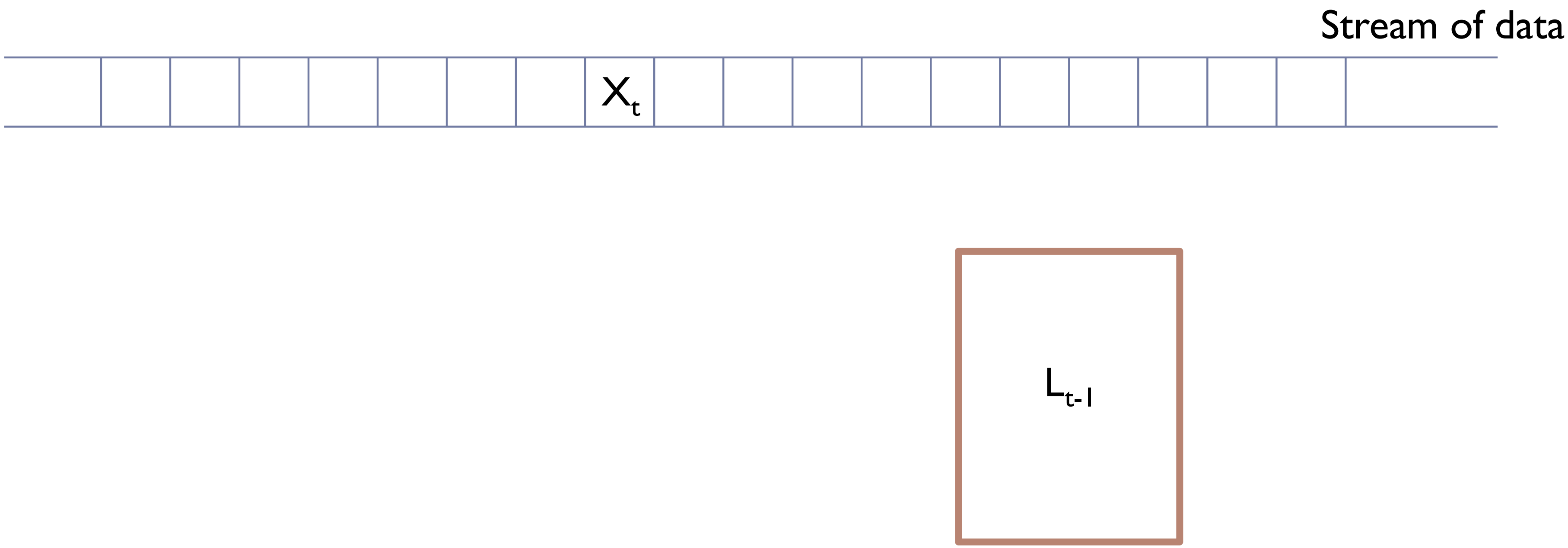
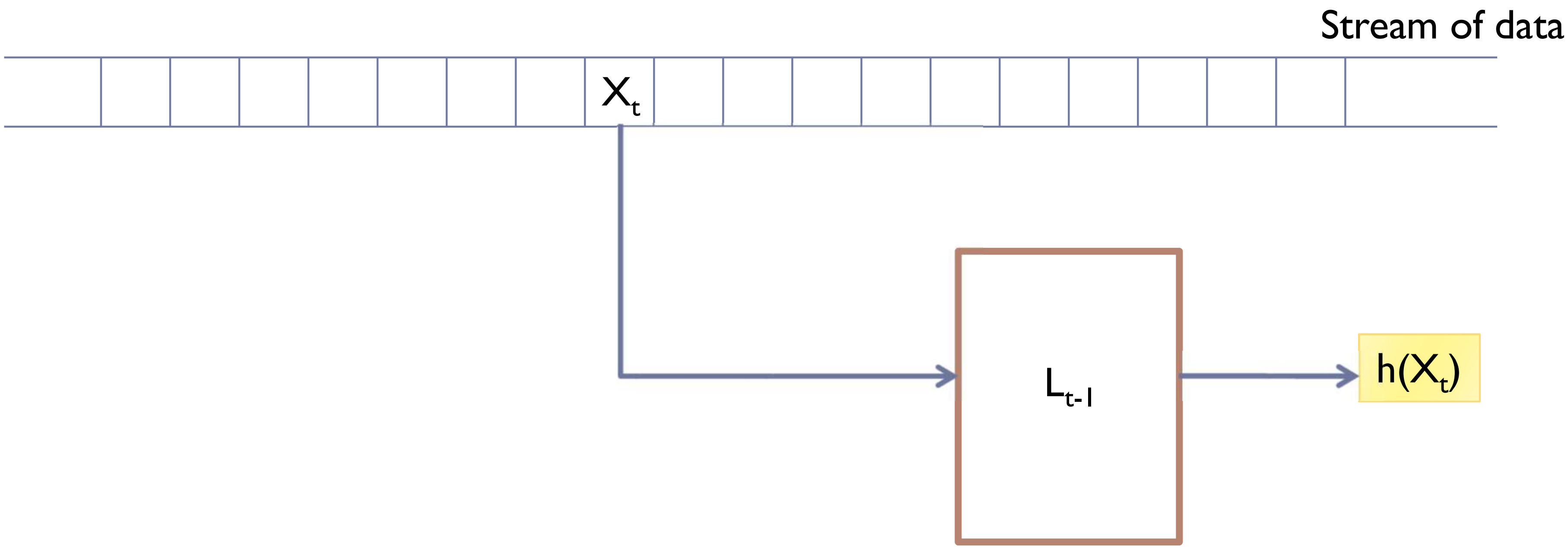
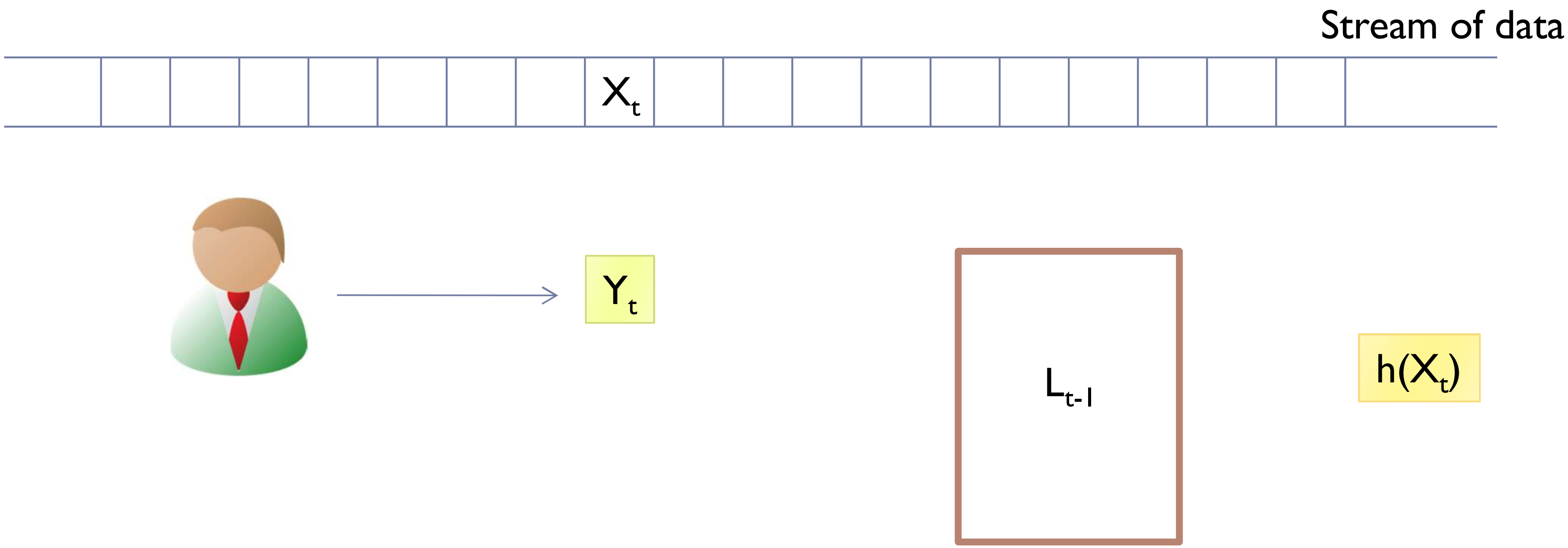
Data stream classification



