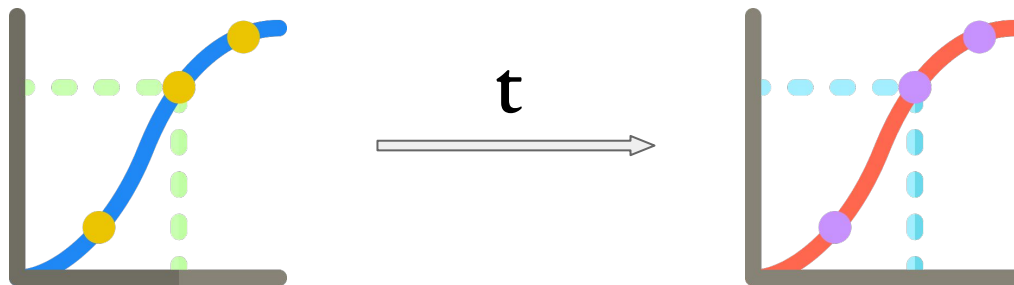


Dynamic weighted majority

Alejandro González

Concept drift



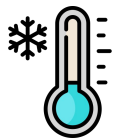
!Data changes over the time!

Concept drift

- Weather data

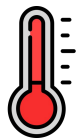
Summer

15 °C



Winter

15 °C



- Email relevance



➤ 2 artificial datasets: **STAGGER** and **SEA**

DWM algorithm

- Weighted ensemble classifier
- Online learners
 - Naive Bayes...
- β : Weight scaling factor
- θ : Weight threshold
- P

DWM algorithm

1. Train first learner
2. Predict input sample
3. Update weights
4. Normalize weights
5. Remove learners (if necessary)
6. If global prediction wrong:
Add new Learner
7. Train learners with input sample
8. Repeat from 2

STAGGER concepts

		Size		
		S	M	L
Green	T			
	C			
	R			
Blue	T			
	C			
	R			
Red	T			
	C			
	R			
Color	Shape			

Target
concept

$t = 1 \dots 40.$

		Size		
		S	M	L
Green	T			
	C			
	R			
Blue	T			
	C			
	R			
Red	T			
	C			
	R			
Color	Shape			

Target
concept

$t = 41 \dots 80.$

		Size		
		S	M	L
Green	T			
	C			
	R			
Blue	T			
	C			
	R			
Red	T			
	C			
	R			
Color	Shape			

Target
concept

$t = 81 \dots 120.$

Stagger dataset

- 3 attributes
- 2 labels (1 or 0)
- 3 concepts

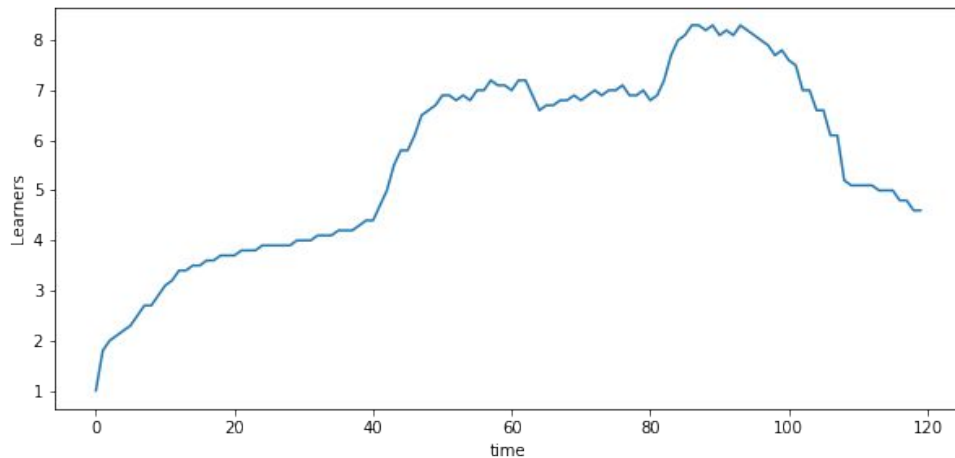
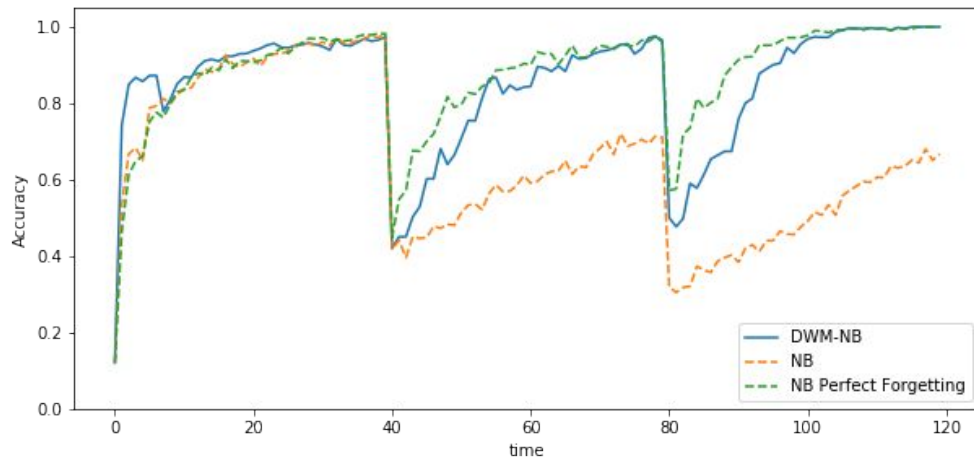
Processing

