Online ensembles for non-stationary data streams

Rohit Uttamchandani

MUIA 2018/2019

Index

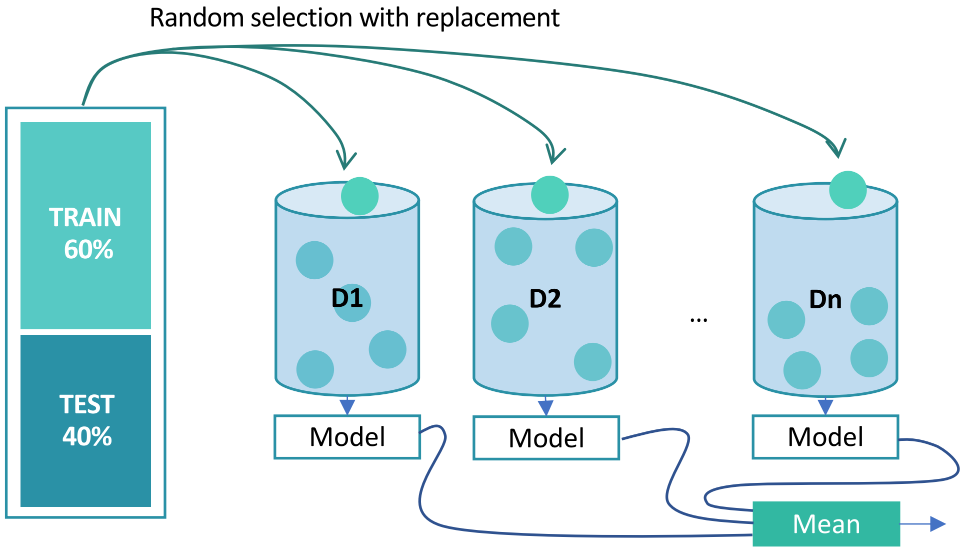
1. Introduction & Taxonomy
2. Active Approaches
3. Reactive Approaches
4. Advanced issues
5. Conclusion and future research

1. INTRODUCTION

- Online Machine Learning (chunk/one-at-a-time)
- Non/stationary data streams
- Concept drift (active/reactive)
- Plasticity (learn new concepts) / Stability (retaining knowledge)
- Diversity in ensembles

1. INTRODUCTION

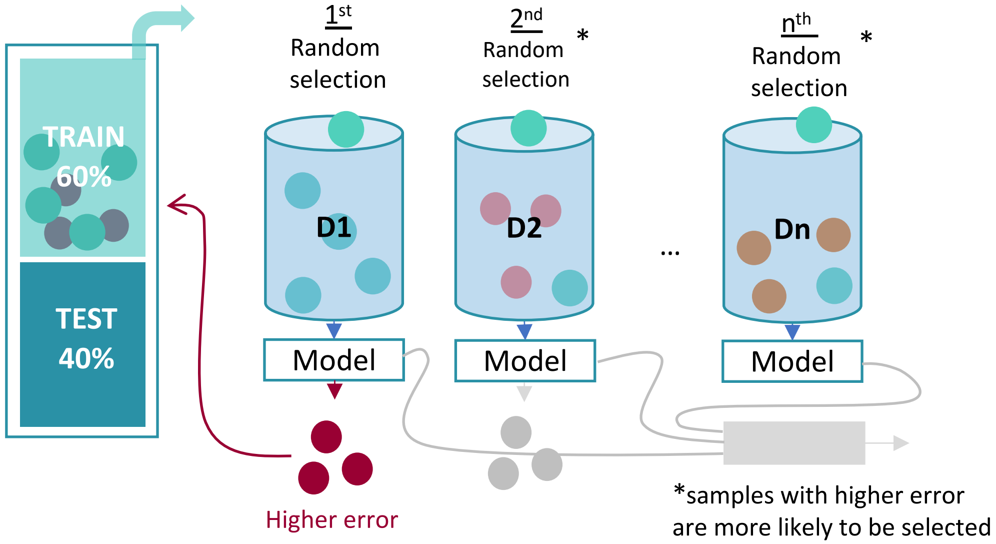
BAGGING



$$P(K=k) = \binom{N}{k}\left(\frac{1}{N}\right)^{k}\left(1-\frac{1}{N}\right)^{N-k}\Bigg|_{N\to\infty} \Rightarrow K \sim \frac{\exp(-1)}{k!} \qquad \underset{(\lambda=1)}{Poisson}$$

