

Un nuovo algoritmo per la costruzione di classificatori
da dati sperimentali con errore di generalizzazione garantito

Seminario per il corso di Teoria e Tecniche di Identificazione e Stima (18 Giugno 2009)

Per contatti, approfondimenti, et cetera...

Algo Carè

Trasparenze: <http://www.ing.unibs.it/~algo.care/>

Mail: algo.care@ing.unibs.it

Telefono @DEA: 5431

Prof. Marco C. Campi

Università degli Studi di Brescia
Facoltà di Ingegneria
Corso di Laurea in Ingegneria Informatica
Dipartimento di Elettronica per l'Automazione



Un nuovo algoritmo per la costruzione di classificatori
da dati sperimentali con errore di generalizzazione
garantito

Relatore:
Prof. **Marco C. Campi**

Tesi di laurea specialistica di:
Algo Carè

Schema della presentazione



“Un nuovo **algoritmo** per la costruzione di **classificatori** da dati sperimentali con **errore di generalizzazione *garantito***”

A. La classificazione (binaria)

B. Il nuovo algoritmo: GPE
(*Guaranteed Probability of Error*)

C. Proprietà teoriche e risultati
sperimentali

Prima Parte



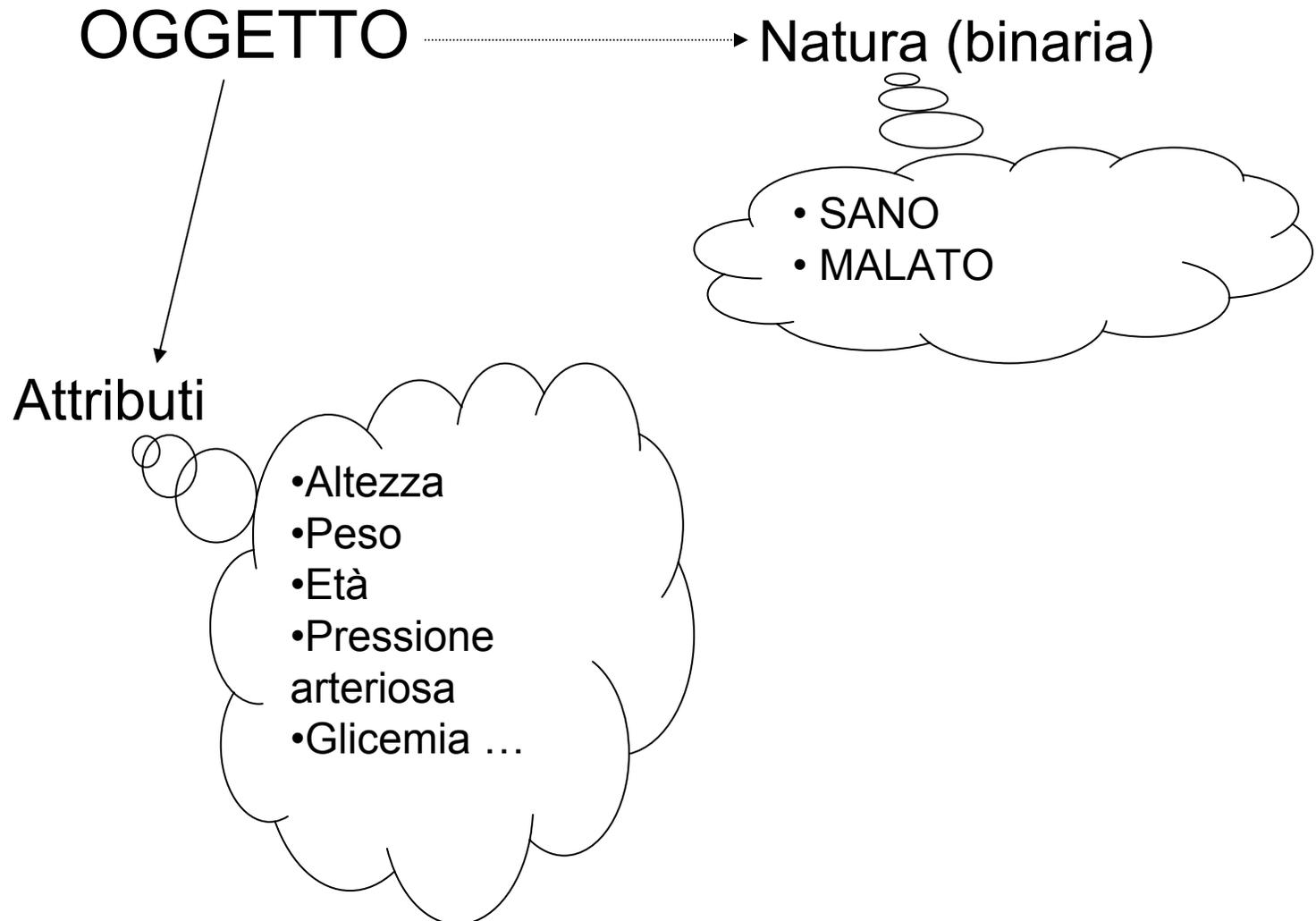
“Un nuovo **algoritmo** per la costruzione di **classificatori** da dati sperimentali con **errore di generalizzazione *garantito***”

A. La classificazione (binaria)

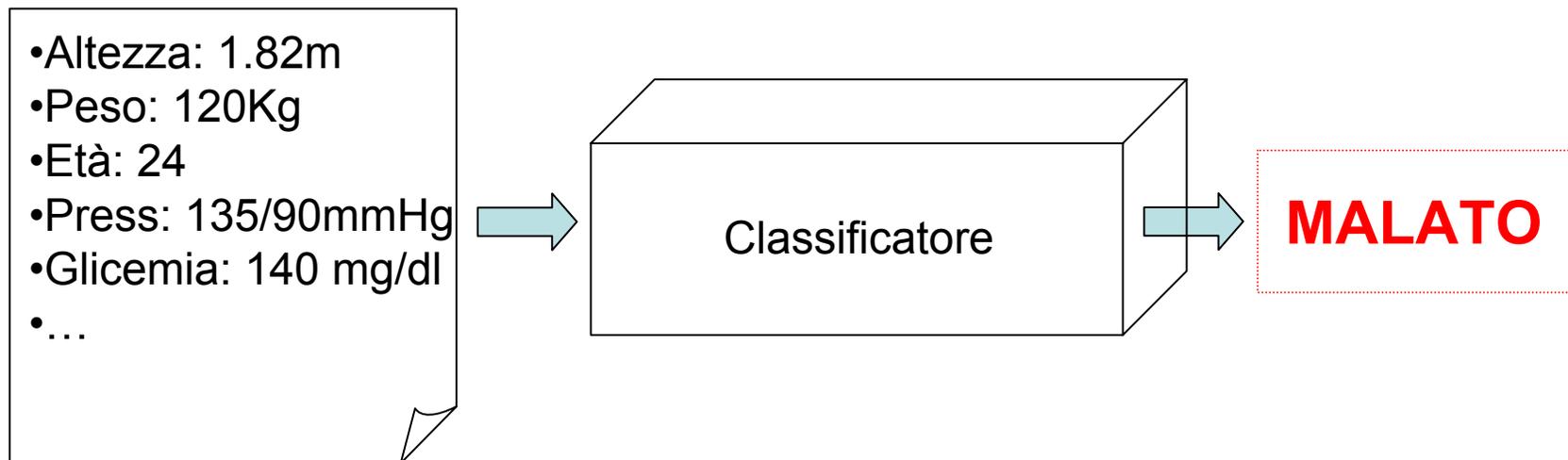
Il nuovo algoritmo: GPE
(*Guaranteed Probability of Error*)

