Streaming Machine Learning (SML)

Alessio Bernardo & Emanuele Della Valle

14-03-2022

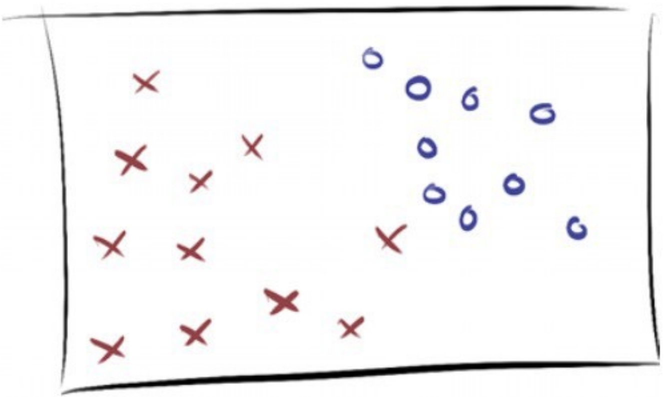
Part IV

Ensemble Classification

Credits

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

SML Ensemble Classification models



Ensemble Classifiers

*"An **ensemble** can be described as a **composition** of **multiple weak** learners to form one with (expected) **higher** predictive **performance** (strong learner), such that a weak learner is loosely defined as a learner that performs slightly better than random guessing"*

Freund and Schapire, 1997

Ensemble Classifiers

- **Diversity:** induce diversity among learners

- **Combination:** combine the predictions

- **Adaptation:** adapt to evolving data

Pro
- High Predictive performance
- Flexibility

Cons
- Computational resources

Gomes, H. M., Barddal, J. P., Enembreck, F., & Bifet, A. (2017). **A survey on ensemble learning for data stream classification.** ACM, 50(2), 1-36.

Induce Diversity

Horizontal Partitioning

- **Bagging:** build a set of ***M*** base models, with a bootstrap sample from the original dataset of size ***N***, created by drawing random samples with replacement. Each bootstrap contains each original sample ***K*** times, where ***Pr(K=k)*** follows a binomial distribution.

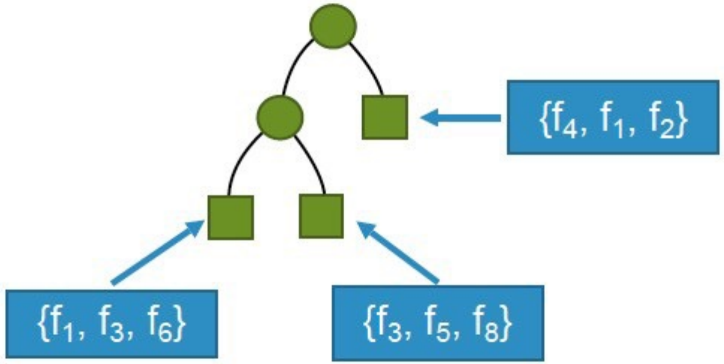
Gomes, H. M., Barddal, J. P., Enembreck, F., & Bifet, A. (2017). **A survey on ensemble learning for data stream classification.** ACM, 50(2), 1-36.

Induce Diversity

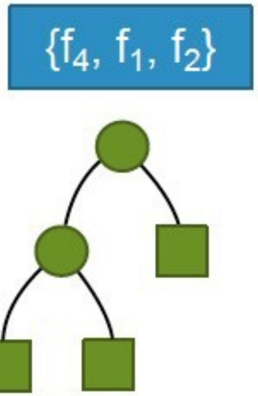
Vertical Partitioning

- **Random Subspaces:** train learners on different subsets of features

Local Randomization



Global Randomization



Gomes, H. M., Barddal, J. P., Enembreck, F., & Bifet, A. (2017). **A survey on ensemble learning for data stream classification.** ACM, 50(2), 1-36.