ONLINE BAGGING (2001)

BOOSTING

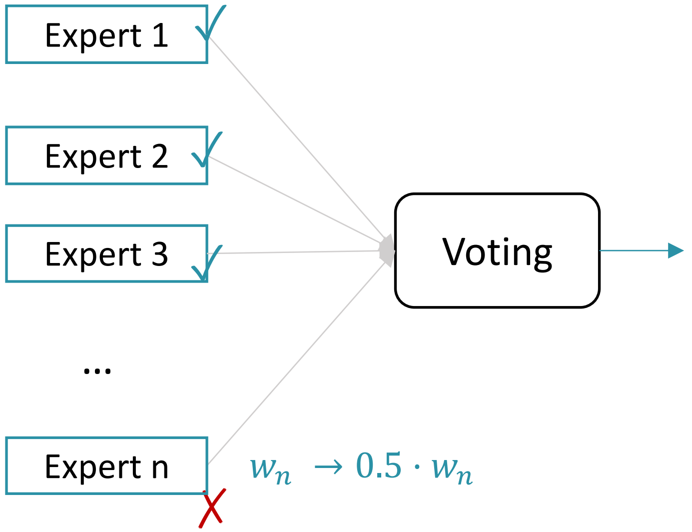


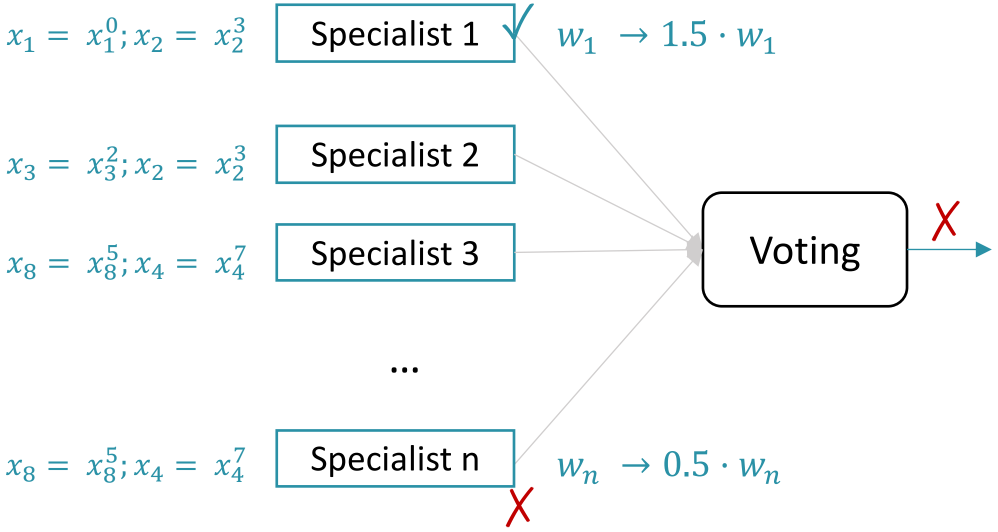
$$P(K=k) \sim \underset{(\lambda=1)}{Poisson} \Rightarrow \begin{cases} \text{Correct classification,} & \lambda \text{ is decreased} \\ \text{Misclassification,} & \lambda \text{ is increased} \end{cases}$$

ONLINE BOOSTING (2001)

1. INTRODUCTION

"Learning simple things really well."- Blum (1997)