Proprietà teoriche e risultati
sperimentali

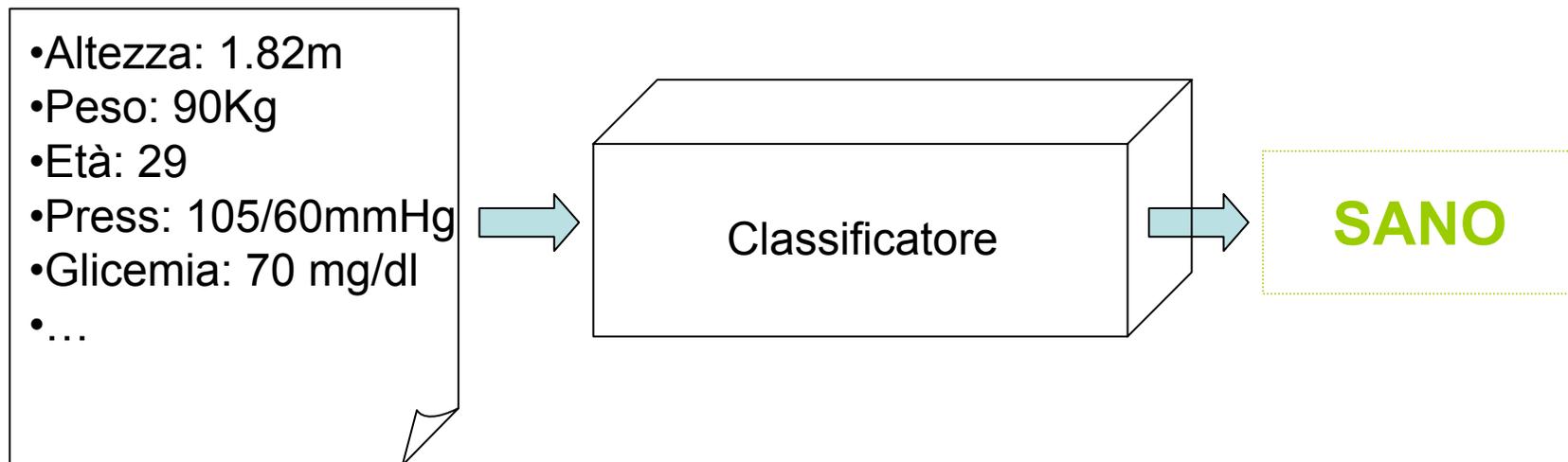
Concetti fondamentali



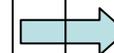
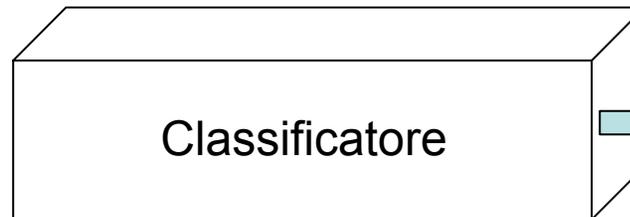
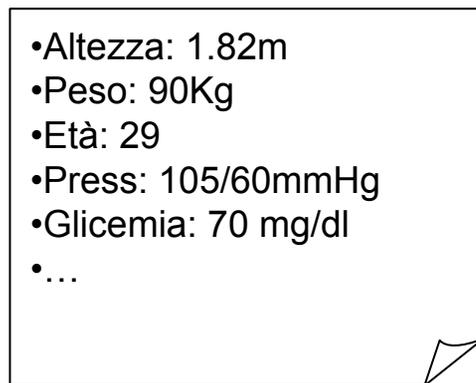
Un classificatore binario



Un classificatore binario



Costruire un classificatore



Valutare un classificatore

k attributi reali \longleftrightarrow oggetto

reale natura degli oggetti $y(\cdot) : \mathbf{R}^k \rightarrow \{0,1\}$

classificatore binario $\hat{y}(\cdot) : \mathbf{R}^k \rightarrow \{0,1\}$

misclassificazione $\hat{y}(x) \neq y(x), x \in \mathbf{R}^k$

ERRORE DI GENERALIZZAZIONE

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

Applicazioni

- Applicazioni medicali
- Applicazioni bancarie
- Riconoscimento di posta elettronica indesiderata
- Riconoscimento di pacchetti di rete
- ...

Schema della presentazione



“Un nuovo **algoritmo** per la costruzione di **classificatori** da dati sperimentali con **errore di generalizzazione *garantito***”

A. La classificazione (binaria)

B. Il nuovo algoritmo: GPE
(*Guaranteed Probability of Error*)

C. Proprietà teoriche e risultati
sperimentali

Seconda Parte



“Un nuovo **algoritmo** per la costruzione di **classificatori** da dati sperimentali con **errore di generalizzazione *garantito***”

La classificazione (binaria)

