The Forgetron: A Kernel-Based Perceptron on a Fixed Budget

Ofer Dekel **Shai Shalev-Shwartz** **Yoram Singer**
School of Computer Science & Engineering
The Hebrew University, Jerusalem 91904, Israel
{oferd,shais,singer}@cs.huji.ac.il

- Presented by : Mourad GUEZZOU

Contents

Introduction

- The online classifications algorithms store a subset of observed example in its internal memory
- It continually changes as learning progresses *(new hypothesis are added)*
- A rapid growth of active set + Bounded memory ⟹ Risk to require more memory than physically available
- problem specially eminent in cases where the online Algorithm is implemented in hardware with small memory such as mobile telephone

- FORGETRON: since its update builds on that of the percepTRON and since it gradually FORGETs active example as learning progresses

Problem Setting

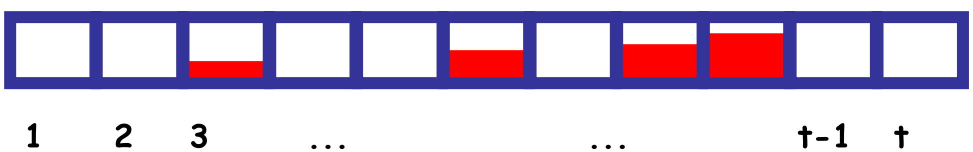
- **Online learning:**
  - Choose an initial hypothesis f_0
  - For t=1,2,…
    - Receive an instance x_t and predict sign($f_t(x_t)$) *determined by a hypothesis, stored in internal memory and updated from round to round*
    - If ($y_t f_t(x_t)$ <= 0) *(f_t denote the hypothesis used in round t)*
      - update the hypothesis f
- **Goal**: minimize the number of prediction mistakes
- **Kernel**-based hypotheses $$f_t(\mathbf{x}) = \sum_{i \in I_t} \sigma_{i,t} y_i K(\mathbf{x}_i, \mathbf{x})$$
- Example: the **dual Perceptron**
  - $\sigma_{i,t}$ is always 1
  - Initial hypothesis: $I_1 = \emptyset$
  - Update rule: $I_{t+1} = I_t \cup \{t\}$

K	kernel Operator
I	subset of {1,…,(t-1)}
X_i	is active on round t if i in I_t
Y_i	in {-1, +1}
B	positif integer, refer budget parameter

The Forgetron

- Initilize: $I_1 = \emptyset\,;\ Q_1 = 0\,;\ M_0 = 0$ *with M number of mistakes*
- For t=1,2…

define $f_t(\mathbf{x}) = \sum_{i \in I_t} \sigma_{i,t} y_i K(\mathbf{x}_i, \mathbf{x})$



Receive an instance x_t, predict sign($f_t(x_t)$), and then receive y_t

If $y_t\ f_t(x_t) <= 0$ set $M_t = M_{t-1} + 1$ and update

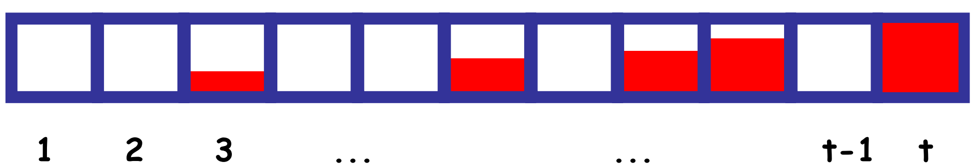
The Forgetron

Step (1) - Perceptron

B positif integer, refer budget parameter

$$I'_t = I_t \cup \{t\}$$

define $$f'_t = f_t + y_t K(\mathbf{x}_t, \cdot)$$



- If $|I'_t| <= B$ skip the next two steps
- define $r_t = \min I_t$

The Forgetron

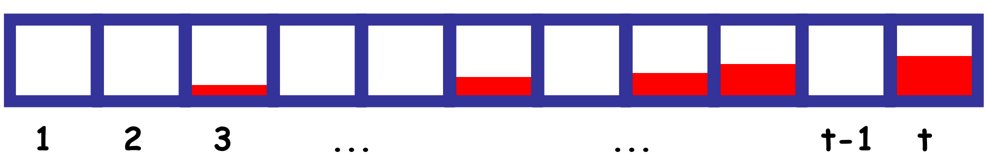
Step (2) - Shrinking

Shrinking coefficient to manage the damage caused by the removal step

$$\phi_t = \max\{\phi \in (0,1] : \Psi(\phi, \sigma_{r_t,t}, \mu_t) + Q_t \;\leq\; (15/32)\, M_t\}$$