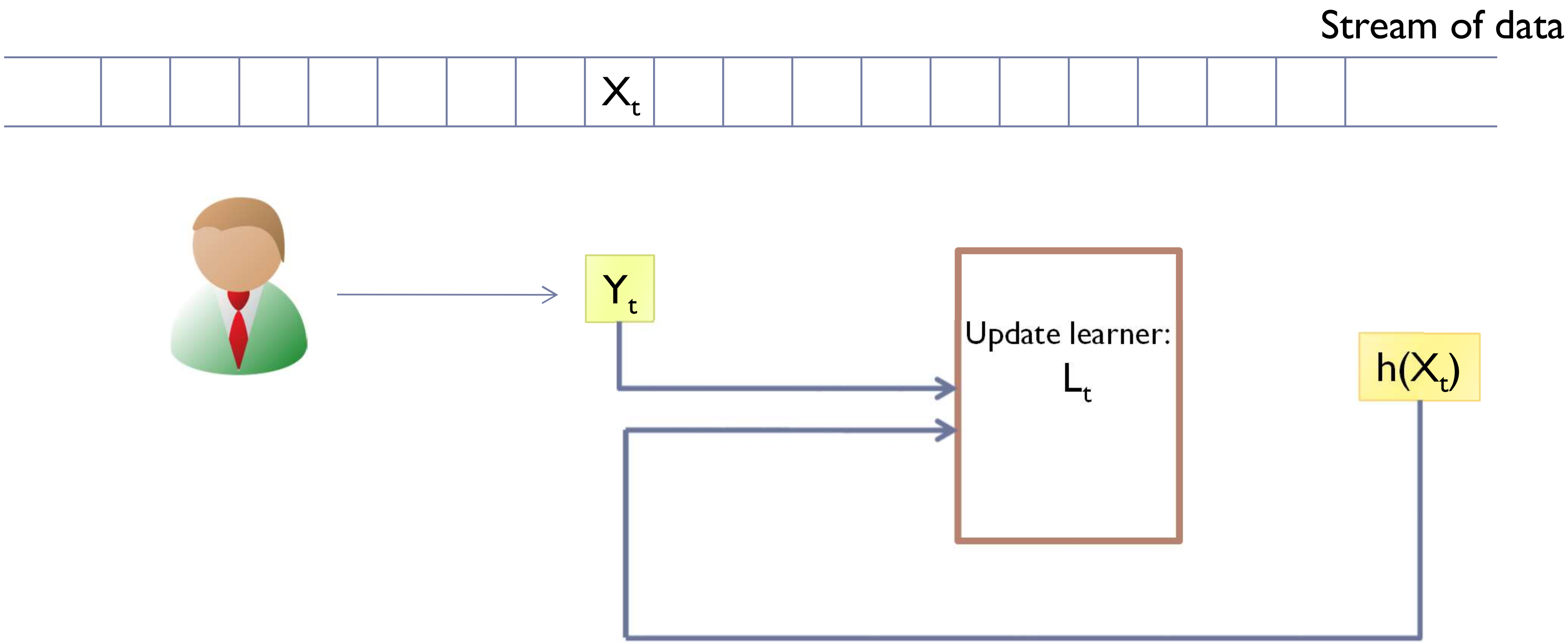
Data stream classification



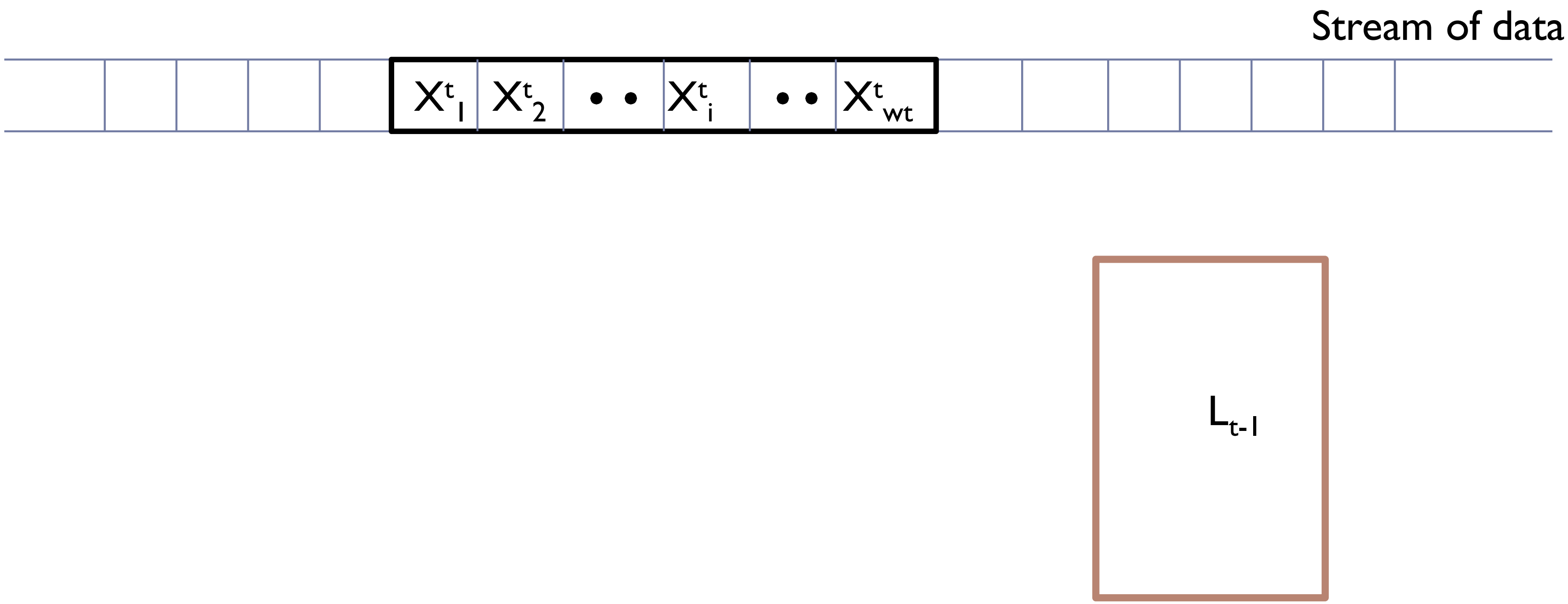
Data stream classification



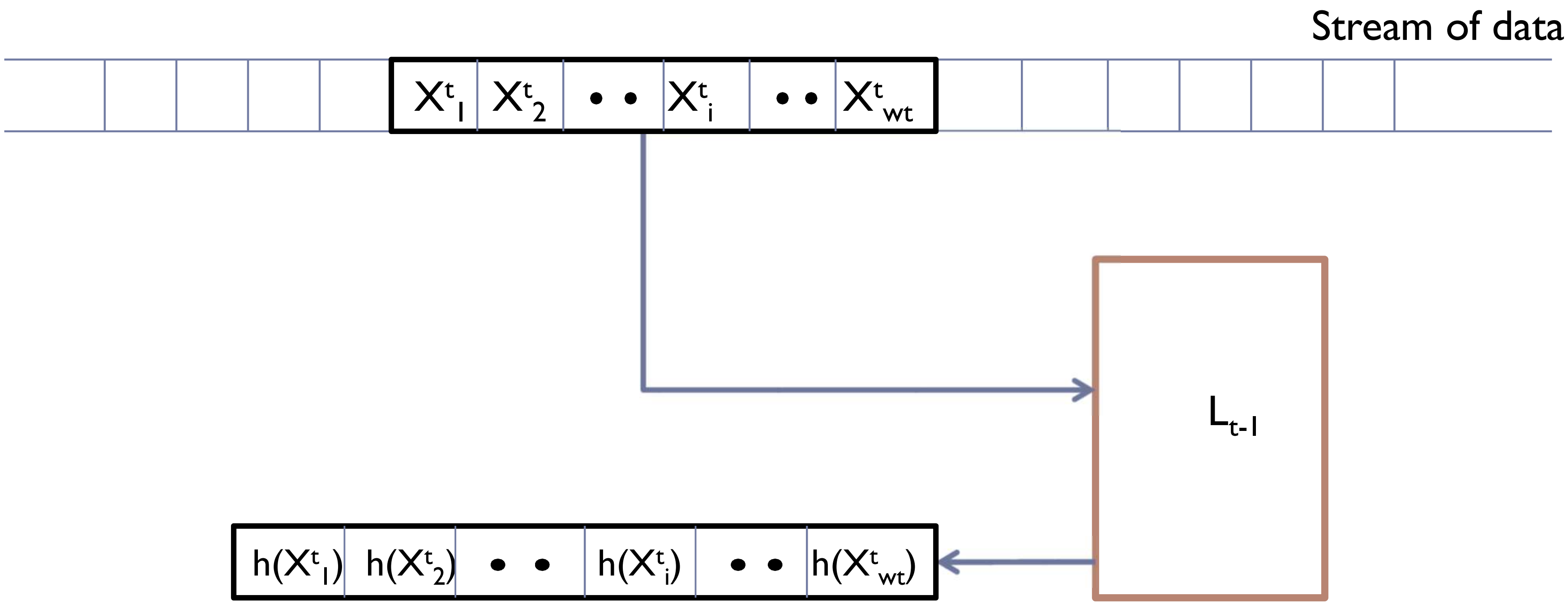
Data stream classification



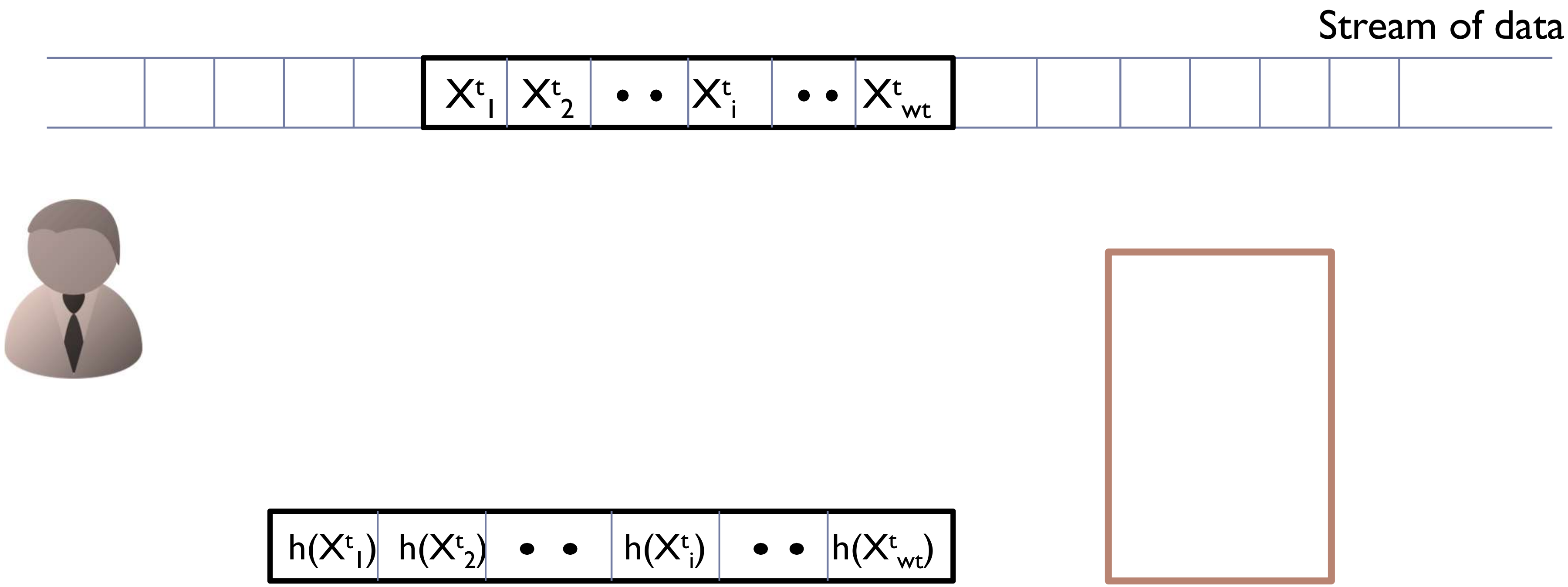
Data stream classification