- One hot encode attributes
- 120 samples for streaming (40 samples per concept)
- Use of Bernoulli NB classifier

STAGGER evaluation

- DWM model
- NB Perfect forgetting
(Perfect case)
- Standard NB

! Averaging results 10 times !



SEA concepts

$X1 = [0,10]$ $X2 = [0,10]$ $X3 = [0,10]$

$X1 + X2 > 8$ \Rightarrow

$X1 + X2 > 9$ \Rightarrow

$X1 + X2 > 7.5$ \Rightarrow

$X1 + X2 > 9.5$ \Rightarrow



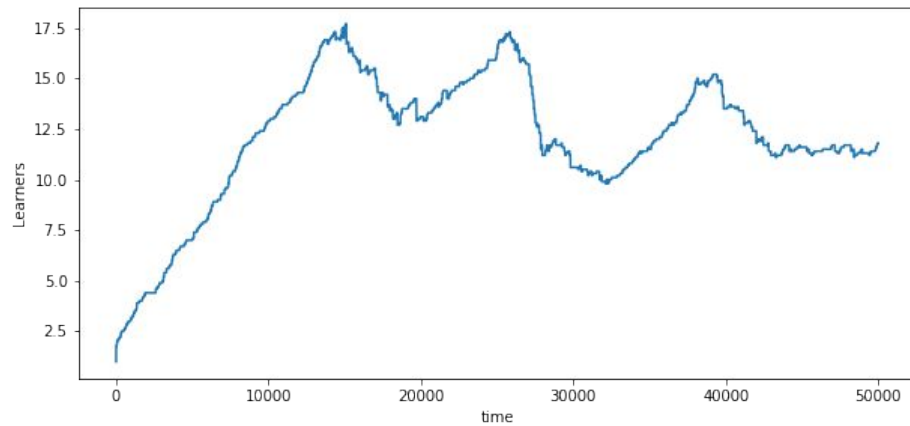
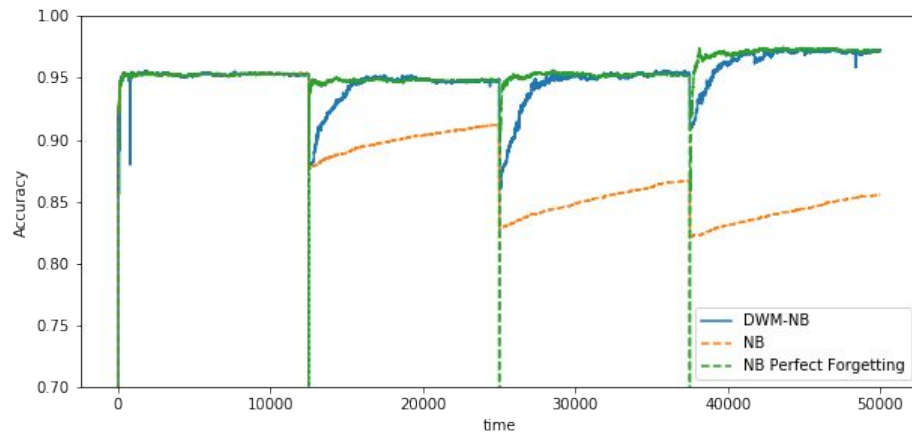
Stagger dataset

- 3 continuous attributes (1 irrelevant)
- 2 labels (1 or 0)
- 4 concepts

Processing

- One test-set for each concept
- 50000 samples for streaming (12500 samples per concept)
- Use of Gaussian NB classifier

