A study on majority-voting classifiers with guarantees on the probability of error

Algo Carè (speaker)
Marco C. Campi
Federico A. Ramponi
Università degli Studi di Brescia (IT)

Simone Garatti
Politecnico di Milano (IT)

A.T.J. Roy Cobbenhagen
University of Technology, Eindhoven (NL)

Classification

$\mathrm{x} \in \mathbb{R}^n$: object ($n$ features)

$y \ \in \{1, 0\}$: true class

$\hat{y}(\cdot) : \mathbb{R}^n \to \{1, 0\}$ classifier

Error

$$\hat{y}(x) \neq y$$

Probability of Error

$$(x, y) \sim \mathbb{P}$$

Unknown!

$$PE(\hat{y}) := \mathbb{P}\{\hat{y}(x) \neq y\}$$

$\mathrm{S}=\{(\mathrm{x}_1, y_1), \ldots, (x_N, y_N)\}$ i.i.d.

check: $\hat{y}(x_i) \neq y_i$

conclude: $PE(\hat{y}) \leq \hat{\epsilon}(S)$

unless $S \in$ "Unlucky set", where $\mathbb{P}^N(\text{"Unlucky set"}) < \beta$

Typical workflow

$$\mathrm{T}=\{(\mathrm{x}_1, y_1), \ldots, (x_N, y_N)\}$$

$\hat{y}(\cdot)$ trained on T

$$\mathrm{S}=\{(\mathrm{x}_{N+1}, y_{N+1}), \ldots, (x_{N+K}, y_{N+K})\} \text{ i.i.d.}$$

check: $\hat{y}(x_{N+i}) \neq y_{N+i}$

conclude: $PE(\hat{y}) \leq \hat{\epsilon}(S)$

unless $S \in$ "Unlucky set", where $\mathbb{P}^K(\text{“Unlucky set”}) < \beta$

Self-testing classifiers

$$\mathrm{T}=\{(\mathrm{x}_1, y_1), \ldots, (x_N, y_N)\} \text{ i.i.d.}$$

$\hat{y}(\cdot)$ trained on T

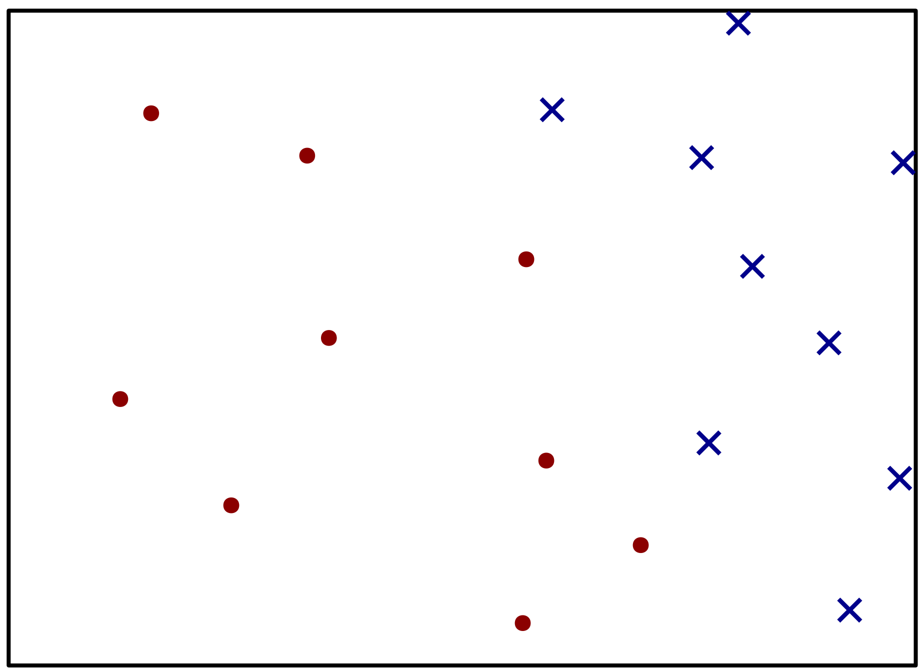
certificate: $PE(\hat{y}) \leq \hat{\epsilon}(T)$