B. Il nuovo algoritmo: GPE
(*Guaranteed Probability of Error*)

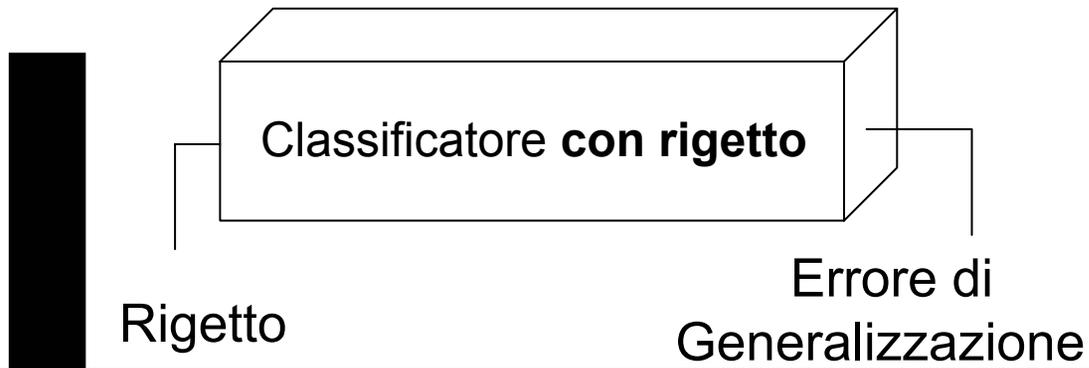
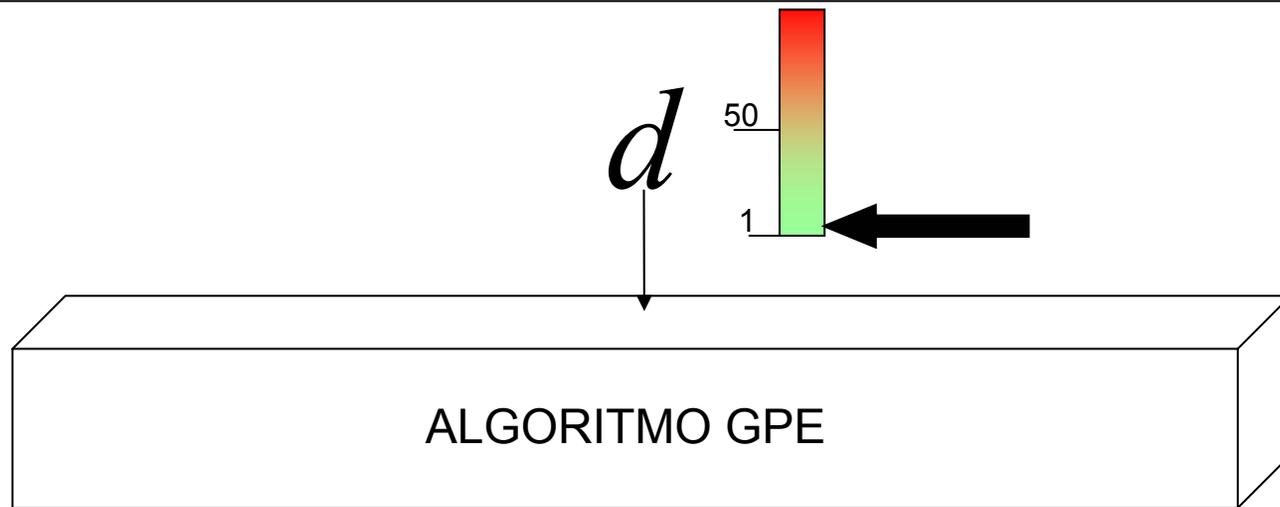
Proprietà teoriche e risultati
sperimentali

Classificatore con rigetto

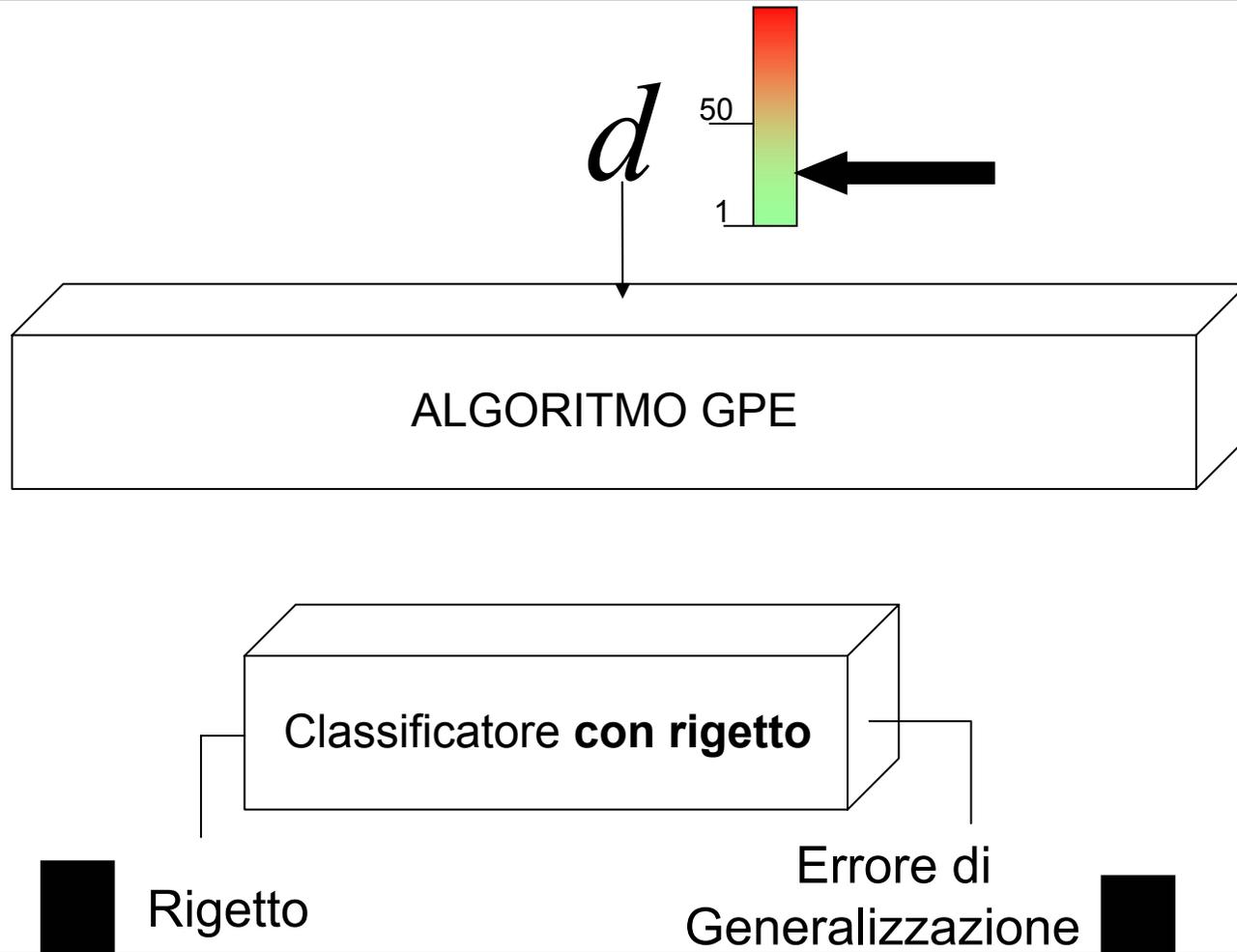
$$\hat{y}(\cdot) : \mathbf{R}^k \rightarrow \{0, 1, \textit{Rigetta}\}$$