Induce Diversity

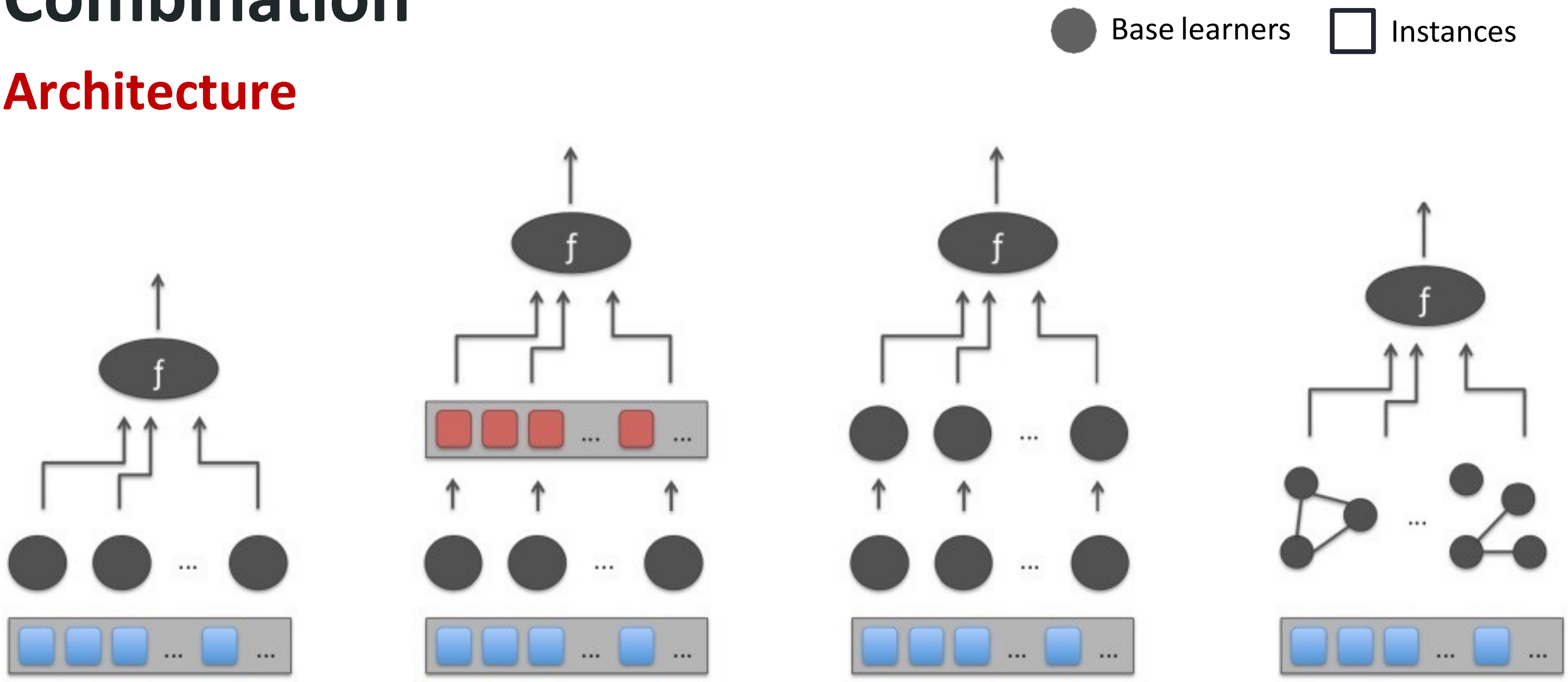
Others

- **Base Learner Manipulation:** varying parameters of the same base learner
- **Heterogeneous Base Learners:** use heterogeneous base learners and obtain ensemble members with different biases

Gomes, H. M., Barddal, J. P., Enembreck, F., & Bifet, A. (2017). **A survey on ensemble learning for data stream classification.** ACM, 50(2), 1-36.

Combination

Architecture



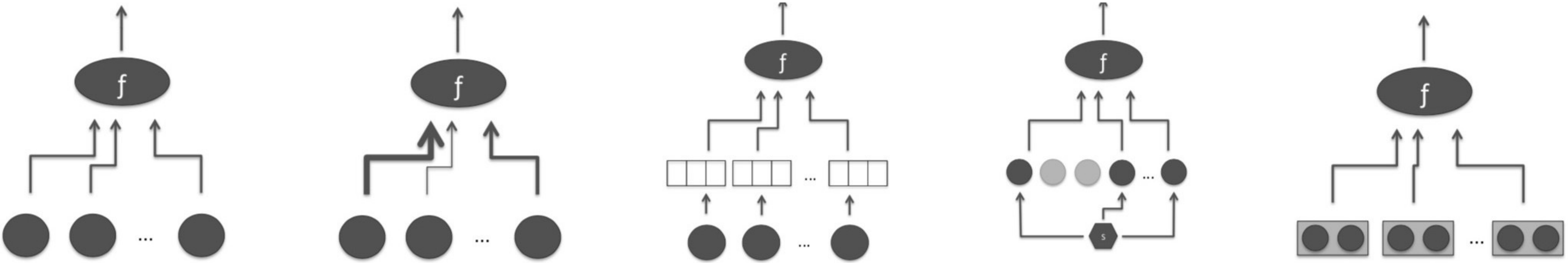
Flat **Meta-Learner** **Hierarchical** **Network**

Gomes, H. M., Barddal, J. P., Enembreck, F., & Bifet, A. (2017). **A survey on ensemble learning for data stream classification.** ACM, 50(2), 1-36.