unless $T \in$ "Unlucky set", where $\mathbb{P}^N($"Unlucky set"$) < \beta$

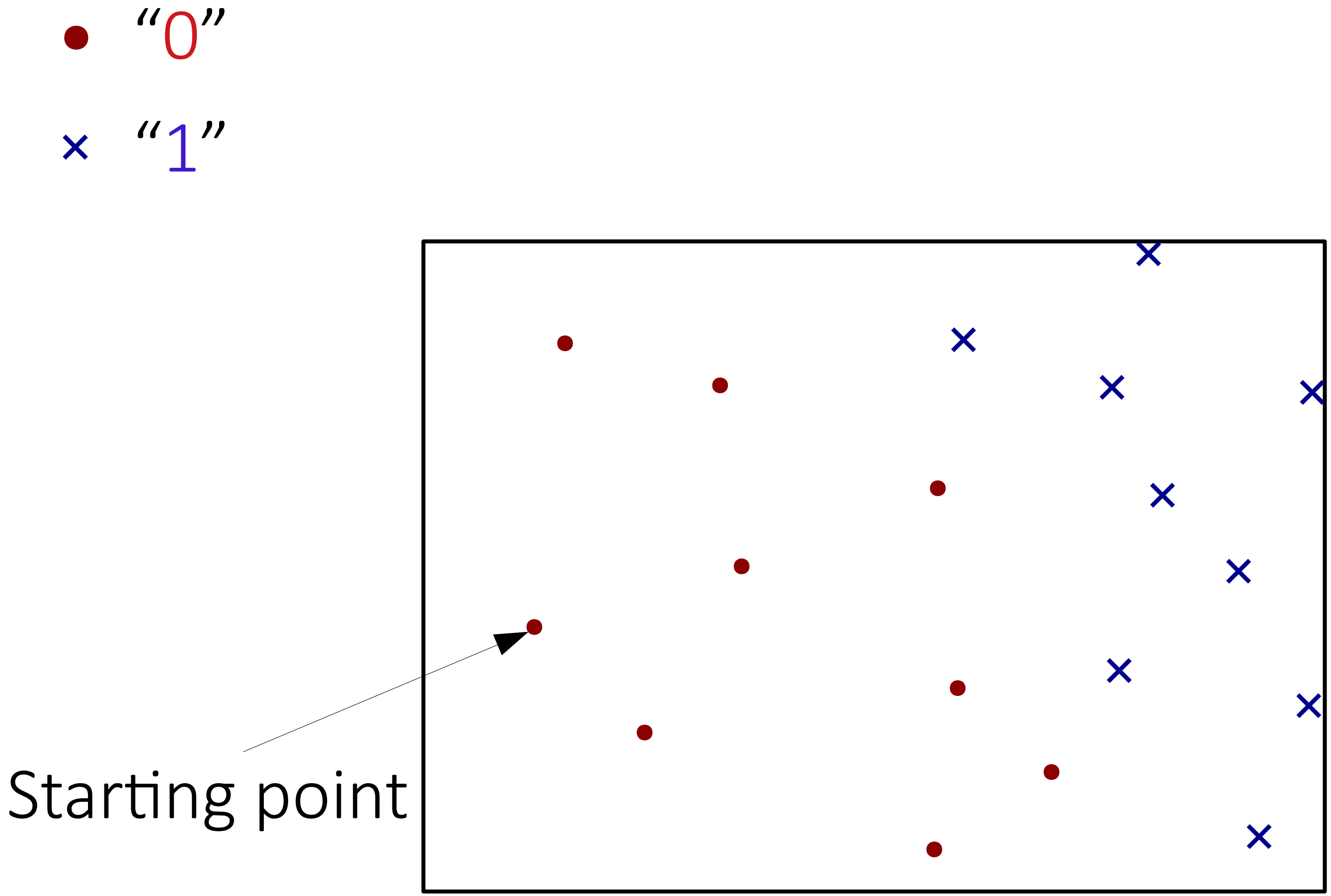
For example…

GEM-BALLS

- ● "0"
- × "1"