Il parametro “d”



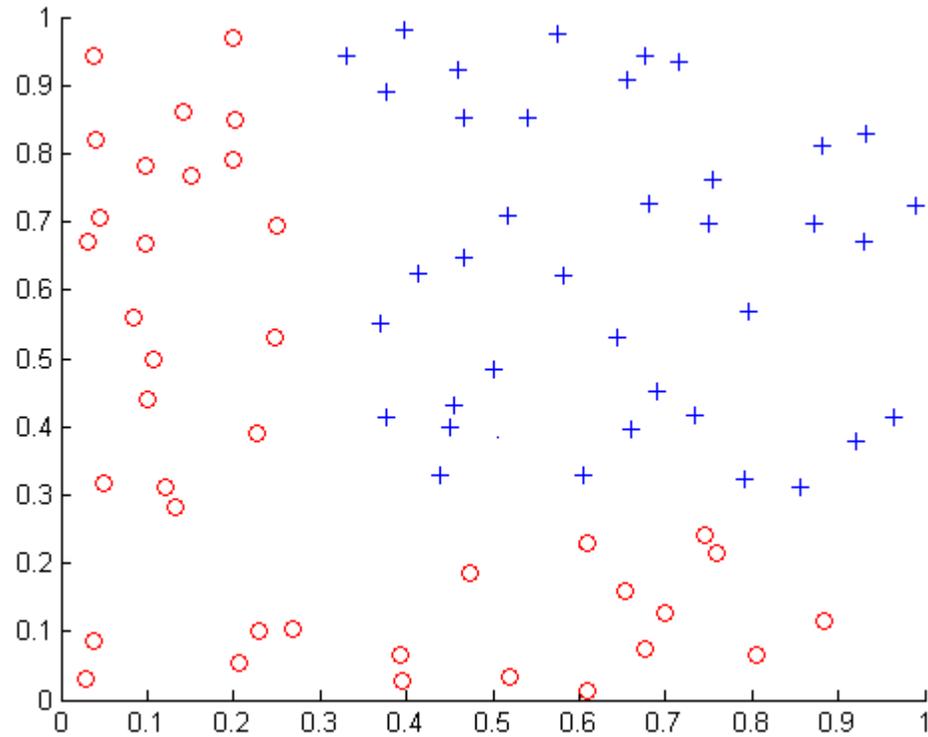
Il parametro “d”



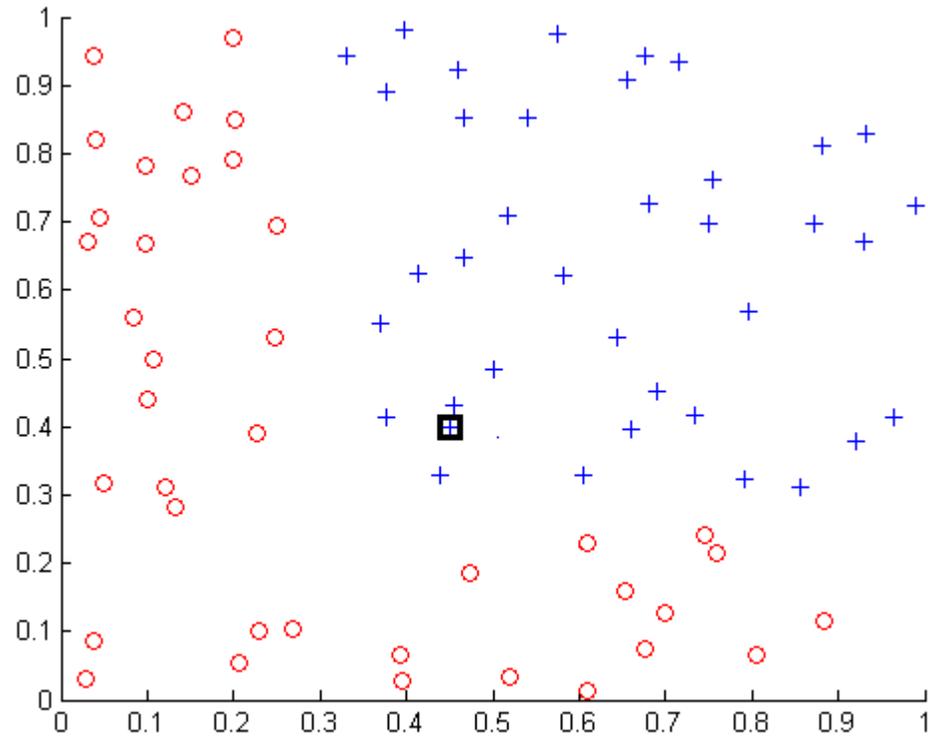
Logica di funzionamento (molto) in breve

- Costruzione incrementale di regioni
 - di natura alternata
 - di complessità decrescente

