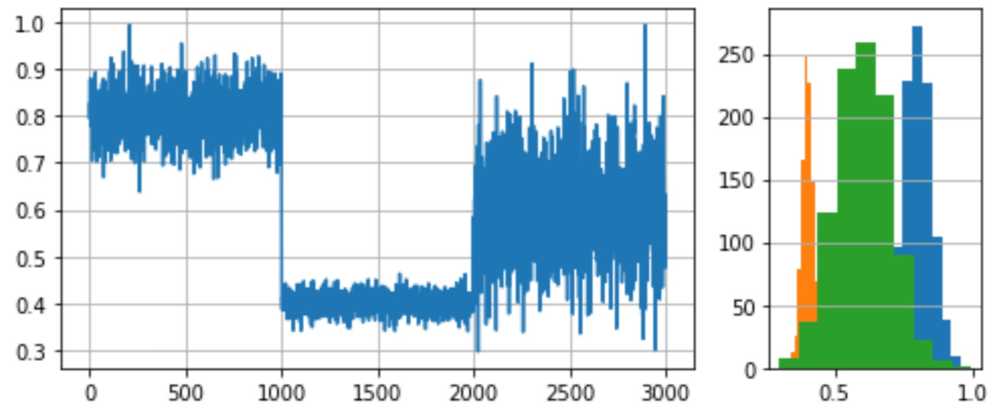


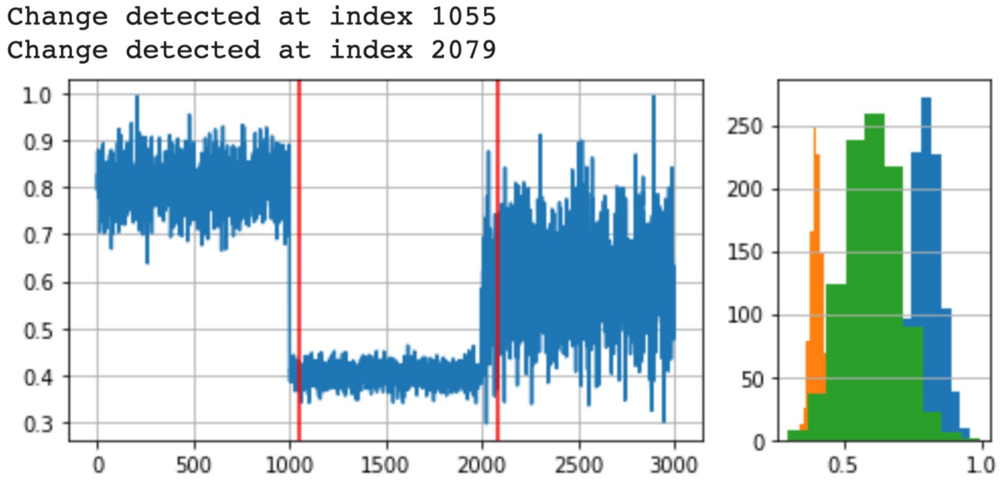
ADaptive WINdowing (ADWIN) Drift Detector

- The test statistic is the difference of the two sample means
  - $\theta\_ADWIN = |\hat{\mu}hist - \hat{\mu}new|$.
- An optimal cut is found when the difference exceeds a threshold with a predefined confidence interval δ.
- Both the false positive rate and false negative rate are bounded by δ.

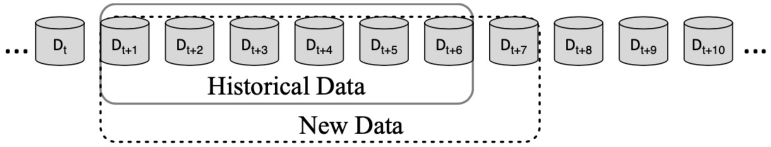
River Toolbox

- https://riverml.xyz





DDM (Drift Detection Method)



- First algorithm to define the warning level and drift level for concept drift detection.
- Stage 1 is implemented by a landmark time window. When a new data instance become available for evaluation, DDM detects whether the overall online error rate within the time window has increased significantly.
- If the confidence level of the observed error rate change reaches the warning level, DDM starts to build a new learner while using the old learner for predictions.
- If the change reached the drift level, the old learner will be replaced by the new learner for further prediction tasks.

DDM (Drift Detection Method) Continued

- If the change reached the drift level, the old learner will be replaced by the new learner for further prediction tasks.
- To acquire the online error rate, DDM needs a classifier to make the predictions.
- This process converts training data to a learning model, which is considered as the Stage 2 (Data Modeling).
- The test statistics in Stage 3 constitute the online error rate.
- The hypothesis test, Stage 4, is conducted by estimating the distribution of the online error rate and calculating the warning level and drift threshold.

Other drift detection methods

- Learning with Local Drift Detection (LLDD)
- Early Drift Detection Method (EDDM)
- Heoffding’s inequality based Drift Detection Method (HDDM)
- Fuzzy Windowing Drift Detection Method (FW-DDM)
- Dynamic Extreme Learning Machine (DELM)
- …
- See river.xyz for other possibilities

CD understanding

- Drift understanding refers to retrieving concept drift information about
  - “When” (the time at which the concept drift occurs and how long the drift lasts),
  - “How” (the severity /degree of concept drift),
  - “Where” (the drift regions of concept drift). This status information is the output of the drift detection algorithms, is used as input for drift adaptation.

CD adaptation

- Research into concept drift adaptation in Types 1-3 focuses on how to minimize the drop in accuracy and achieve the fastest recovery rate during the concept transformation process.
- In contrast, the study of Type 4 drift emphasizes the use of historical concepts, that is, how to find the best matched historical concepts with the shortest time. The new concept may suddenly, incrementally, or gradually reoccur.

CD adaptation

- simple retraining
  - An explicit concept drift detector is required to decide when to retrain the model
- ensemble retraining
  - Good for recurrent drift
- model adjusting
  - Incremental and online learning

Retraining: Paired learners

- two learners:
  - the *stable learner (trained with a lot of data)* and
  - the *reactive learner (trained with recent data, but with much less examples)*.
- If the stable learner frequently misclassifies instances that the reactive learner correctly classifies, a new concept is detected and the stable learner will be replaced with the reactive learner.

- Generally uses ADWIN

Model ensemble for recurring drift

- What are ensemble methods?
- When a drift is detected, learn one additional base learner
- Add the new base learner to the ensemble