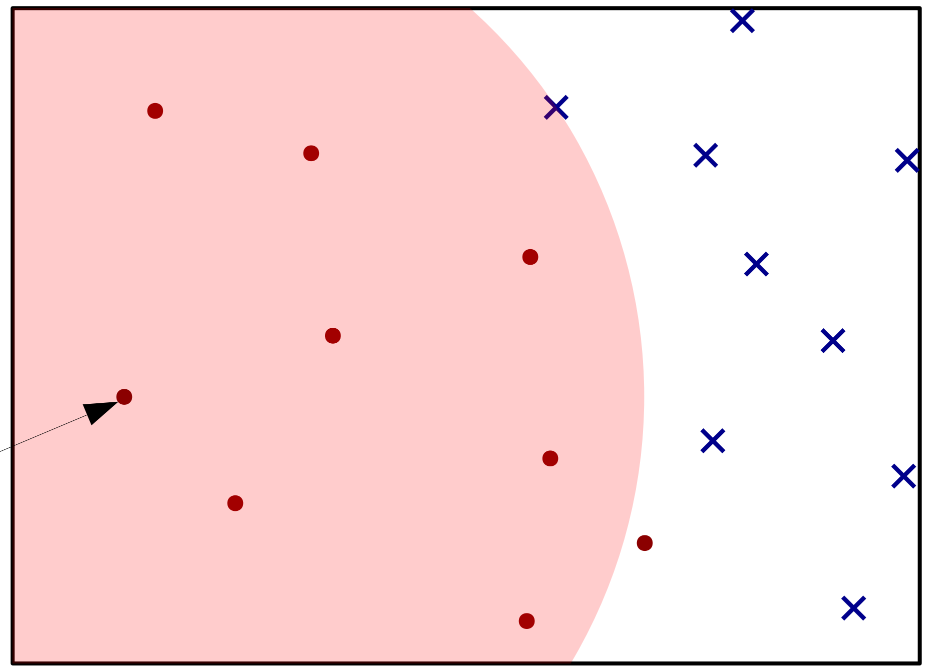


GEM-BALLS



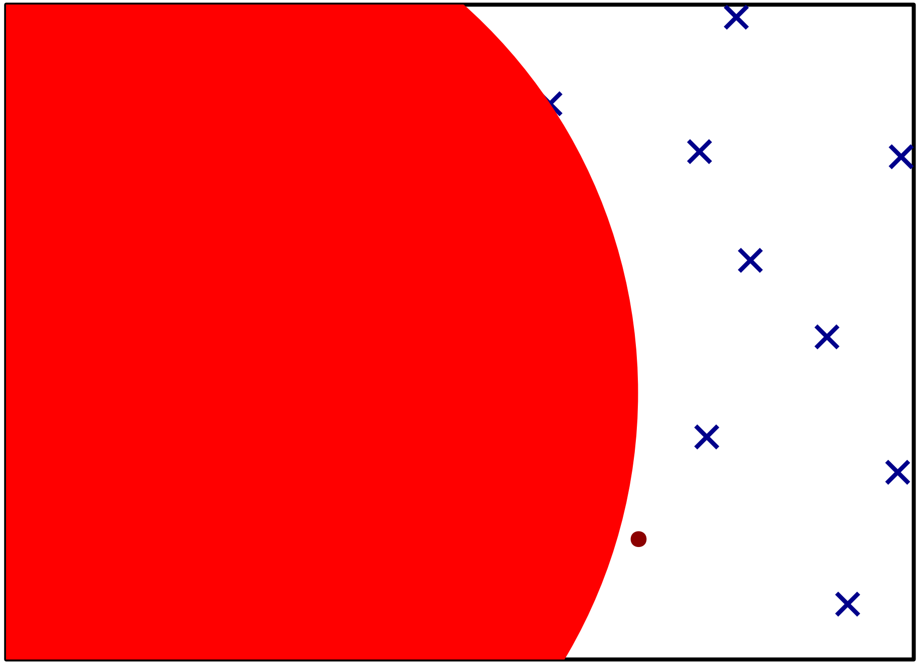
GEM-BALLS

Largest ball that does not include blue points



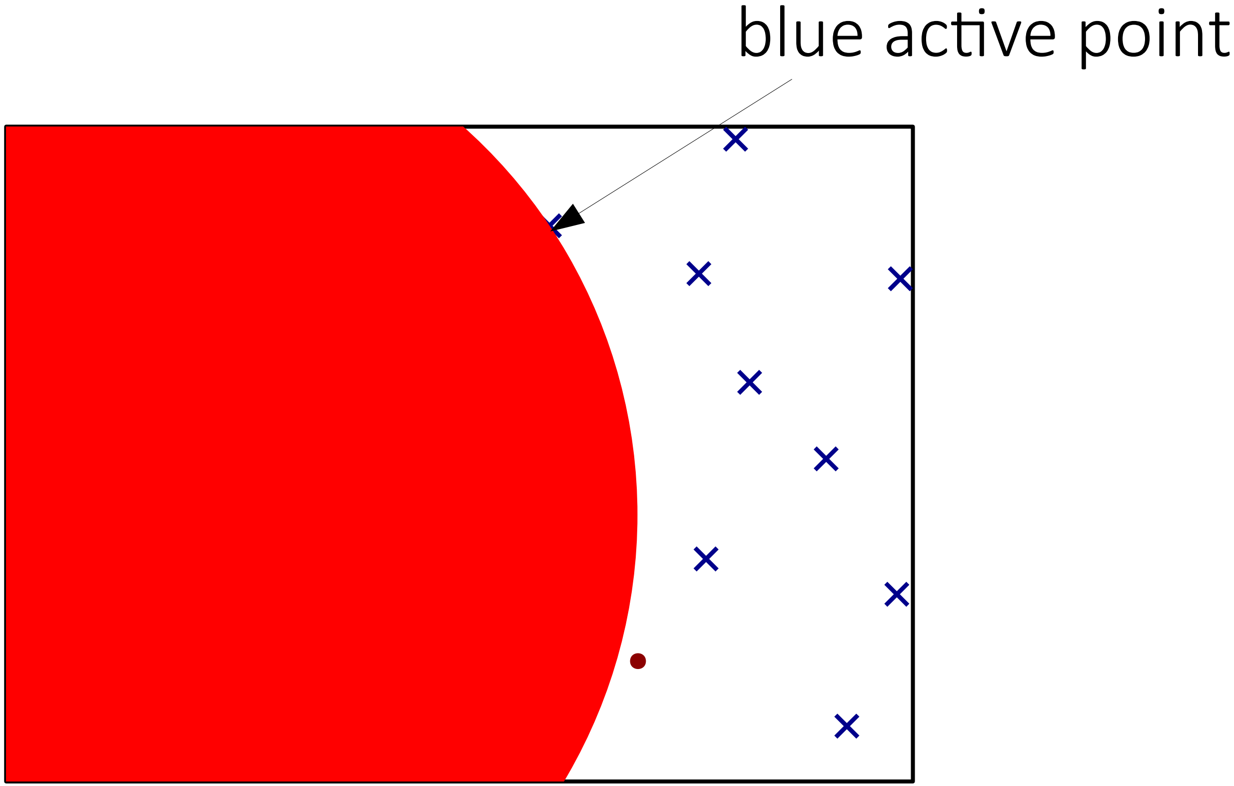
Starting point

GEM-BALLS



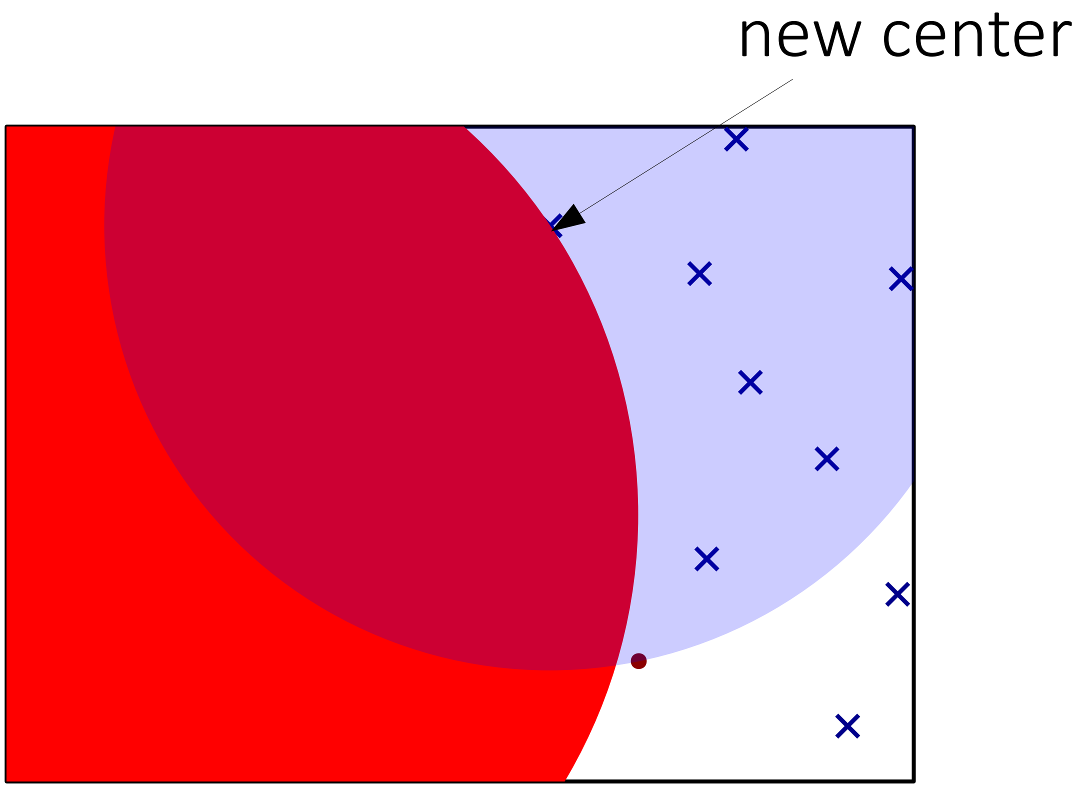
GEM-BALLS