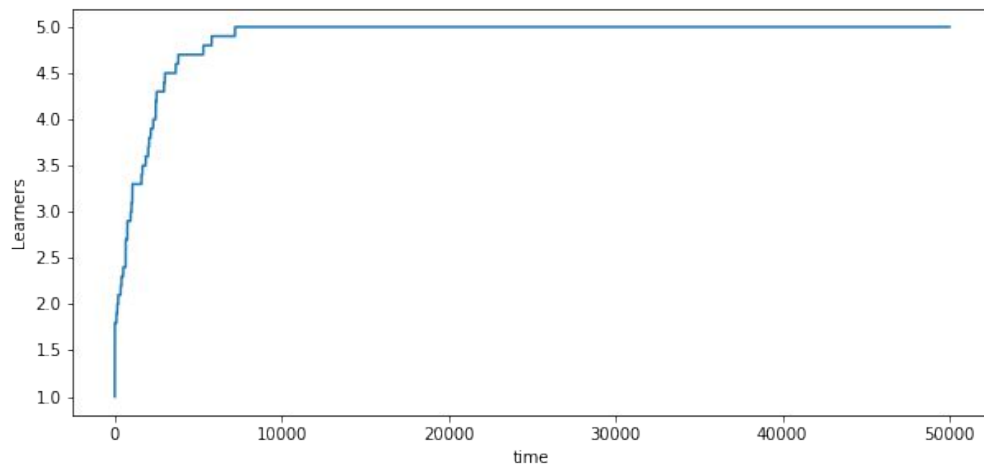
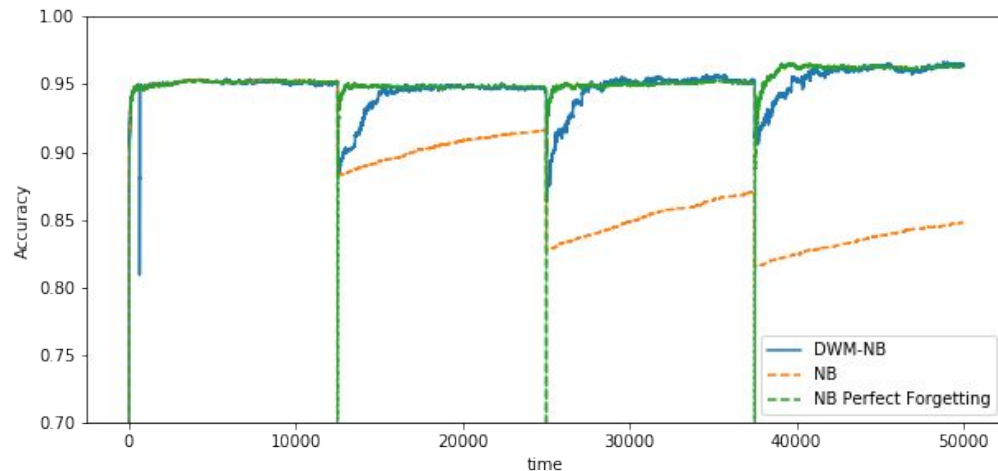
SEA evaluation



SEA evaluation

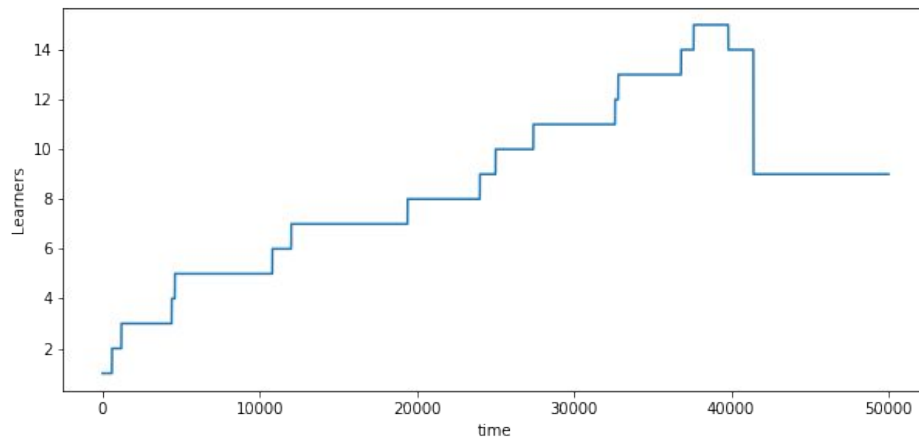
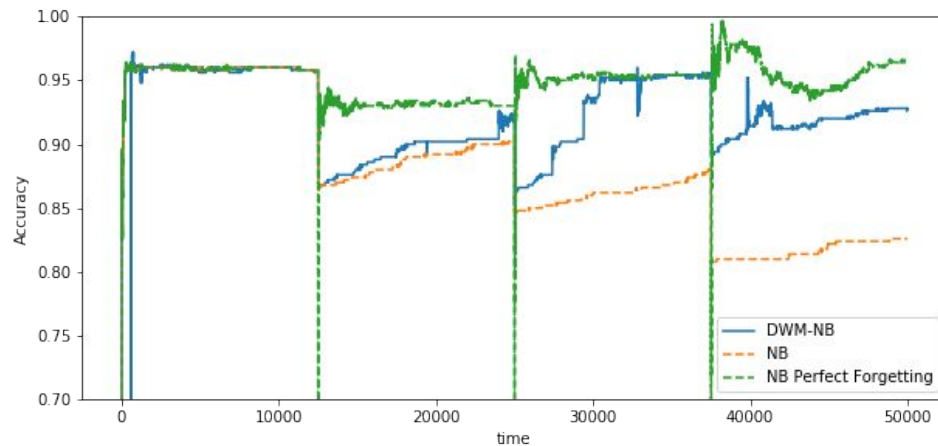
- Parameter k: Maximum number of learners
 - Faster computation
 - Similar performance