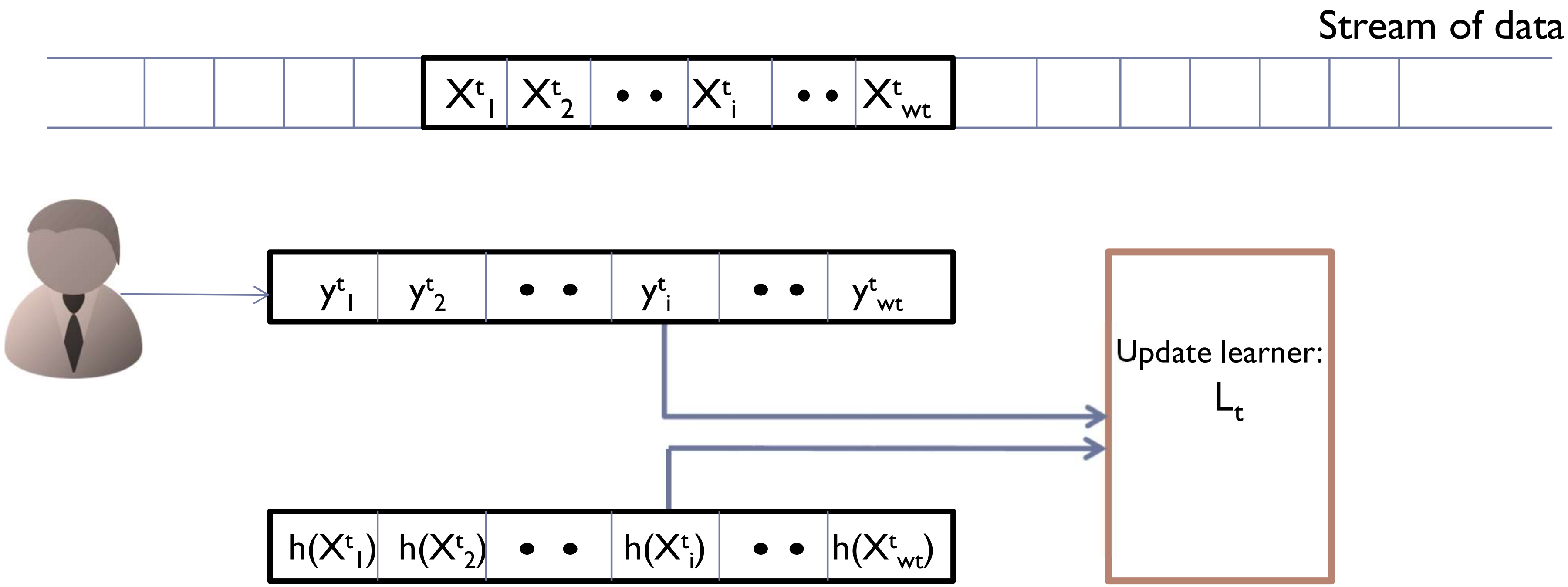
Data stream classification



Data stream classification



Data stream classification



Data stream challenges

- Unlimited stream of data → storage limit
  → labeling all data

Data stream challenges

- Unlimited stream of data → storage limit
  → labeling all data
- Online process of data → algorithms with low time complexity

Data stream challenges

- Unlimited stream of data → storage limit
  → labeling all data

- Online process of data → algorithms with low time complexity

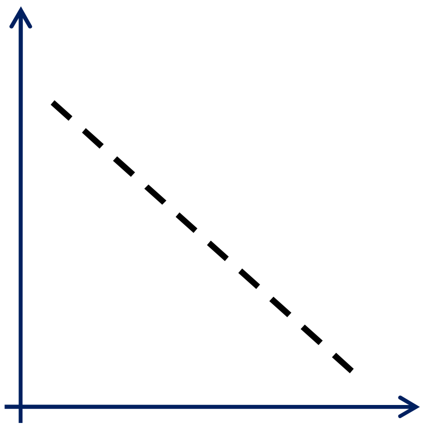
- Dynamic target functions (concept drift)