#blue active points so far =1



GEM-BALLS

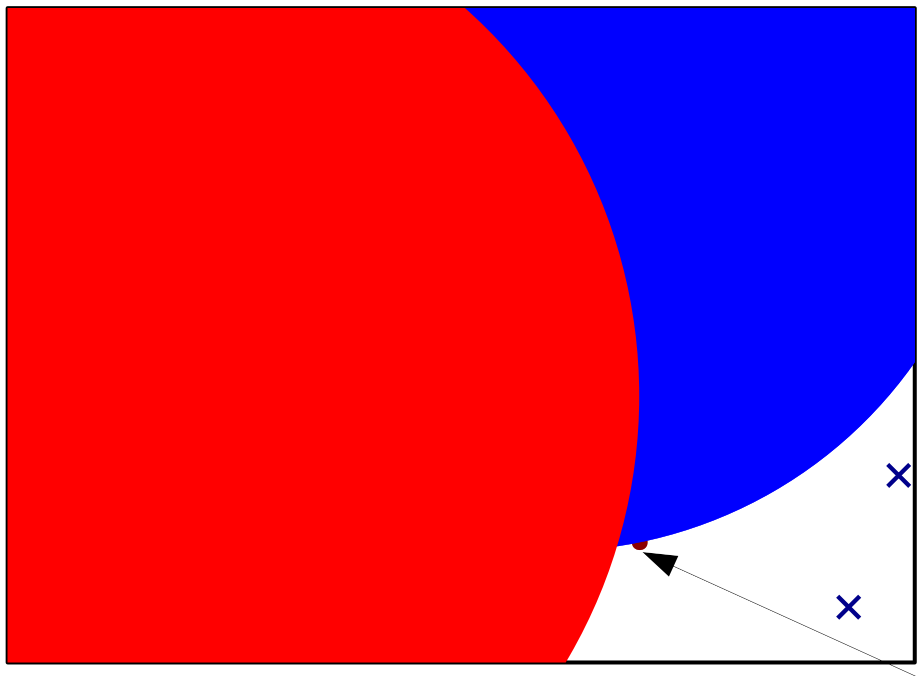
#blue active points so far =1



GEM-BALLS

#blue active points so far =1

#red active points so far =1

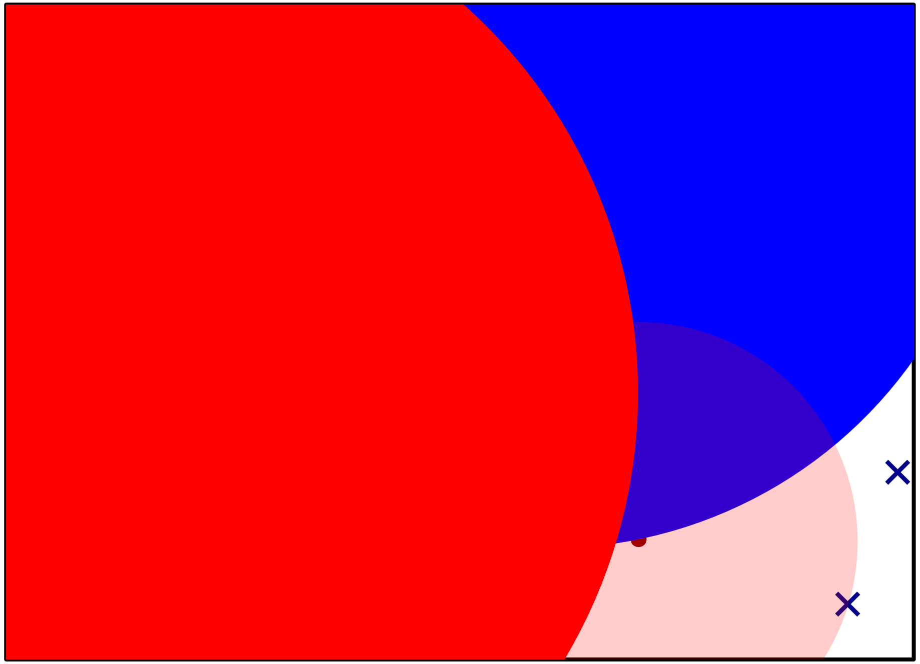


red active point

GEM-BALLS

#blue active points so far =1