Combination

Voting

Base learners Instances



Majority **Weighted Majority** **Rank** **Abstaining** **Relational**

Gomes, H. M., Barddal, J. P., Enembreck, F., & Bifet, A. (2017). **A survey on ensemble learning for data stream classification.** ACM, 50(2), 1-36.

Adaptation

Cardinality

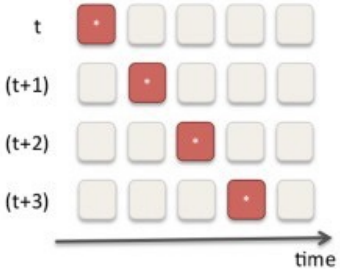
- **Fixed:** fixed numbers of base learners
- **Dynamic:** add classifiers on the fly

Gomes, H. M., Barddal, J. P., Enembreck, F., & Bifet, A. (2017). **A survey on ensemble learning for data stream classification.** ACM, 50(2), 1-36.

Adaptation

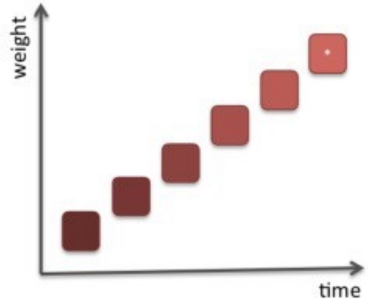
Learning Mode

Instances



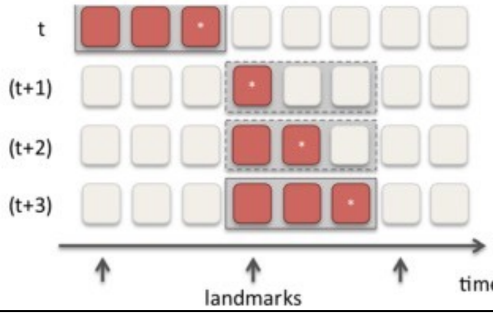
Incremental

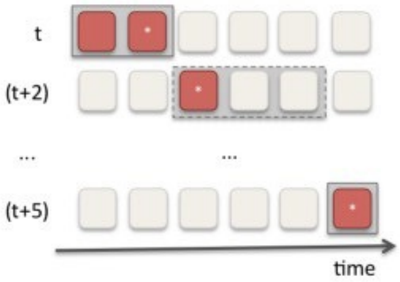
Sliding window



Damped Window

Landmark window



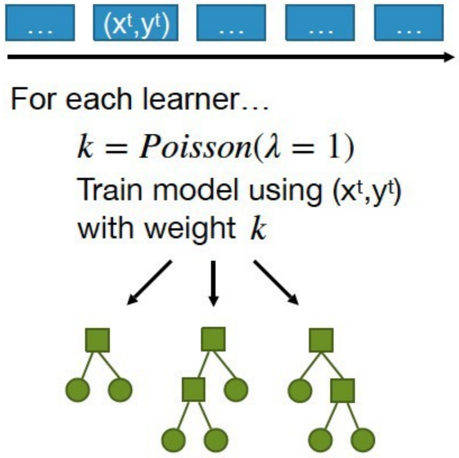


Adaptive window

Gomes, H. M., Barddal, J. P., Enembreck, F., & Bifet, A. (2017). **A survey on ensemble learning for data stream classification.** ACM, 50(2), 1-36.

Online Bagging

- Since data streams are supposed to be unbounded (large ***N***), the binomial distribution tends to a **Poisson(1)** distribution.