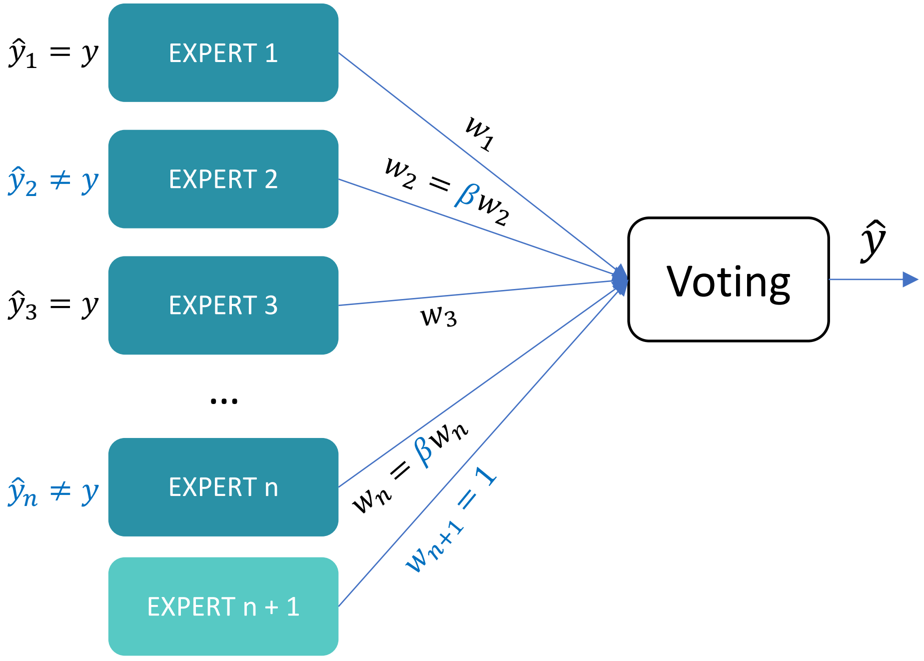


$\binom{x}{2}$: # models $\quad$ x : # features

WEIGHTED MAJORITY(1988)

WINNOW SPECIALIST (1997) – version of WINNOW (1988)

2. REACTIVE APPROACHES



Hyperparameters:

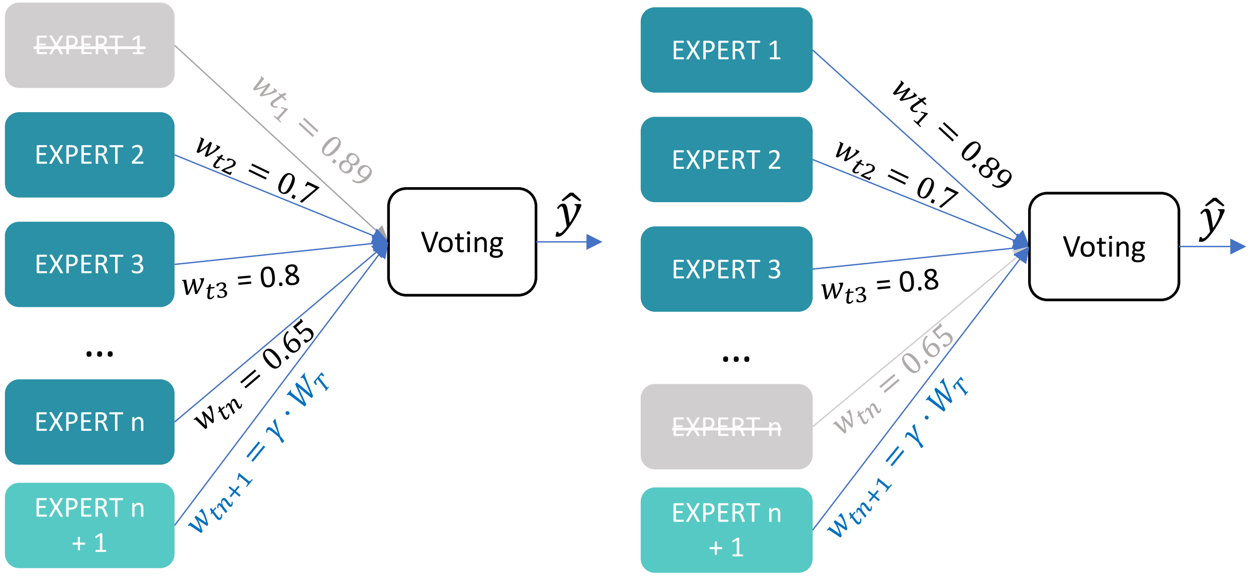
$\beta\ (reduction\ factor), \theta(threshold), p(period)$