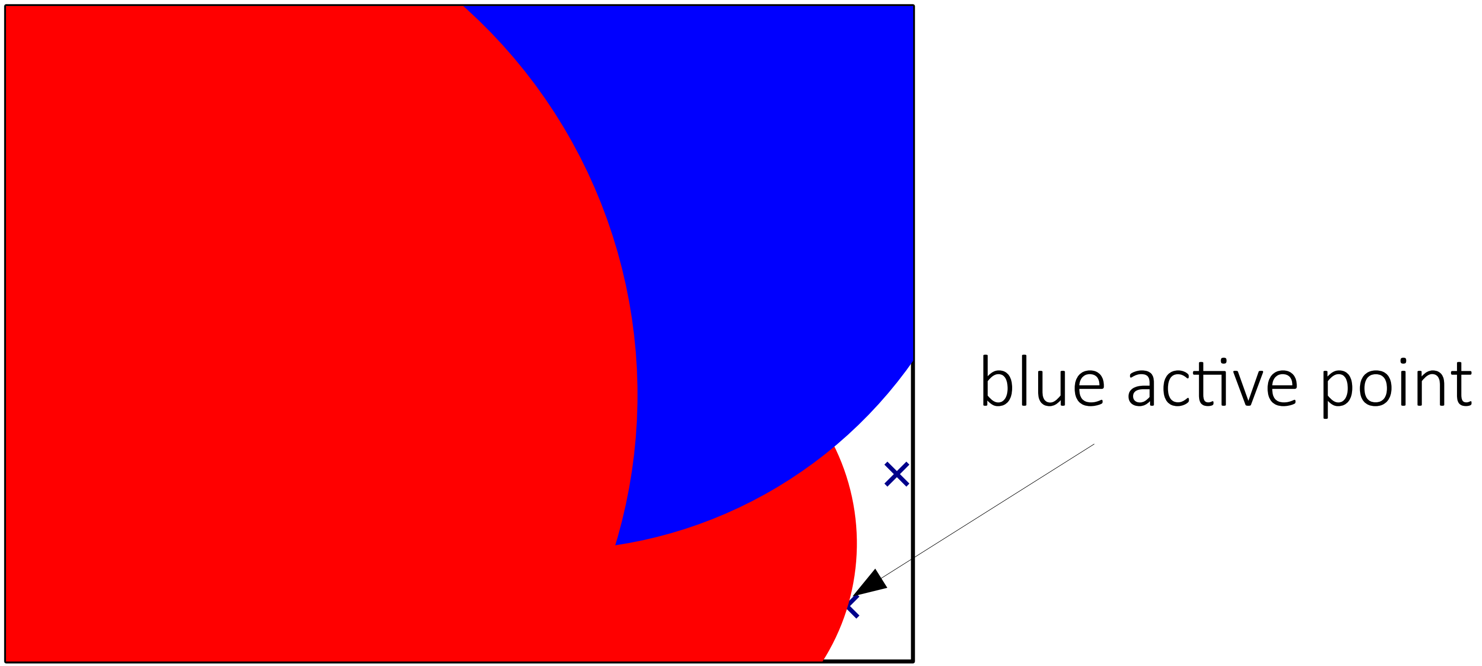
#red active points so far =1



GEM-BALLS

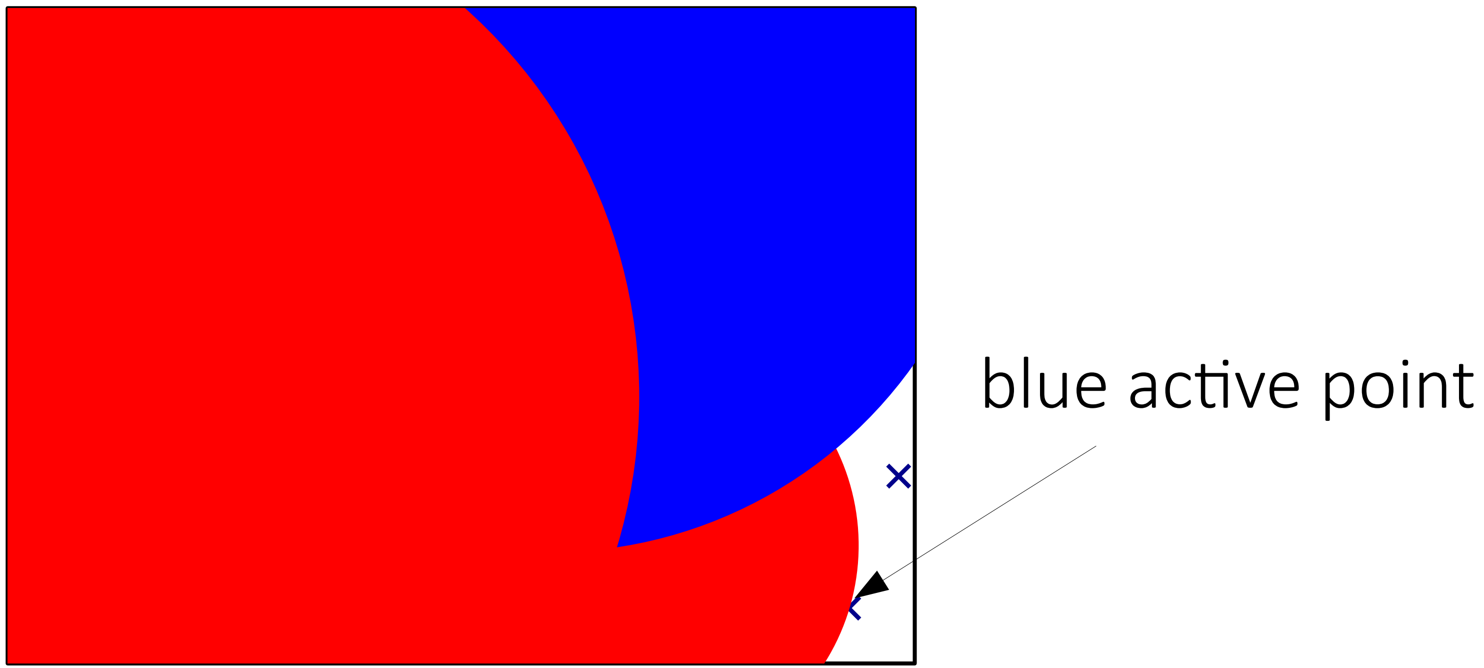
#blue active points so far =1+1

#red active points so far =1



GEM-BALLS

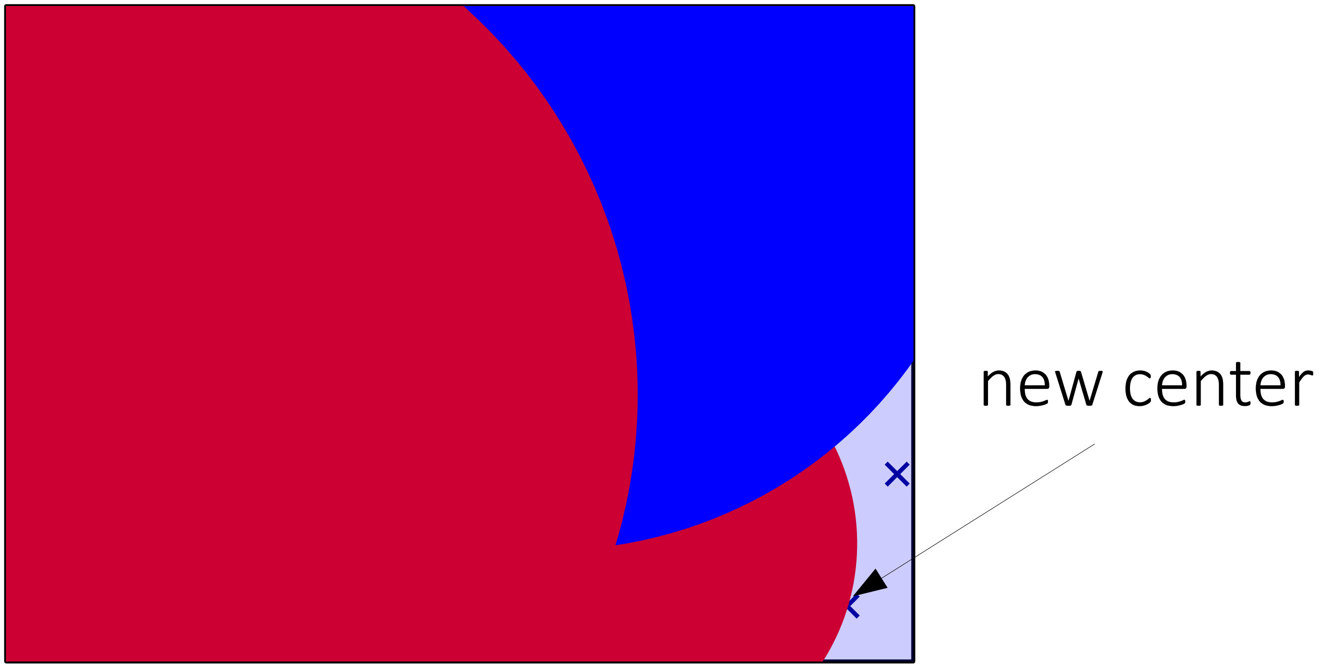
#blue active points so far =2
#red active points so far =1