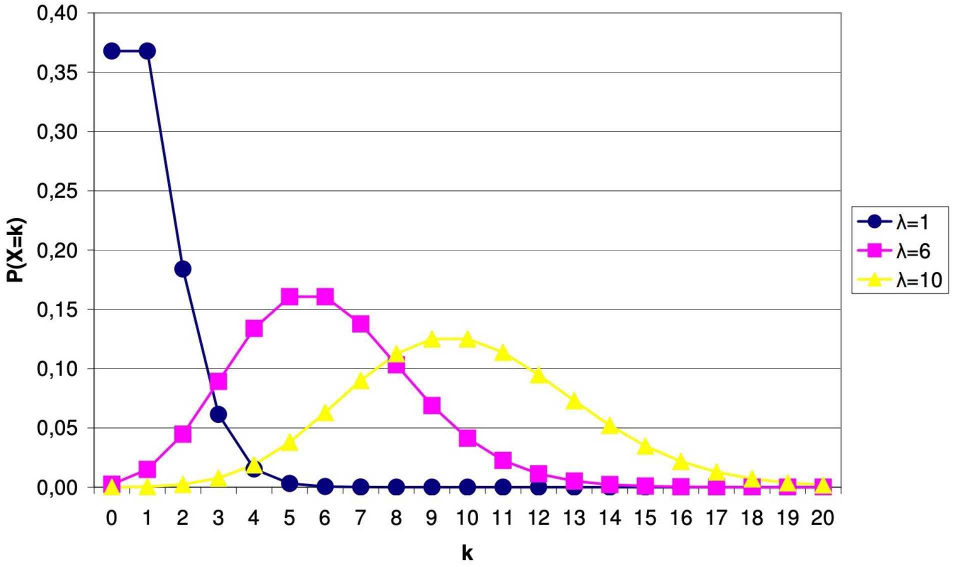


Train learners on different subsets of instances

Oza and Russel, **“Online bagging and boosting,”** in Artificial Intelligence and Statistics 2001.

Leveraging Bagging

- Add an **ADWIN** drift detector per base learner
- Use more weight during training - **Poisson(6)**



Bifet, G. Holmes, and B. Pfahringer, **“Leveraging bagging for evolving data streams,”** in PKDD, 2010

Adaptive Random Forest (ARF)

- **Base Learners**: Hoeffding Trees
- **Diversity:** Leveraging Bagging + **Local** Random Subspaces
- **Combination:**
  - Flat architecture
  - Weighted majority voting
- **Adaptation:** Adaptive window + warning period (train background learners)

H. M. Gomes et al, **“Adaptive random forests for evolving data stream classification,”** Machine Learning, 2017.

Streaming Random Patches (SRP)

- **Base Learners**: User choice
- **Diversity:** Leveraging Bagging + **Global** Random Subspaces
- **Combination:**
  - Flat architecture
  - Weighted majority voting
- **Adaptation:** Adaptive window + warning period

Gomes, Read and Bifet, **"Streaming Random Patches for Evolving Data Stream Classification"**, ICDM, 2019

QUIZ

1. What is the difference between Online Bagging and Leveraging Bagging?
    a. They give the same weights to the instances
    b. The former gives higher weights to the instances, inducing more diversity
    c. The latter gives higher weights to the instances, inducing more diversity
2. What are the **2** most important differences between ARF and SRP?
    a. ARF uses only HT as base learners and leveraging bagging, SRP uses HT as base learners and online bagging
    b. ARF uses ADWIN and local random subspaces, SRP does not use any CD detector and uses global random subspaces
    c. ARF uses only HT as base learners and local random subspaces, SRP can use everything as base learners and global random subspaces

QUIZ

1. What is the difference between Online Bagging and Leveraging Bagging?
    a. They give the same weights to the instances
    b. The former gives higher weights to the instances, inducing more diversity
    c. The latter gives higher weights to the instances, inducing more diversity
2. What are the **2** most important differences between ARF and SRP?
    a. ARF uses only HT as base learners and leveraging bagging, SRP uses HT as base learners and online bagging
    b. ARF uses ADWIN and local random subspaces, SRP does not use any CD detector and uses global random subspaces
    c. ARF uses only HT as base learners and local random subspaces, SRP can use everything as base learners and global random subspaces

QUIZ

1. What is the difference between Online Bagging and Leveraging Bagging?
   a. They give the same weights to the instances
   b. The former gives higher weights to the instances, inducing more diversity
   c. **The latter gives higher weights to the instances, inducing more diversity**
2. What are the **2** most important differences between ARF and SRP?
   a. ARF uses only HT as base learners and leveraging bagging, SRP uses HT as base learners and online bagging
   b. ARF uses ADWIN and local random subspaces, SRP does not use any CD detector and uses global random subspaces
   c. **ARF uses only HT as base learners and local random subspaces, SRP can use everything as base learners and global random subspaces**

EXERCISE 4: Stream Ensemble Classification
LAB 4: Final Challenge