$K = 5$



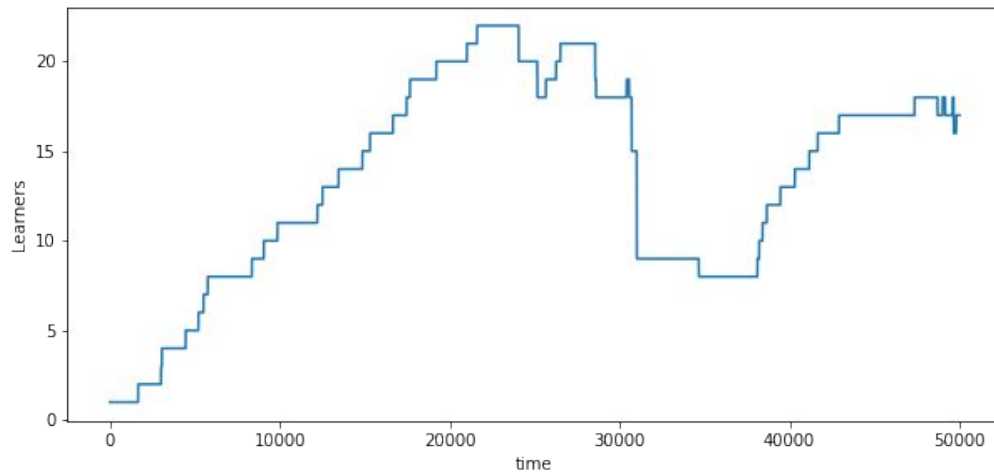
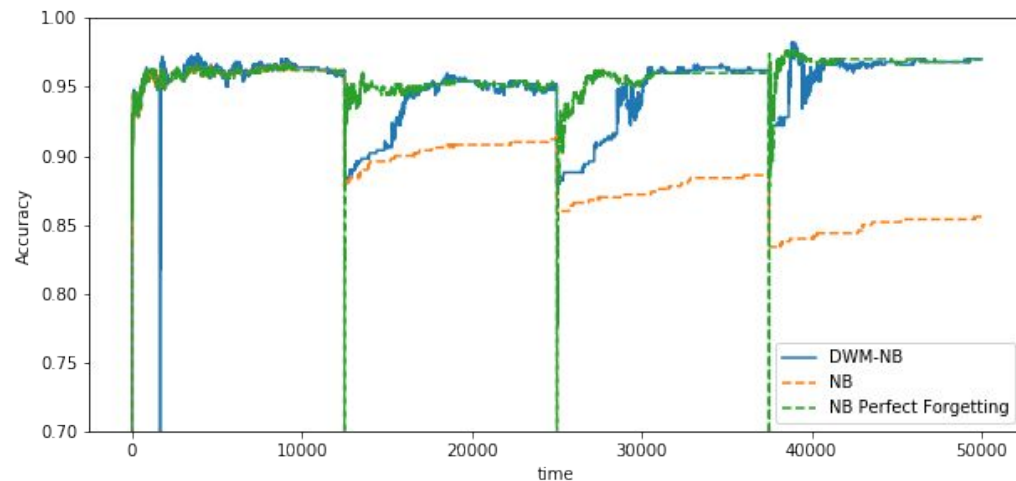
SEA evaluation

$P = 200$



SEA evaluation

$$\beta = 0.7$$



DEMO TIME!

Questions?