MISTAKE'S BOUND

$$M_{t_2} - M_{t_1} \le \frac{\log(W_{t_1}/W_{t_2})}{\log(\frac{2}{1+\beta+2\gamma})}$$

$$where\ \beta + 2 \cdot \gamma < 1$$

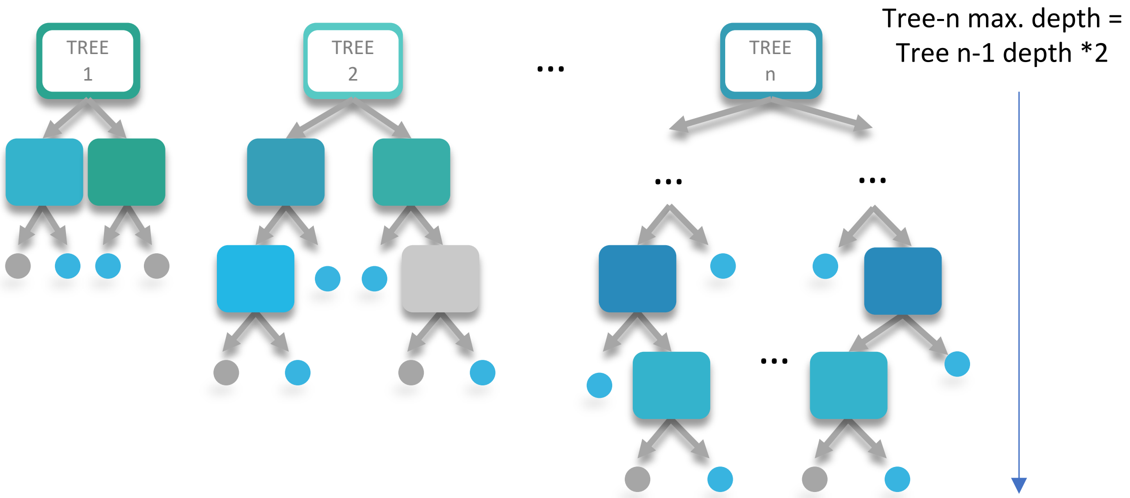
PRUNING TECHNIQUES:



Dynamic Weighted Majority (2003)

Additive Expert Ensembles (2005)

2. REACTIVE APPROACHES



HOEFFDING'S BOUND

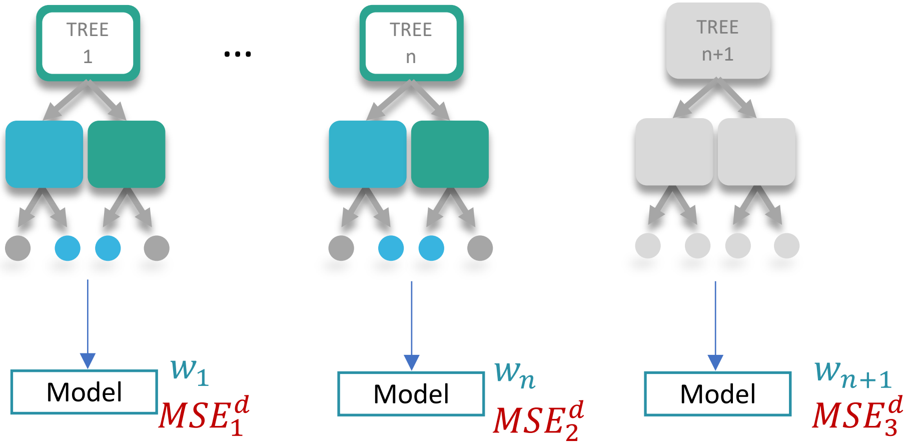
$$P(\overline{X} - E[\overline{X}] \geq t) \leq e^{-2\pi t^2}$$

DELETE NODES

Two options:

- Delete oldest node (top)
- Restart Tree

Adaptive Size Hoeffding Trees (2009)



CHUNKS or BLOCKS (size *d*)

B_1	B_2	B_3	B_4	...	B_n

Accuracy Updated Ensemble (2014)

ONLINE VERSION

- Windowing strategy (size d)
- AUE + incremental learner (last d)
- Drift detector (Active approach)

WEIGHTING FUNCTION

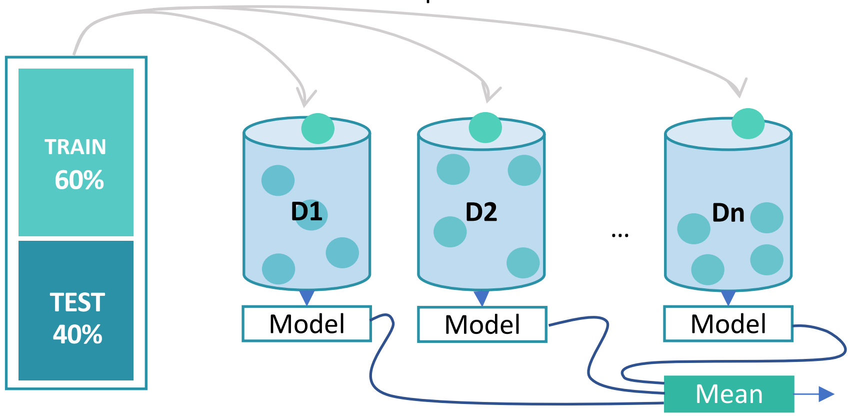
$$f(\tau_{created}, MSE_i^d, d, p_{iy}^t(x^t))$$