Concept drift

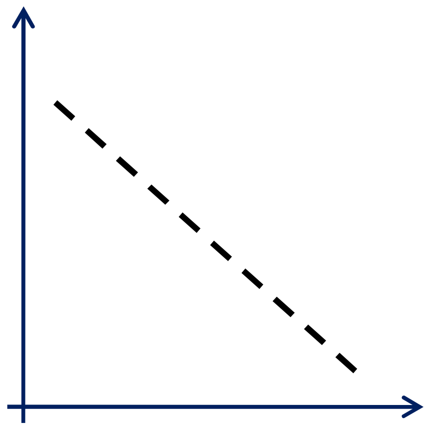
Instances which are near together in feature space, may belong to different target classes in different time passes

Example:

Binary classification of instances with two features



Online learning



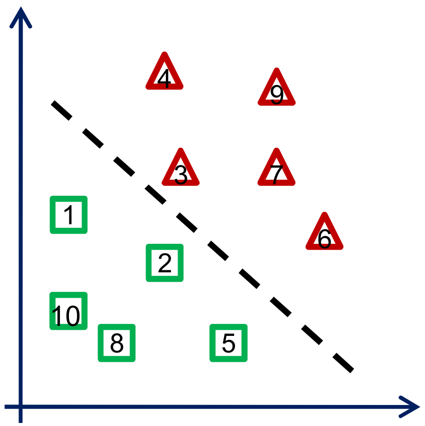
Online learning with concept drift

Concept drift

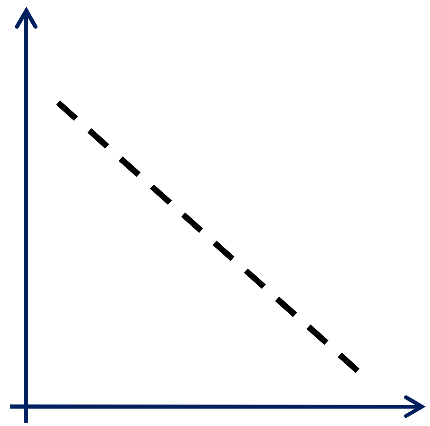
Instances which are near together in feature space, may belong to different target classes in different time passes

Example:

Binary classification of instances with two features



Online learning



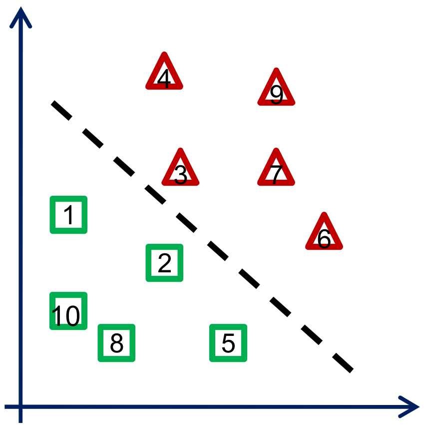
Online learning with concept drift

Concept drift

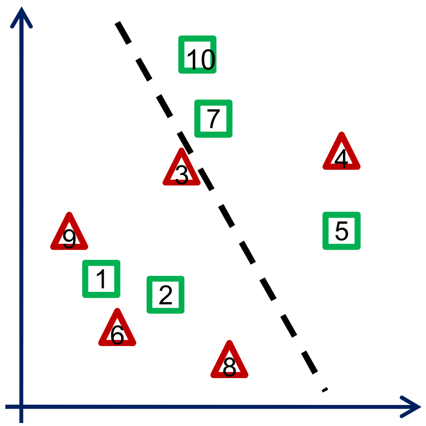
Instances which are near together in feature space, may belong to different target classes in different time passes

Example:

Binary classification of instances with two features



Online learning



Online learning with concept drift

$$P(X,C) = P(C|X) \times P(X)$$

Concept drift

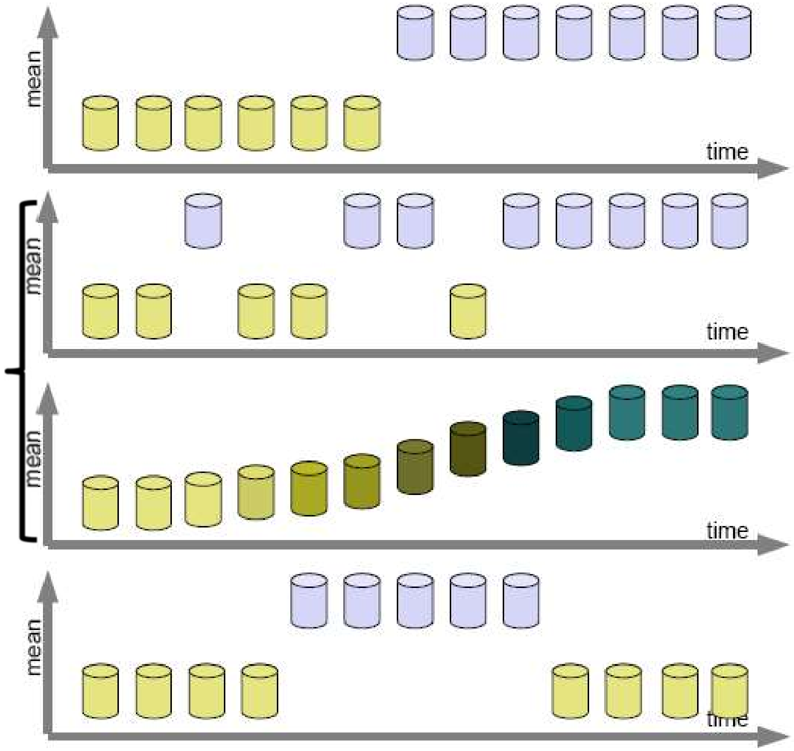
Instances which are near together in feature space, may belong to different target classes in different time passes

$$P(X,C) = P(C|X) \times P(X)$$

Abrupt drift

Gradual drift

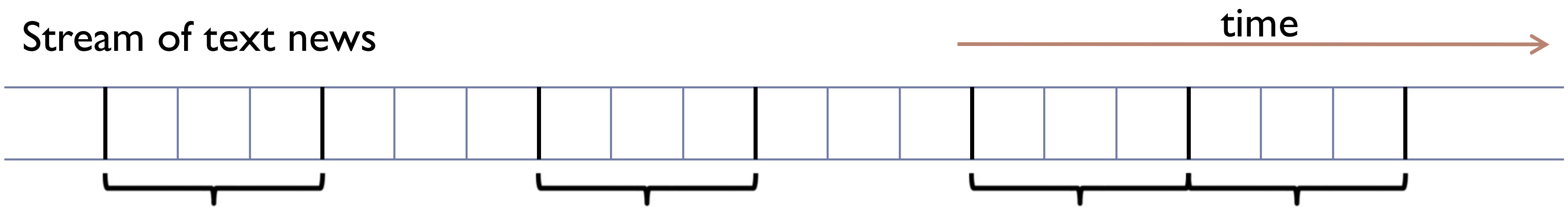
Recurring concepts



[Zliobaite, 2010]