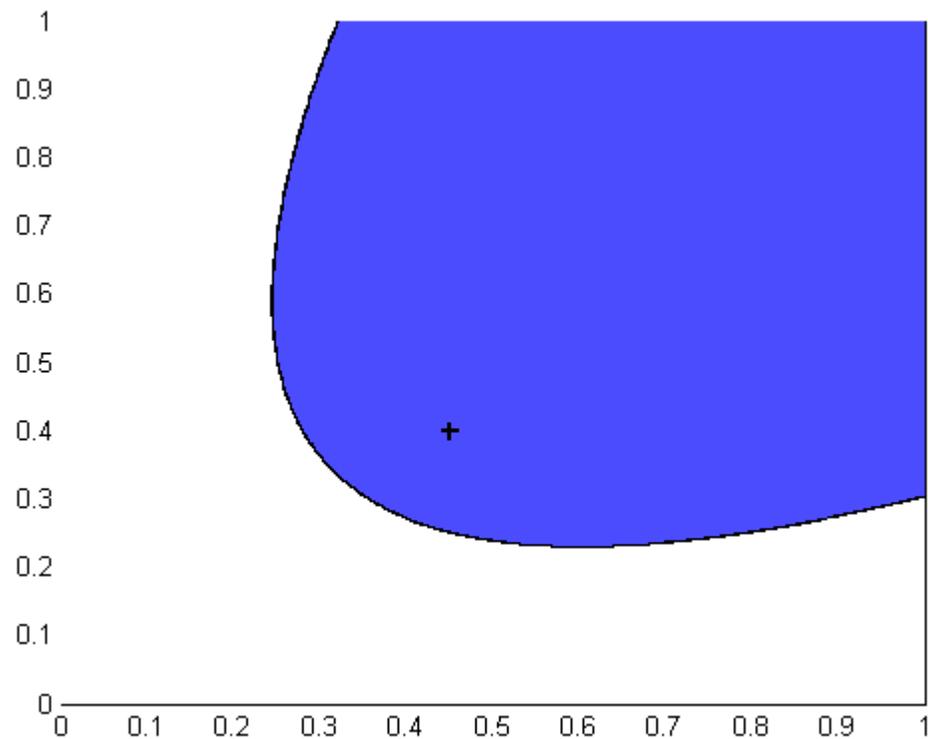
Un esempio (molto) semplice



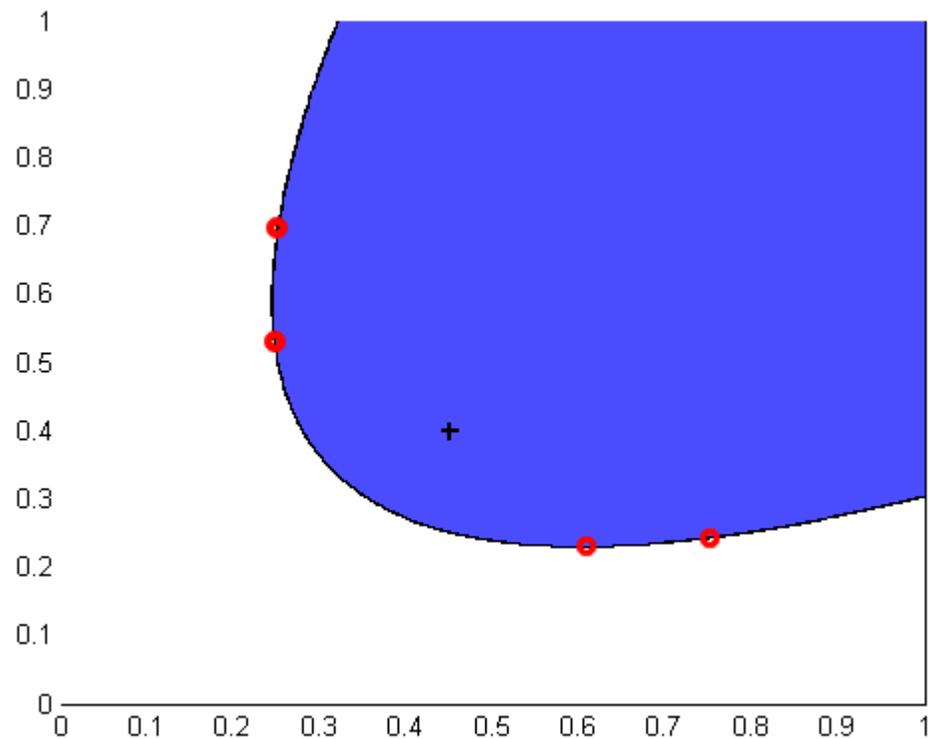
Un esempio (molto) semplice



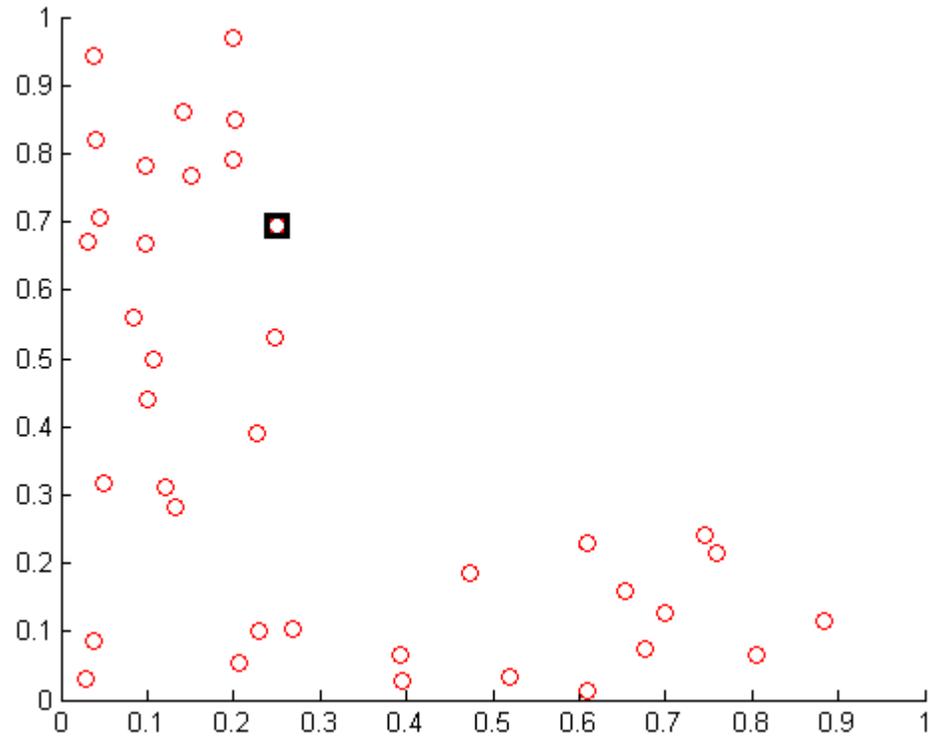
Un esempio (molto) semplice



Un esempio (molto) semplice

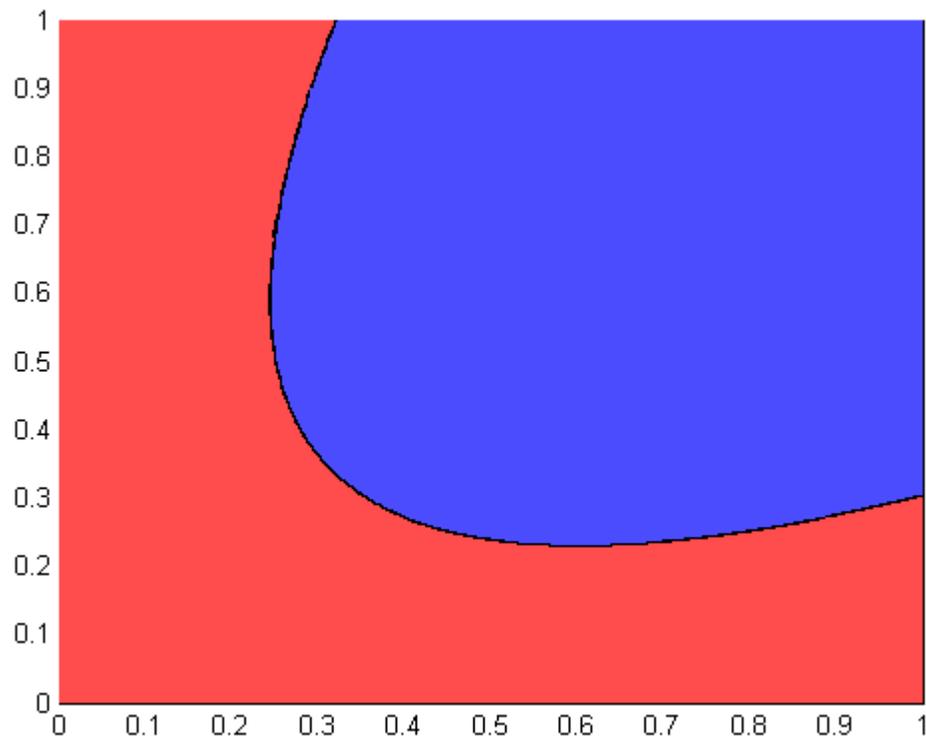


Un esempio (molto) semplice



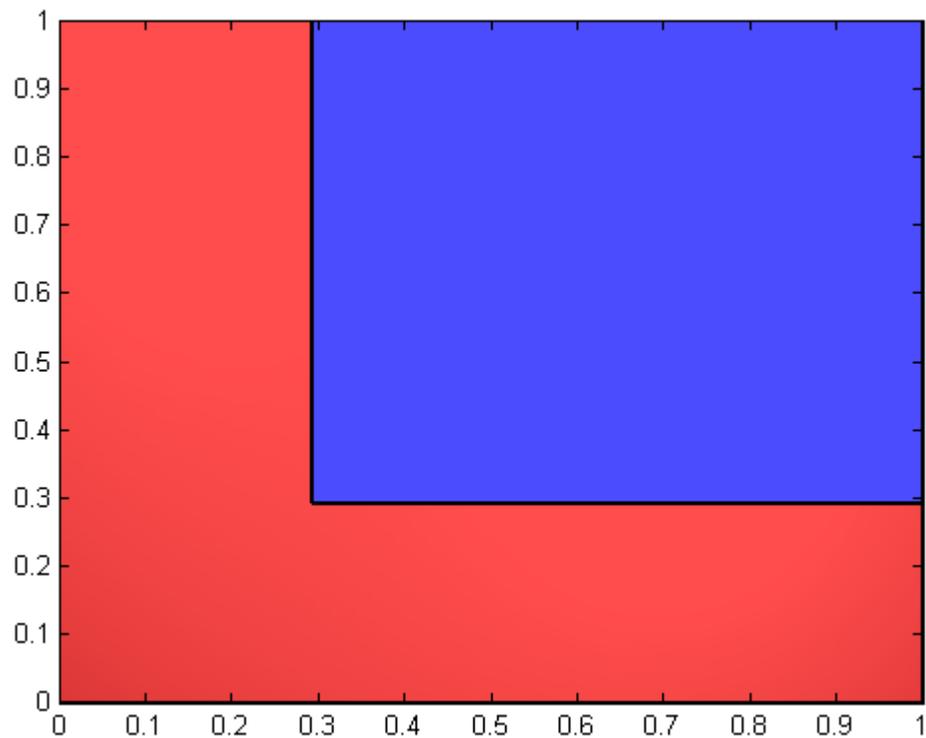
Un esempio (molto) semplice

\hat{y}

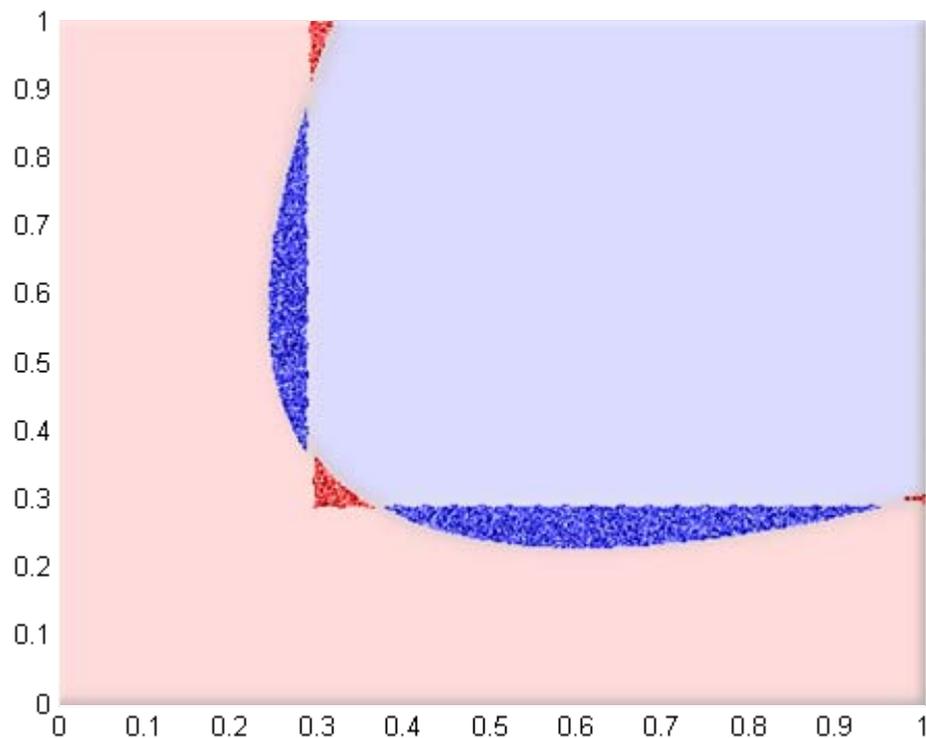


Un esempio (molto) semplice

y



Un esempio (molto) semplice

 $PE(\hat{y})$ 

PROMEMORIA

ERRORE DI GENERALIZZAZIONE

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

Schema della presentazione



“Un nuovo **algoritmo** per la costruzione di **classificatori** da dati sperimentali con **errore di generalizzazione *garantito***”

A. La classificazione (binaria)

B. Il nuovo algoritmo: GPE
(*Guaranteed Probability of Error*)

C. Proprietà teoriche e risultati
sperimentali

Terza Parte



“Un nuovo **algoritmo** per la costruzione di **classificatori** da dati sperimentali con **errore di generalizzazione *garantito***”

La classificazione (binaria)