Online Accuracy Updated Ensemble (2014)

3. ACTIVE APPROACHES

ONLINE BAGGING

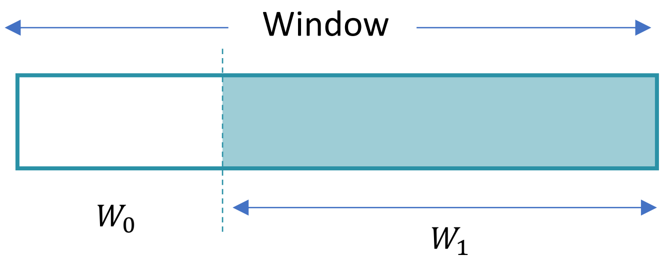
Random selection with replacement



$$P(K=k) = \binom{N}{k}\left(\frac{1}{N}\right)^k\left(1-\frac{1}{N}\right)^{N-k}\Bigg|_{N\to\infty} \Rightarrow K \sim \frac{\exp(-1)}{k!}$$

Poisson ($\lambda = 1$)

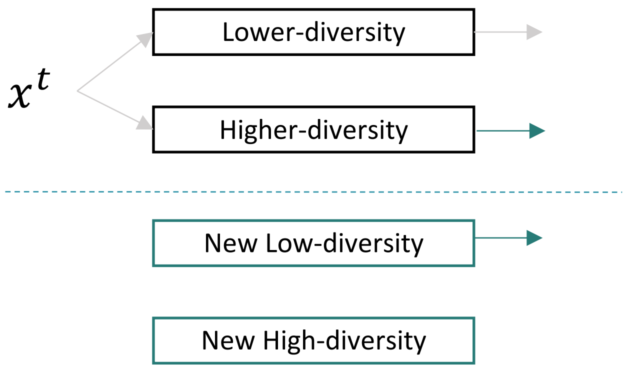
ADWIN



ADWINBagging (2001)