$$\forall i \in I'_t,\;\; \sigma_{i,t+1} = \phi_t\, \sigma_{i,t}$$

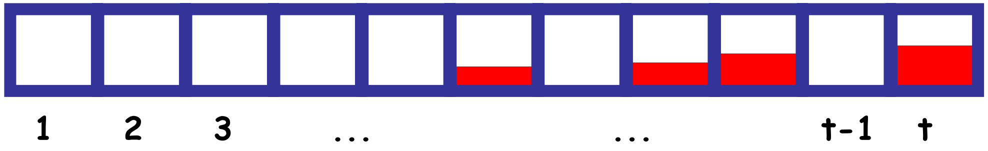
define $\quad f''_t = \phi_t f'_t$



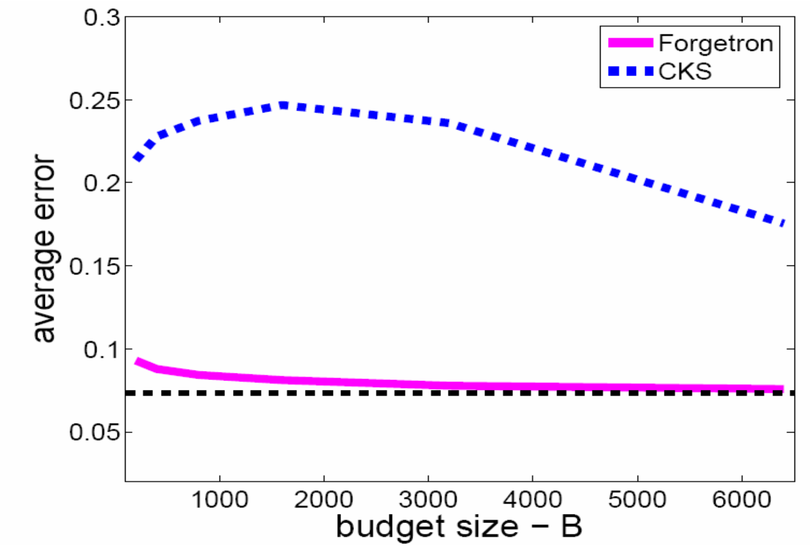
The Forgetron

Step (3) - Removal

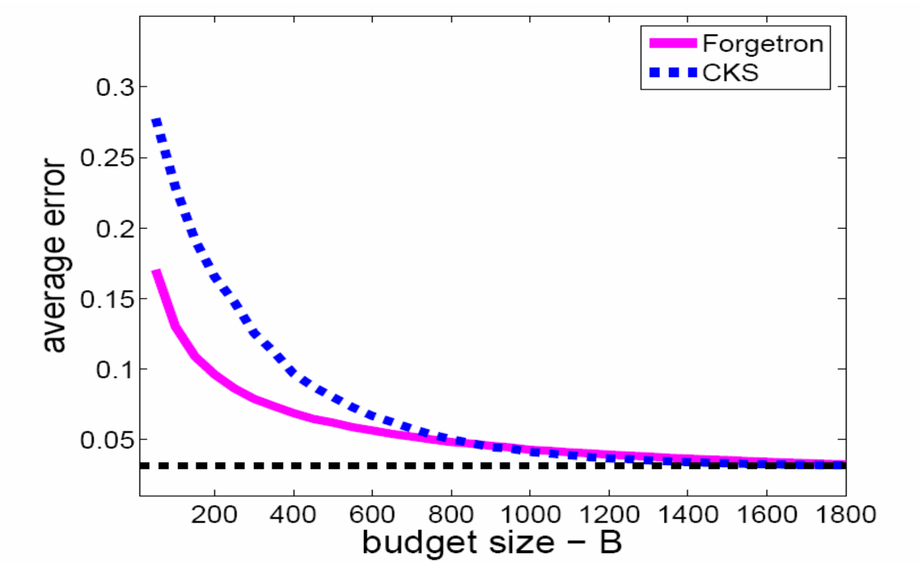
$$I_{t+1} = I_t \setminus \{r_t\}$$



Experiments



Census-income Dataset



MNIST Dataset

Note that the Forgetron outperforms CKS on both datasets, especially when the value of B is small.

Conclusion

- Describe the FORGETRON algorithm which is kernel-based online learning with a fixed memory budget
- The analysis presented in this paper can be used to derive a family of online algorithms of which the Forgetron is only one special case.