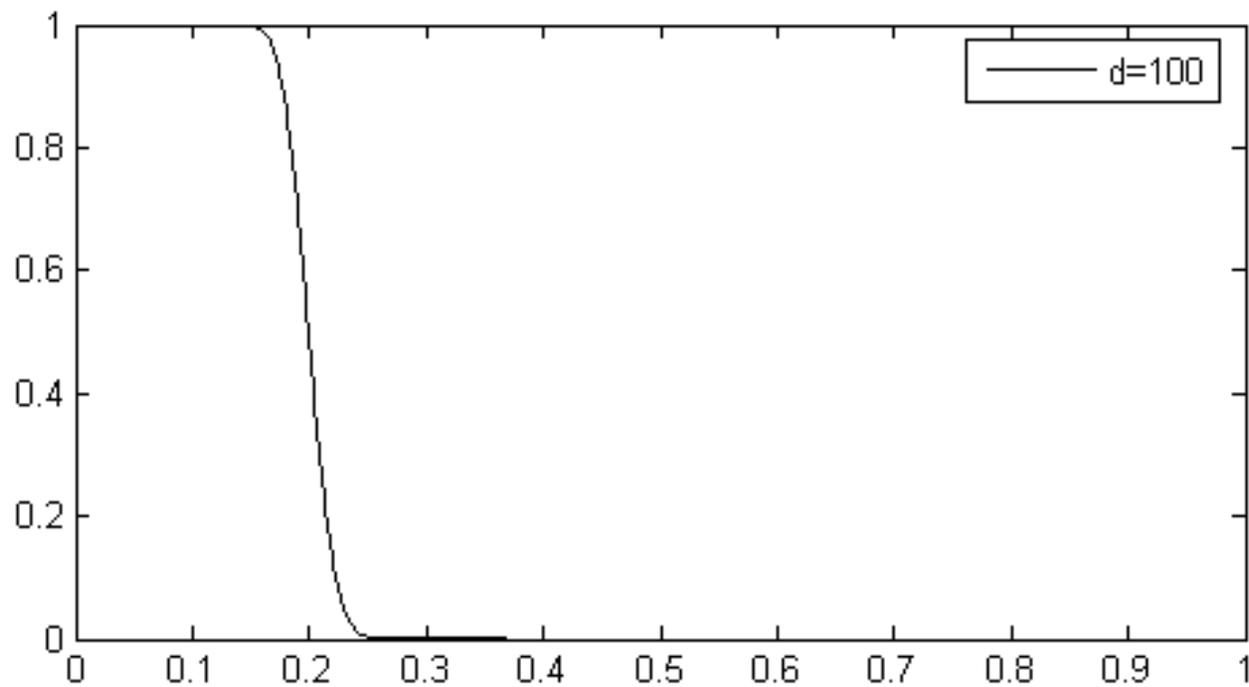
Il nuovo algoritmo: GPE
(*Guaranteed Probability of Error*)

C. Proprietà teoriche e risultati
sperimentali

Il teorema fondamentale

La distribuzione dell'errore di generalizzazione di \hat{y} è nota a priori ed è:

$$F_{PE}(z) = 1 - \sum_{i=0}^{d-1} \binom{N-1}{i} z^i (1-z)^{N-i-1}$$



Esperimenti su dati reali (I)

© Foto di Danielle Langlois



© Foto John D. Byrd



- **Lunghezza** dei sepali
- **Larghezza** dei sepali
- **Lunghezza** dei petali
- **Larghezza** dei petali

Esperimenti su dati reali (II)

d	V.A.	Tot. Rigetti	Tot. Errori	Tot. Corretti
2	2.22%	58	2 (2.22%)	30
3	3.33%	55	5 (5.56%)	30
4	4.44%	48	6 (6.67%)	36
5	5.56%	38	5 (5.56%)	47
6	6.67%	32	6 (6.67%)	52
10	11.11%	0	8 (8.89%)	82

Esperimenti su dati reali (III)

Breast: 683 campioni; 10 attributi (interi)

Haberman: 294 campioni; 3 attributi (interi)

Pima: 768 campioni; 8 attributi (interi e reali)

Bupa: 345 campioni; 6 attributi (interi e reali)

Credit: 653 campioni; 15 attributi (categorici, reali, interi)

Glass: 163 campioni; 9 attributi (**reali**)

Iris Setosa VS Versicolor e Virginica: 149 campioni; 4 attributi (**reali**)

Iris Versicolor VS Virginica: 99 campioni; 4 attributi (**reali**)

Lettere AH: 1523 campioni; 16 attributi (interi)

Lettere FT: 1571 campioni; 16 attributi (interi)

FINE

La classificazione (binaria)

Il nuovo algoritmo: GPE (*Guaranteed Probability of Error*)

Proprietà teoriche e risultati sperimentali



“Un nuovo algoritmo per la costruzione di classificatori da dati sperimentali con errore di generalizzazione *garantito*”

Grazie per l'attenzione

Università degli Studi di Brescia
Facoltà di Ingegneria
Corso di Laurea in Ingegneria Informatica
Dipartimento di Elettronica per l'Automazione



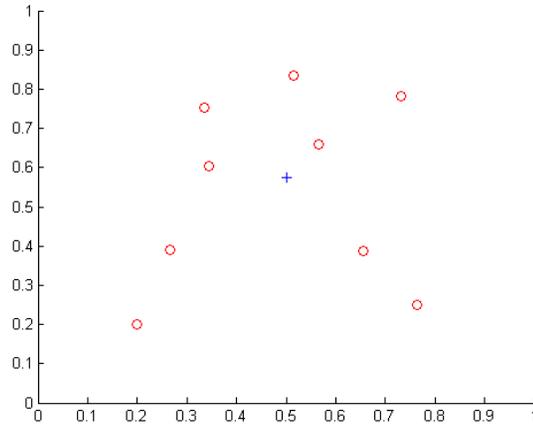
Bonus

Relatore:
Prof. Marco C. Campi

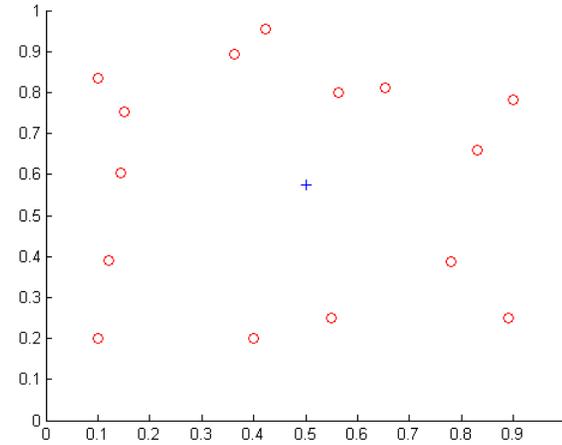
Tesi di laurea specialistica di:
Algo Carè
Matr. 60772