GEM-BALLS

#blue active points so far =2

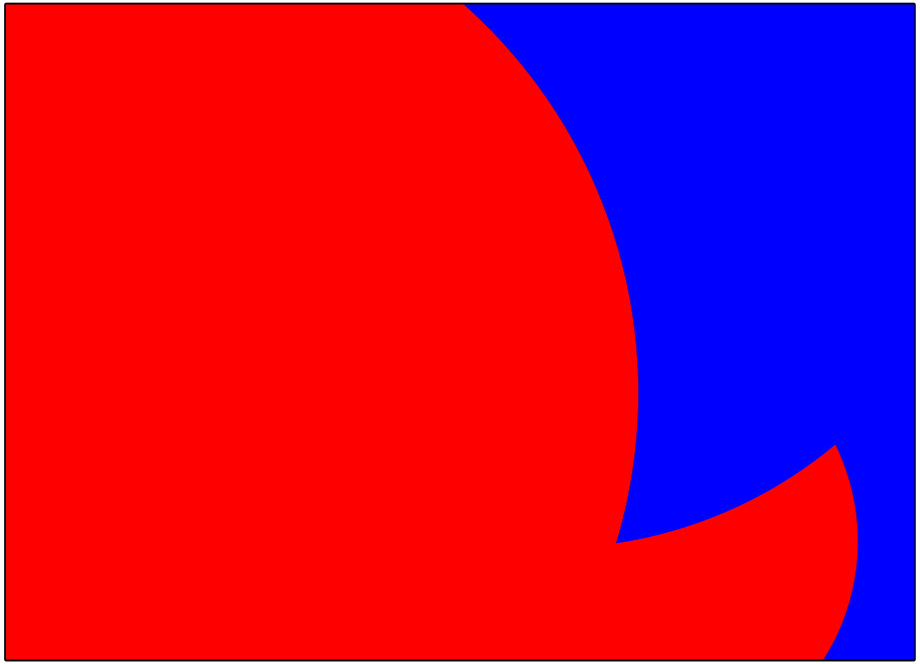
#red active points so far =1



GEM-BALLS

#blue active points so far =2

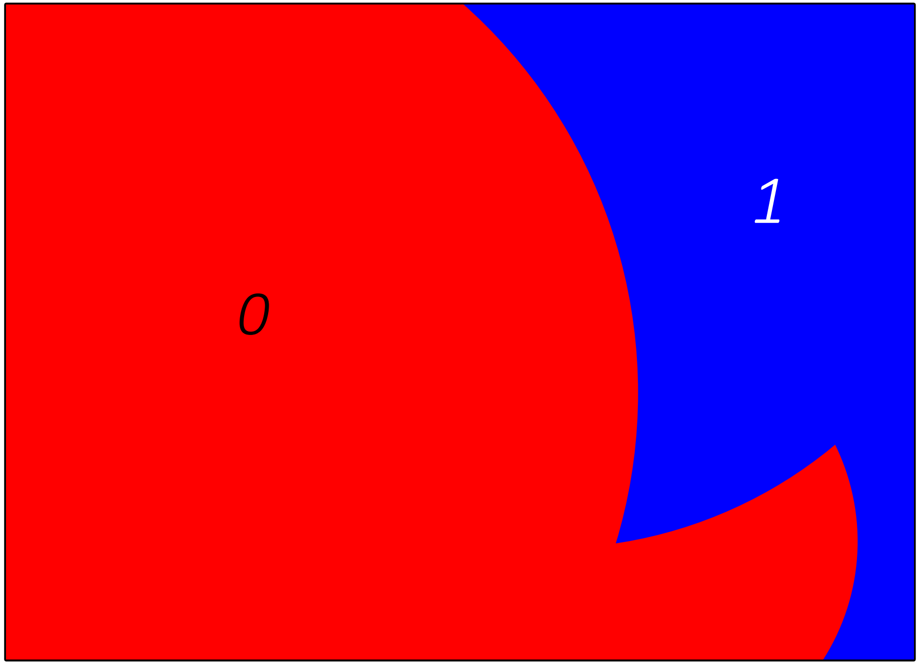
#red active points so far =1



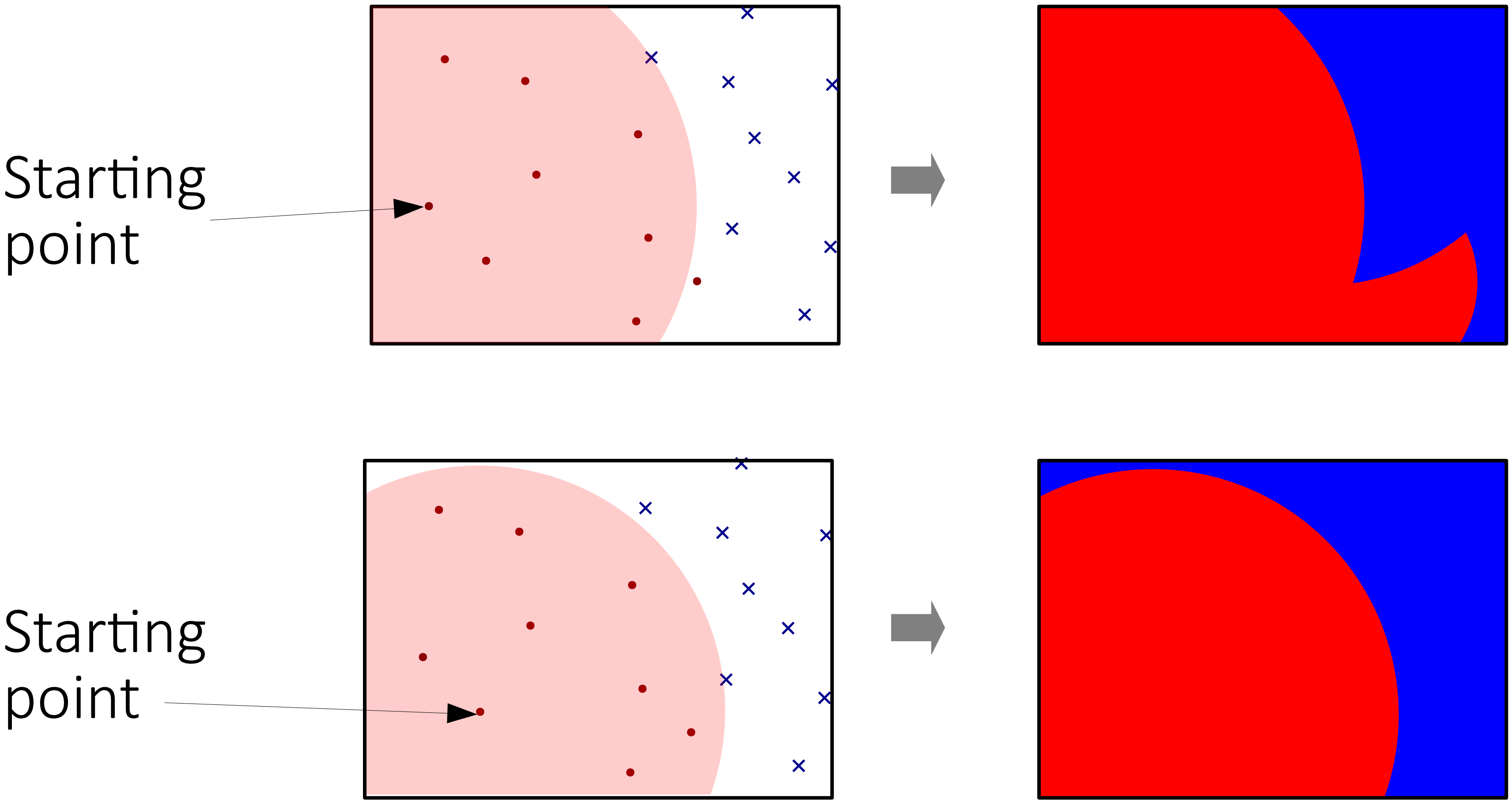
GEM-BALLS

#blue active points so far =2

#red active points so far =1



Motivating example



The set-up of this paper

Classifier:	$\hat{y}_1(\cdot)$	$\hat{y}_2(\cdot)$	$\hat{y}_3(\cdot)$	$\cdots$	$\hat{y}_M(\cdot)$
$PE \leq$	$\hat{\epsilon}_1$	$\hat{\epsilon}_2$	$\hat{\epsilon}_3$	$\cdots$	$\hat{\epsilon}_M$
P.o.F. :	β	β	β	$\cdots$	β