Q-STATISTIC

$$Q_{i,k} = \frac{N^{11}N^{00} - N^{01}N^{10}}{N^{11}N^{00} + N^{01}N^{10}}$$

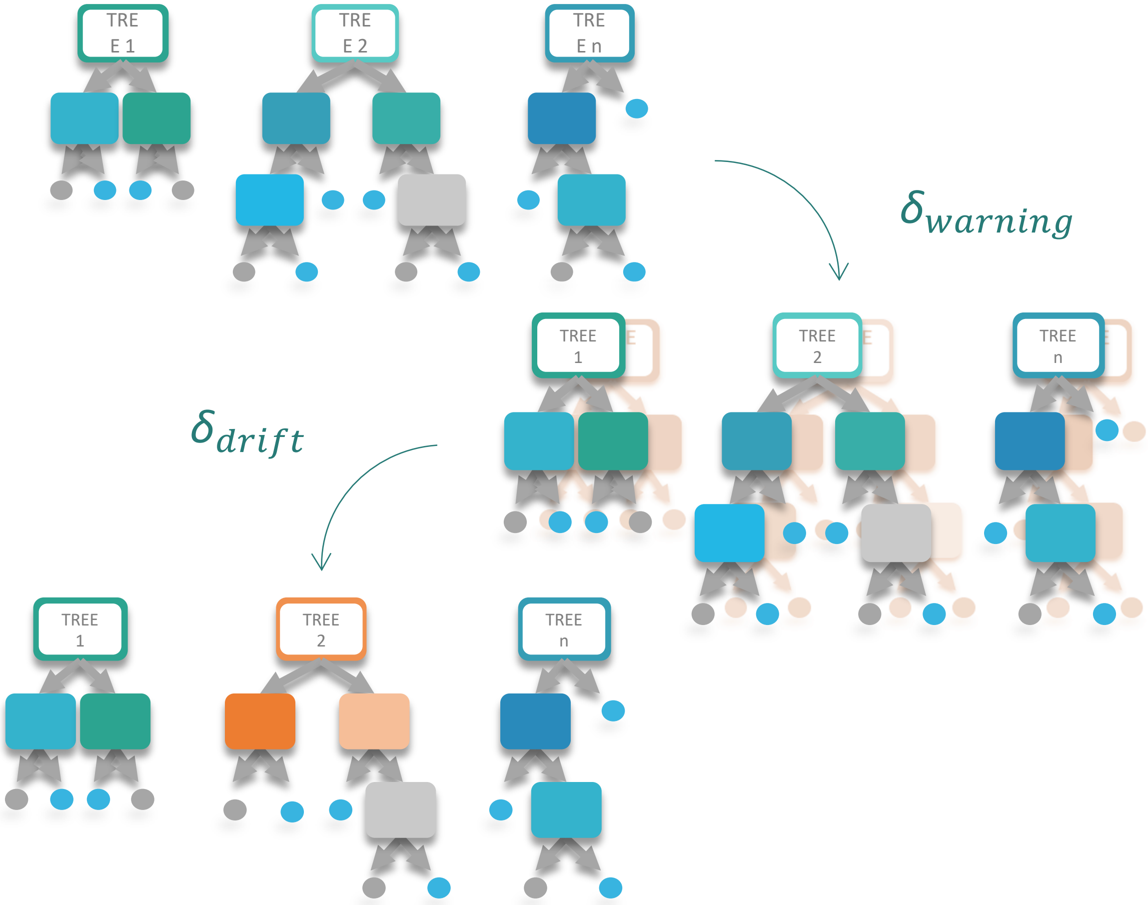


DRIFT DETECTED

Hyperparameters: p_l $(diversity)$, $drift\ detector$, $W_{old/new}$

Diversity for Dealing with Drifts (2012)

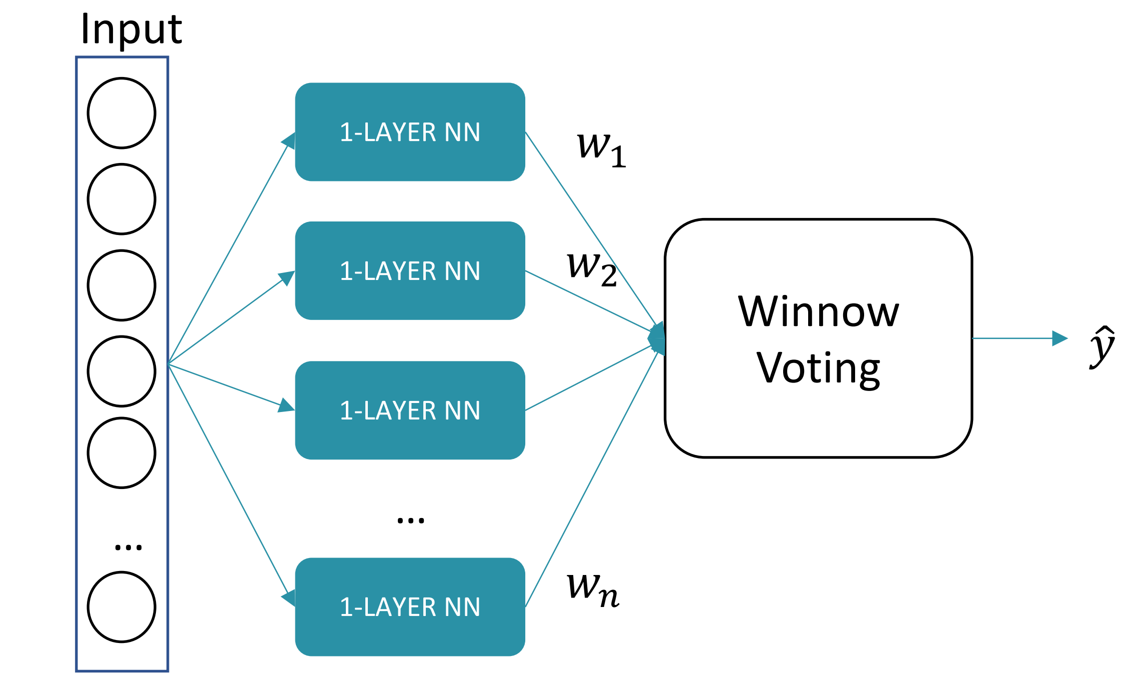
3. ACTIVE APPROACHES



- Hoeffding Trees
- ADWIN
- (δ_w): warning sent
- (δ_d) : drift is confirmed

Adaptive random forests for evolving data stream classification (2017)