Example for concept drift

- Purpose: detect hot news topics for a party

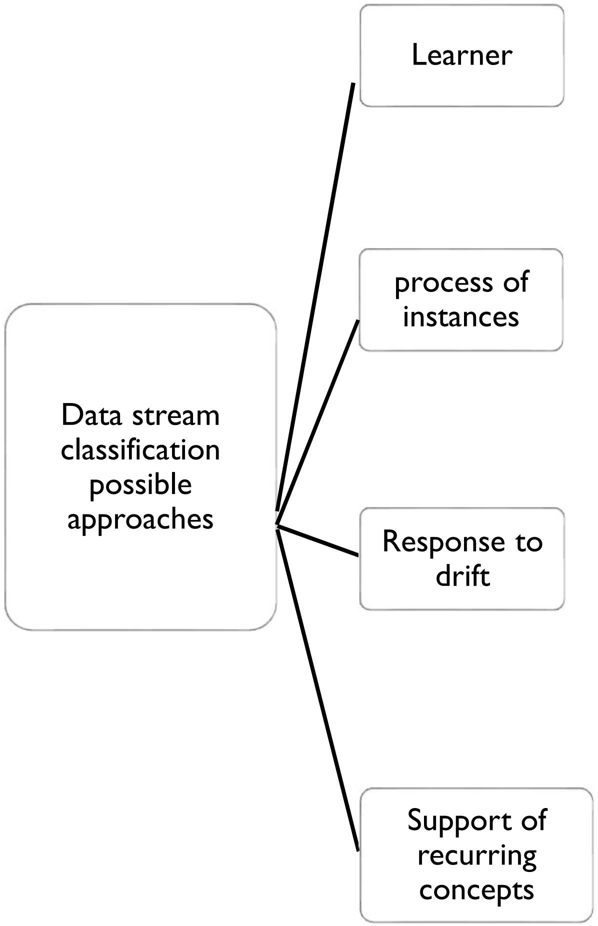


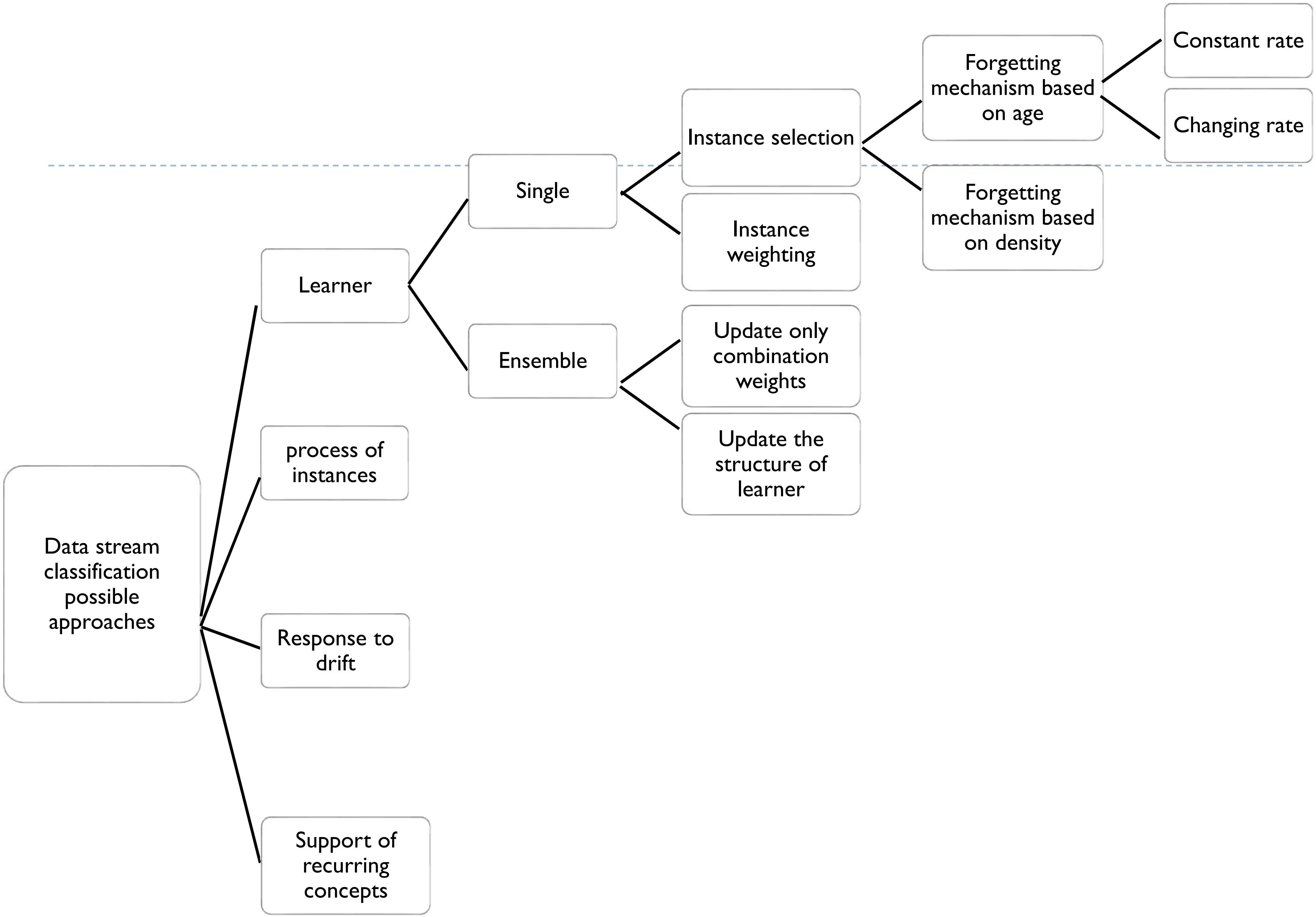
2016-01-18
afd men pictures germany members pegida petry fugitives leave …