$$\hat{y}^*(x) = \begin{cases} 0 & \text{if } \frac{1}{M}\sum_{i=1}^{M} \hat{y}_i(x) < 0.5 \\ 1 & \text{otherwise} \end{cases}$$

$$PE(\hat{y}^*) \leq ?$$

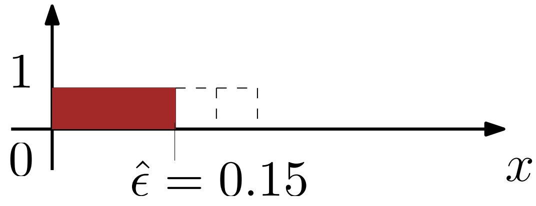
Is the majority right?

Classifier:	$\hat{y}_1(\cdot)$	$\hat{y}_2(\cdot)$	$\hat{y}_3(\cdot)$	$\cdots$	$\hat{y}_M(\cdot)$
$PE \leq$	$\hat{\epsilon}$	$\hat{\epsilon}$	$\hat{\epsilon}$	$\ldots$	$\hat{\epsilon}$

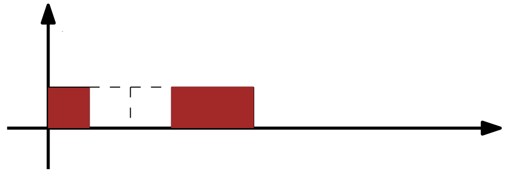
$$PE(\hat{y}^*) \leq \hat{\epsilon} \quad ?$$

Counter-example

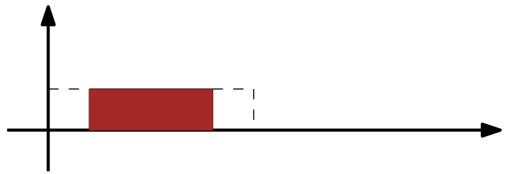
$Err(x, \hat{y}_1(\cdot))$



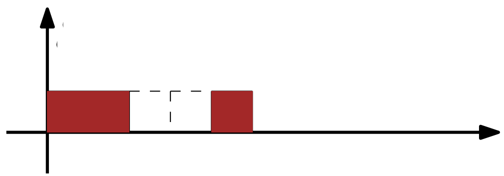
$Err(x, \hat{y}_2(\cdot))$



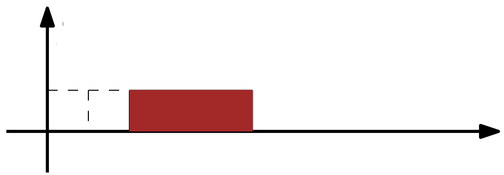
$Err(x, \hat{y}_3(\cdot))$



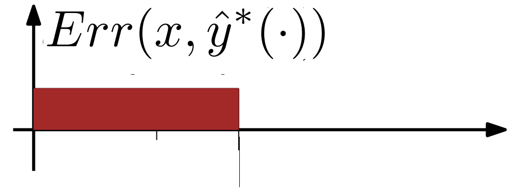
$Err(x, \hat{y}_4(\cdot))$



$Err(x, \hat{y}_5(\cdot))$



$Err(x, \hat{y}^*(\cdot))$



$$PE(\hat{y}^*) = \frac{5}{3} 0.15 = 0.25$$

$$\frac{2M}{M+1}\hat{\epsilon}$$

Is the majority right?

Classifier: $\hat{y}_1(\cdot)$ $\hat{y}_2(\cdot)$ $\hat{y}_3(\cdot)$ $\cdots$ $\hat{y}_M(\cdot)$

$PE \leq$ $\hat{\epsilon}$ $\hat{\epsilon}$ $\hat{\epsilon}$ $\ldots$ $\hat{\epsilon}$

$$PE(\hat{y}^*) \leq 2\hat{\epsilon}$$

Tight!

FACT: It is possible that the performance of a majority of equally skilled experts is worse than the performance of any individual expert!

Bound n.1

Classifier:	$\hat{y}_1(\cdot)$	$\hat{y}_2(\cdot)$	$\hat{y}_3(\cdot)$	$\cdots$	$\hat{y}_M(\cdot)$
$PE \leq$	$\hat{\epsilon}$	$\hat{\epsilon}$	$\hat{\epsilon}$	$\ldots$	$\hat{\epsilon}$