4. OTHER APPROACHES



$$COST\ FUNCTION_j = \sum_{s=1}^{s} c_j(s) \left(f_j'(d_{js}) \left(f_j^{-1}(d_{js}) - \sum_{i=0}^{l} w_{ji} x_{is} \right) \right)^2$$

Ensemble of online neural networks for non-stationary and imbalanced datastreams (2013)

<u>Original Winnow algorithm:</u>

If $\hat{y} = y$:

  Do nothing

Elif $\hat{y}$ =1 and y=0:

  $\forall x_i = 1, w_i = 0$ (demotion)

Elif $\hat{y}$ =0 and y=1:

  $\forall x_i = 1, w_i = \alpha \cdot w_i$ (promotion)

<u>Proposed Winnow algorithm:</u>

Separate weights for each class (wmin and wmaj):

$$\hat{y} = argmax \left(\sum_{j=1}^{n} (p_j^t = 1) \cdot (\gamma \cdot wmin_j + wmaj_j) + (p_j^t = 2) \cdot (wmin_j + wmaj_j) \right)$$

Three parameters: α (promotion), β (demotion) and γ (imbalance)

5. CONCLUSION & FUTURE RESEARCH

- Extreme class imbalance (anomaly detection)
- Delayed or non-available data
- Online cross-validation
- Concept-drift interdependencies (HMM?)
- Fine-tuning online heuristics

[1] Littlestone, N. (1988) Learning Quickly When Irrelevant Attributes Abound: A New Linear-Threshold Algorithm. *Machine Learning*. Vol. 2, pp. 285--318.

[2] Littlestone, N. and Warmuth, M.K. (1994) The Weighted Majority Algorithm. *Information and Computation*. Vol. 108, pp. 285--318.

[3] Blum, A. (1997) Empirical Support for Winnow and Weighted Majority Algorithms: Results on a Calendar Scheduling Domain. Machine Learning, 26.

[4] Oza, N. and Russel, S. (2001) Online Bagging and Boosting. In *Articial Intelligence and Statistics* 2001. Morgan Kaufmann, pp. 105--112.

[5] Kolter, J. Z., and Maloof, M. (2003) Dynamic weighted majority: A new ensemble method for tracking concept drift. In *Third IEEE international conference on data mining*. ICDM 2003, pp. 123--130. IEEE.

[6] Kolter, J. Z., and Maloof, M. (2005) Using additive expert ensembles to cope with concept drift. In *Procedings of the 22nd international conference on Machine Learning*. ICML 2005, pp. 449--456.

[7] Baena-García, M., del Campo Ávila, J., Fidalgo, R., Bifet, A., Gavalda, R., and Morales-Bueno, R. (2006) Early drift detection method. In *Fourth international workshop on knowledge discovery from data streams.* Vol. 6, pp. 7786.

[8] Bifet, A. and Gavaldà, R. (2007) Learning from time-changing data with adaptive windowing. In *SIAM Internationall Conference in Data Mining.* pp. 443--448.

[9] Bifet, A., Holmes, G., Pfahringer, B., Kirkby, R., and Gavaldà, R. (2009) New Ensemble Methods For Evolving Data Streams. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining.* ACM SIGKDD 2009, pp. 139--148.

[10] Minku, L. L. and Yao, X. (2012) Dynamic weighted majority: A new ensemble method for tracking concept drift. In *Third IEEE international conference on data mining*. ICDM 2003, pp. 123--130. IEEE.

[11] Ghazikhani, A., Monse, R. and SadoghiYazdi, H. (2013) Ensemble of online neural networks for non-stationary and imbalanced datastreams. *Neurocomputing*, 122, 535--544.

[12] Brzezinski, D. and Stefanowski, J. (2014) Combining block-based and online methods in learning ensembles from concept drifting data-streams. *Information Sciences*., 265, pp. 50--67

[13] Gomes, H.M., Bifet, A., Read, J. et al. (2017) Adaptive random forests for evolving data stream classication. *Machine Learning*, 106, pp. 1469--1495