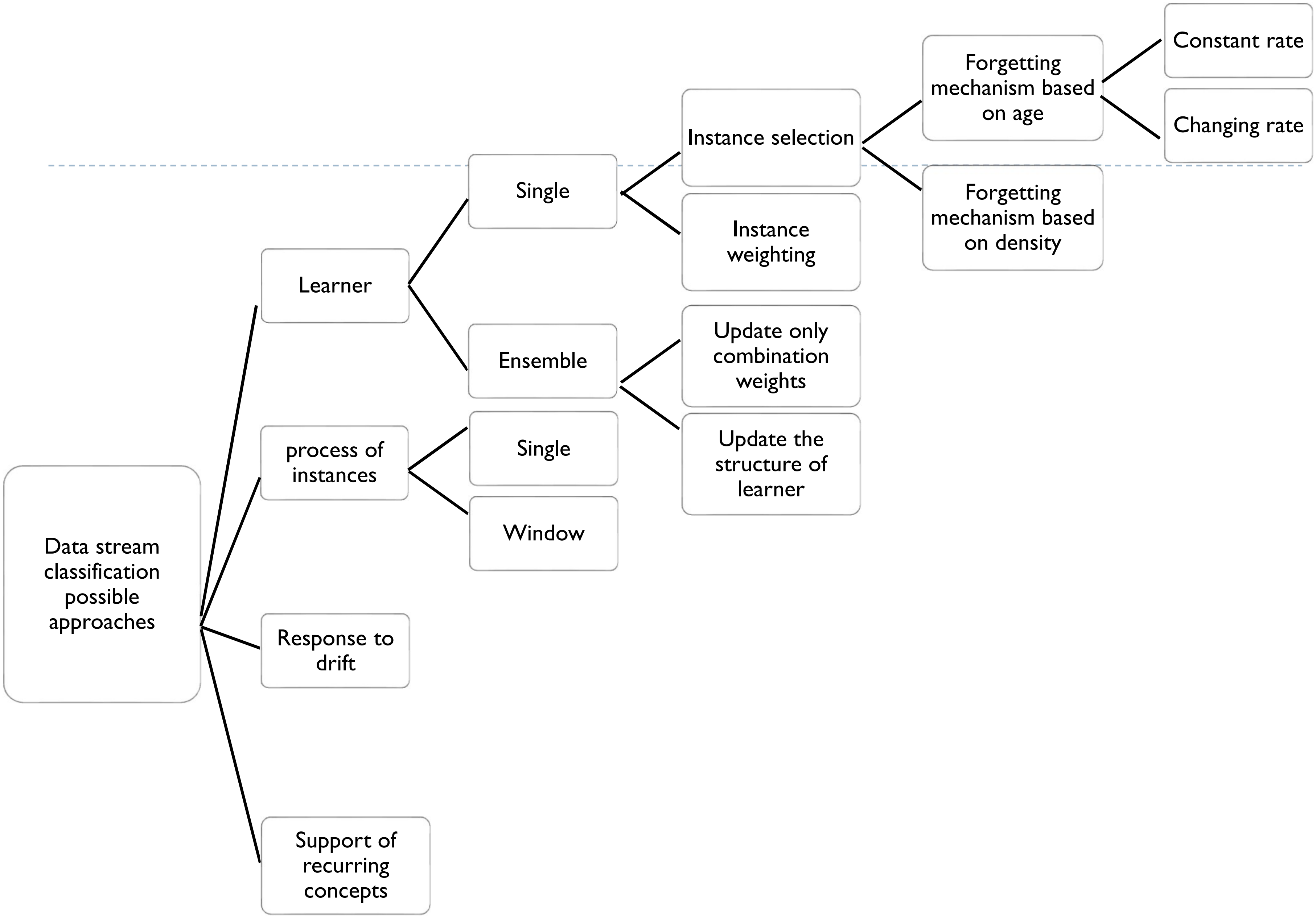
2016-07-05
afd **party_convention** april elected named remain perceived germany best racism relationships nationalism level effort color …