$$PE(\hat{y}^*) \leq \frac{1}{A}\hat{\epsilon}$$

$$A := \min_x \{\text{Agreement}(\mathrm{x})\}$$

An example

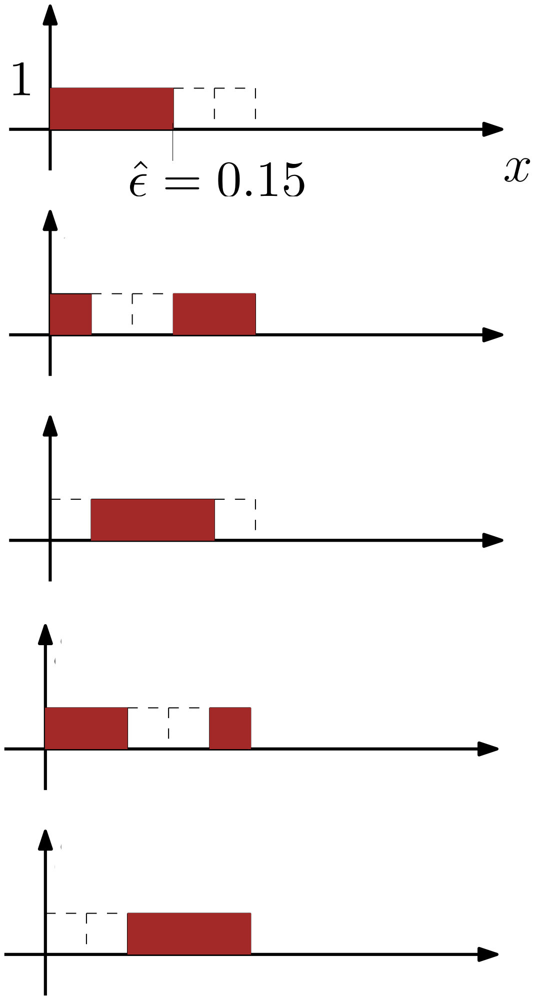
$Err(x, \hat{y}_1(\cdot))$

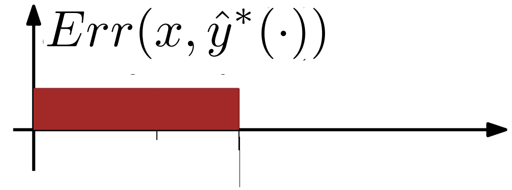
$Err(x, \hat{y}_2(\cdot))$

$Err(x, \hat{y}_3(\cdot))$

$Err(x, \hat{y}_4(\cdot))$

$Err(x, \hat{y}_5(\cdot))$





$$PE(\hat{y}^*) = \frac{5}{3} 0.15 = 0.25$$

$$A := \min_x \{\text{Agreement}(x)\} = \frac{3}{5}$$

Bound n.1

Classifier:	$\hat{y}_1(\cdot)$	$\hat{y}_2(\cdot)$	$\hat{y}_3(\cdot)$	$\cdots$	$\hat{y}_M(\cdot)$
$PE \leq$	$\hat{\epsilon}_1$	$\hat{\epsilon}_2$	$\hat{\epsilon}_3$	$\cdots$	$\hat{\epsilon}_M$

$$PE(\hat{y}^*) \leq \frac{1}{A}\left(\frac{1}{M}\sum_{i=1}^{M}\hat{\epsilon}_i\right)$$

$$A := \min_x\{\mathrm{Agreement(x)}\}$$

Let's go back to our set-up

Classifier:	$\hat{y}_1(\cdot)$	$\hat{y}_2(\cdot)$	$\hat{y}_3(\cdot)$	$\cdots$	$\hat{y}_M(\cdot)$
$PE \leq$	$\hat{\epsilon}_1(T)$	$\hat{\epsilon}_2(T)$	$\hat{\epsilon}_3(T)$	$\cdots$	$\hat{\epsilon}_M(T)$
P.o.F. :	β	β	β	$\cdots$	β

$$\mathbb{P}^N(\text{“Unlucky set”}) < M\beta$$

Let’s go back to our set-up

$$PE(\hat{y}^*) \leq \frac{1}{A}\left(\frac{1}{M}\sum_{i=1}^{M}\hat{\epsilon}_i\right)$$

unless $T \in$ “Unlucky set”, where $\mathbb{P}^N($“Unlucky set”$) < M\beta$

OK if $M\beta$ is small

Bound n.2

$$PE(\hat{y}^*) \leq \frac{1}{A}\left(\frac{1}{M}\sum_{i=1}^{M}\hat{\epsilon}_i\right) + \varrho$$

unless $T \in$ "Unlucky set", where $\mathbb{P}^N(\text{"Unlucky set"}) < \frac{1}{\varrho}\beta$

Bound n.2 for weighted majority

$$PE(\hat{y}^*) \leq \frac{1}{A}\left(\sum_{i=1}^{M} \hat{w}_i(T)\hat{\epsilon}_i\right) + \varrho$$

unless $T \in$ "Unlucky set", where $\mathbb{P}^N(\text{"Unlucky set"}) < \frac{k}{\varrho}\beta$

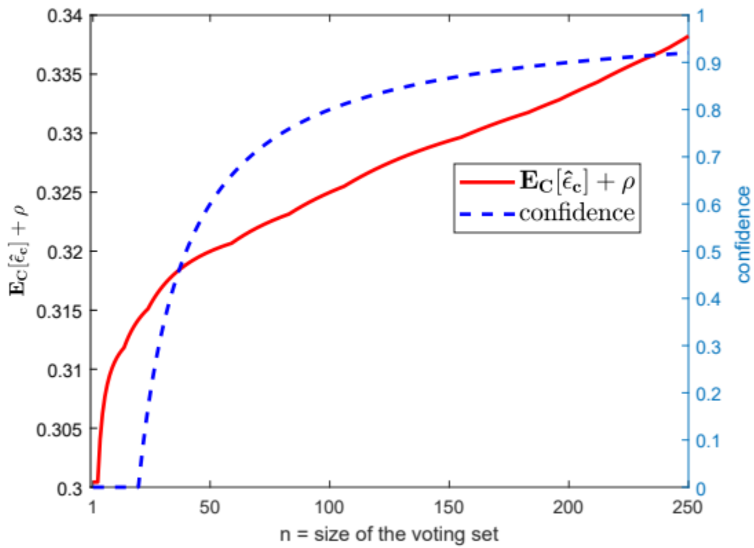
$k = 1 \leftarrow \hat{w}_i(T)$ are democratic $\hat{w}_i(T) = 1/M$

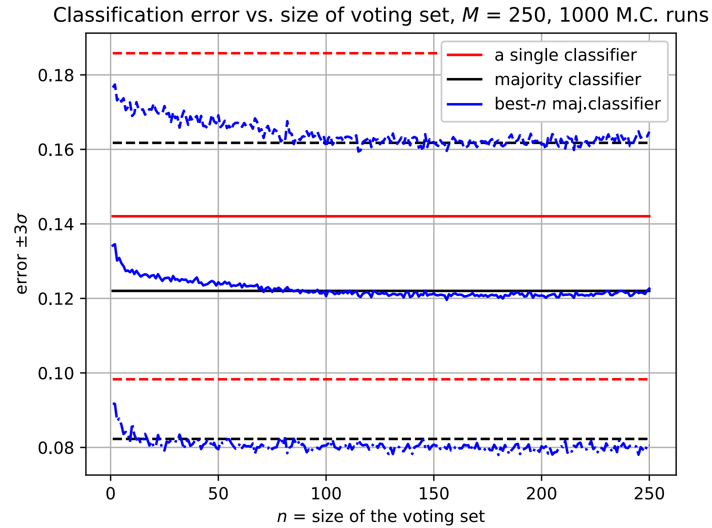
$k = 2 \leftarrow \hat{w}_i(T)$ can be zero for half of the classifiers

$\vdots$

$k = M \leftarrow \hat{w}_i(T)$ are unconstrained

Error-confidence trade-off





Take-home messages

- ➢There exist self-testing classifiers

- ➢They can be used as base classifiers in majority-voting classification schemes

- ➢Although empirical studies suggest that majority voting is good, the performance of the majority **can** in principle be worse than any individual performance.
  We made steps towards:
  → protecting against bad situations
  and even
  → detecting favourable situations

- ➢The proliferation of base classifiers is not a real issue
- ➢Data-dependent weighted majority voting schemes are possible (with care)

Ongoing and future research

- Improve the bounds for favourable situations
  ("$1/A$" is just the starting point)

- Extensions to the regression framework

CODE AND MORE INFORMATION: http://www.algocare.it/GEM-BALLS/

Thank you!

algo.care@unibs.it