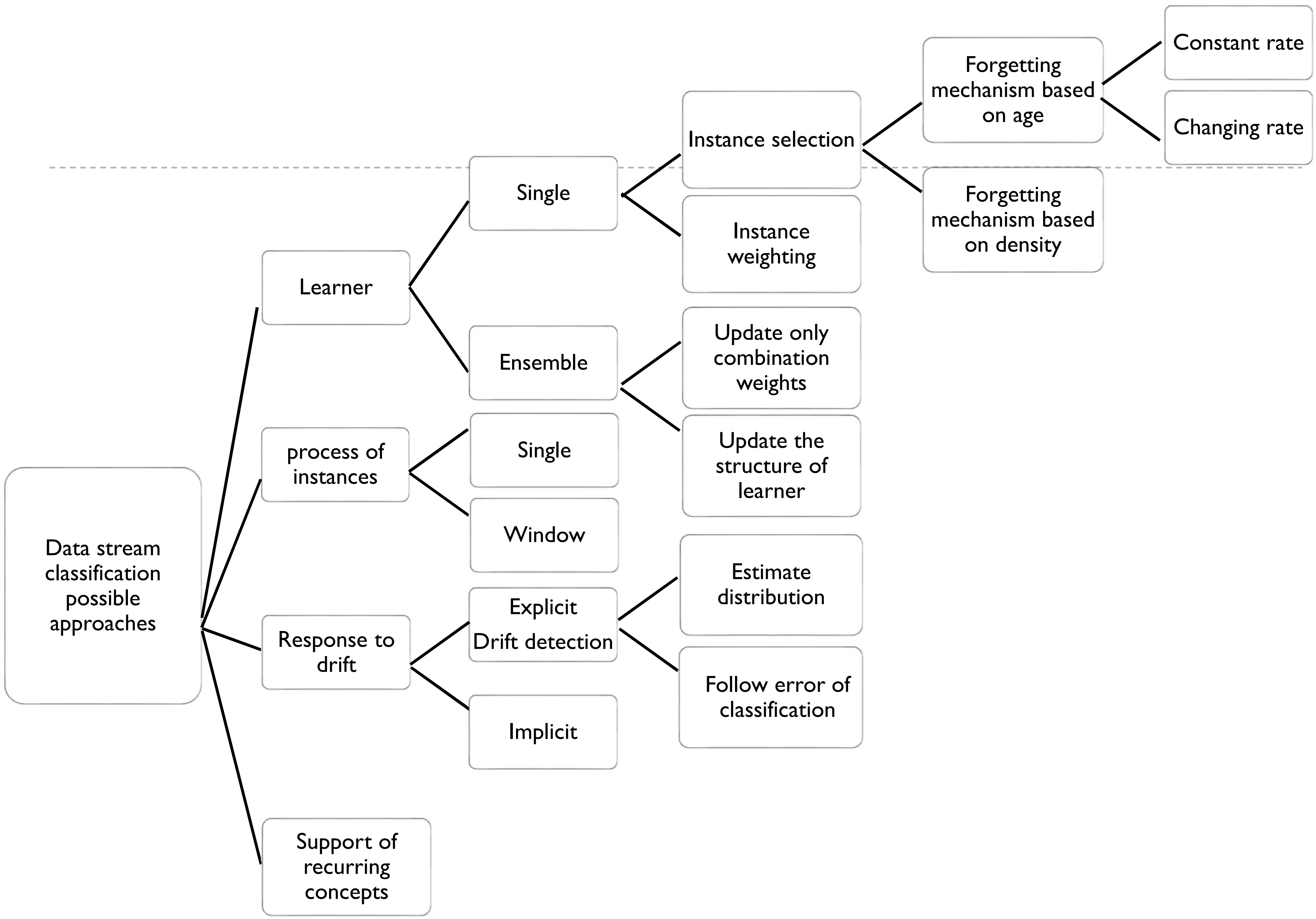
2017-01-11
afd party **bundestag_election** petry percent polls april poll **party_convention** frauke union lucke government cdu grünen linken …

2017-03-08
afd german members government majority keeps participates stop planned us president exit **climate_change** **paris** …


Data stream classification possible approaches
Learner
process of instances
Response to drift
Support of recurring concepts


Data stream classification possible approaches
Learner
process of instances
Response to drift
Support of recurring concepts
Single
Ensemble
Instance selection
Instance weighting
Update only combination weights
Update the structure of learner
Forgetting mechanism based on age
Forgetting mechanism based on density
Constant rate
Changing rate


Data stream classification possible approaches
Learner
process of instances
Response to drift
Support of recurring concepts
Single
Ensemble
Single
Window
Instance selection
Instance weighting
Update only combination weights
Update the structure of learner
Forgetting mechanism based on age
Forgetting mechanism based on density
Constant rate
Changing rate



Data stream classification possible approaches
Learner
Single
Instance selection
Forgetting mechanism based on age
Constant rate
Changing rate
Forgetting mechanism based on density
Instance weighting
Ensemble
Update only combination weights
Update the structure of learner
process of instances
Single
Window
Response to drift
Explicit Drift detection
Estimate distribution
Follow error of classification
Implicit
Support of recurring concepts

ADWIN: ADAptive WINdowing

- Keep a sliding window with the most recent observations
- A drift is reported whenever two "large enough" subwindows of W exhibit "distinct enough" averages
  - The older portion of the window is dropped

ADWIN: ADAPTIVE WINDOWING ALGORITHM

Initialize Window W
for each $t > 0$
    do $W \leftarrow W \cup \{x_t\}$ (i.e., add x_t to the head of W)
        repeat Drop elements from the tail of W
            until $|\hat{\mu}_{W_0} - \hat{\mu}_{W_1}| \geq \epsilon_{cut}$ holds
                for every split of W into $W = W_0 \cdot W_1$
        output $\hat{\mu}_W$

ADWIN: ADAptive WINdowing

ADWIN: ADAPTIVE WINDOWING ALGORITHM

```
Initialize Window W
for each t > 0
    do W ← W ∪ {x_t} (i.e., add x_t to the head of W)
        repeat Drop elements from the tail of W
          until |μ̂_W0 − μ̂_W1| ≥ ε_cut holds
            for every split of W into W = W_0 · W_1
        output μ̂_W
```

$$m = \frac{1}{1/n_0 + 1/n_1} \text{ (harmonic mean of } n_0 \text{ and } n_1\text{)}$$

$$\delta' = \frac{\delta}{n}, \text{ and } \quad \epsilon_{cut} = \sqrt{\frac{1}{2m} \cdot \ln \frac{4}{\delta'}}\ .$$

ADWIN: ADAptive WINdowing

ADWIN: ADAPTIVE WINDOWING ALGORITHM

Initialize Window W
for each $t > 0$
    do $W \leftarrow W \cup \{x_t\}$ (i.e., add x_t to the head of W)
        repeat Drop elements from the tail of W
            until $|\hat{\mu}_{W_0} - \hat{\mu}_{W_1}| \geq \epsilon_{cut}$ holds
                for every split of W into $W = W_0 \cdot W_1$
        output $\hat{\mu}_W$

length of W1

$$m = \frac{1}{1/n_0 + 1/n_1} \text{ (harmonic mean of } n_0 \text{ and } n_1)$$

length of W0

$$\delta' = \frac{\delta}{n}, \text{ and } \quad \epsilon_{cut} = \sqrt{\frac{1}{2m} \cdot \ln \frac{4}{\delta'}}\ .$$

n = n0 + n1

ADWIN for LDA

- The likelihood measures the generative quality of the model with respect to the observed data:

Given parameters of LDA

$$\mathcal{L}(\mathbf{w}) = \frac{\Gamma\left(\sum_i \alpha_i\right)}{\prod_i \Gamma(\alpha_i)} \int \left(\prod_{i=1}^{k} \theta_i^{\alpha_i - 1}\right) \left(\prod_{n=1}^{N} \sum_{i=1}^{k} \prod_{j=1}^{V} (\theta_i \beta_{ij})^{w_n^j}\right) d\theta$$

count of w_j in document n

Modeling recurring concepts