Tipi di regioni costruibili (I)

Parabole



Ellissi



$$P_{\frac{n(n+1)}{2}+n}(x_b, y(x_b)) :$$

$$\min_{A \in \mathbb{R}^{n \times n}, B \in \mathbb{R}^{1 \times n}} \text{Traccia}(A)$$

con vincoli:

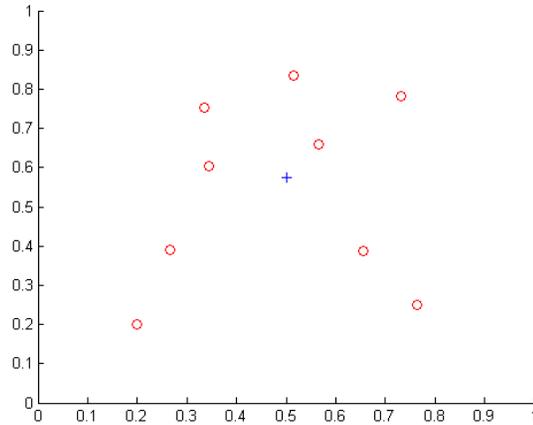
$$(x_j - x_b)^T A (x_j - x_b) + B(x_j - x_b) \geq 1, \\ \forall x_j \in L, y(x_j) \neq y(x_b)$$

$$A = A^T$$

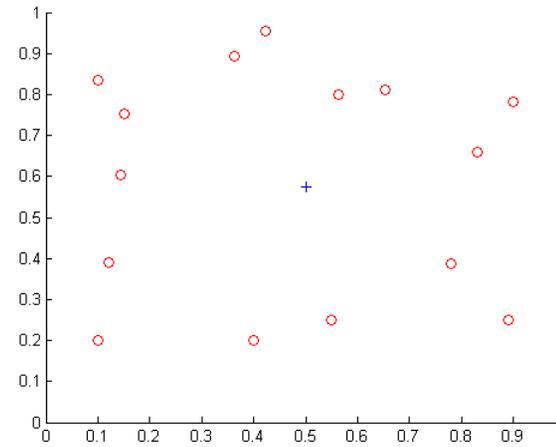
$$A \succeq 0$$

Tipi di regioni costruibili (I)

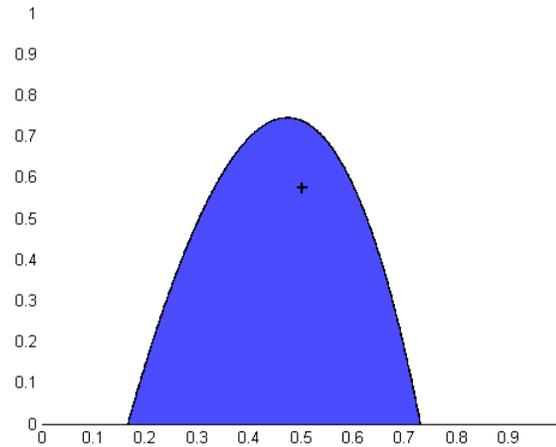
Parabole



Ellissi



I

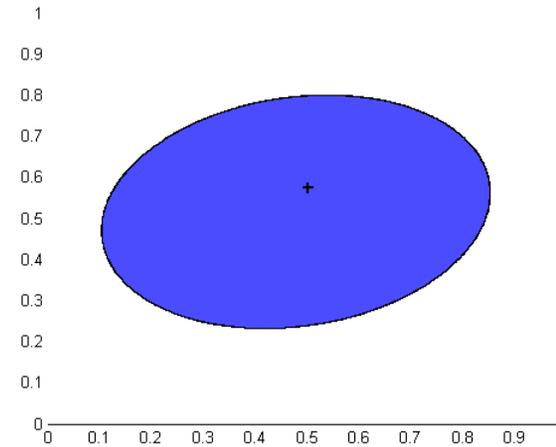


$\exists c_i$

$-$

A

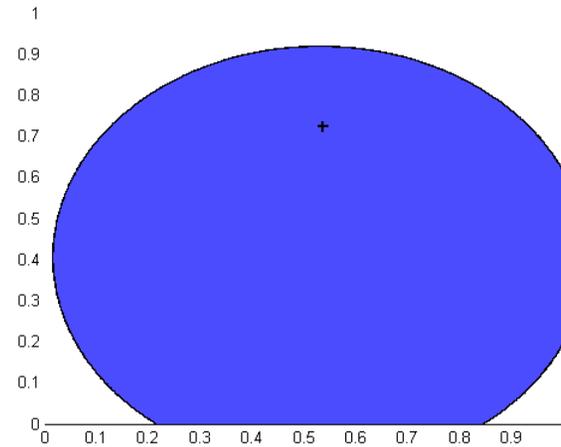
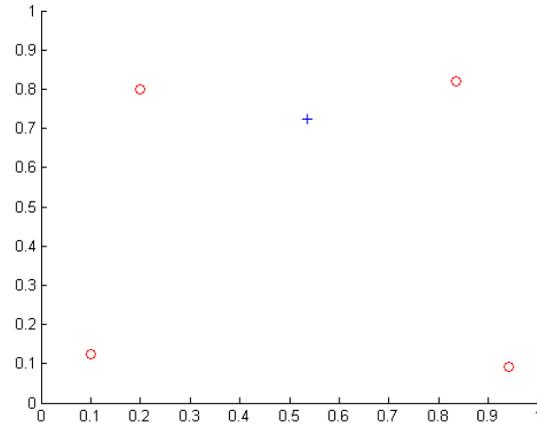
0



$1,$

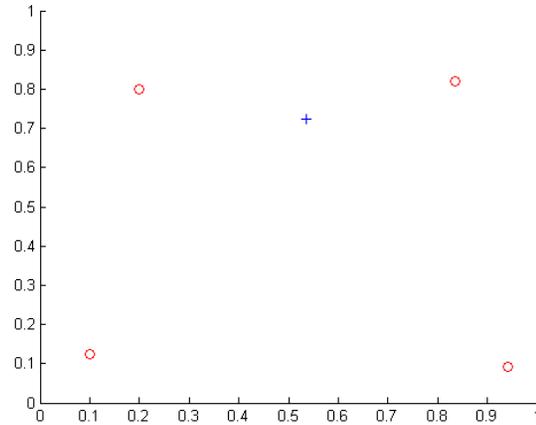
Tipi di regioni costruibili (II)

Cerchi decentrati



Tipi di regioni costruibili (II)

Cerchi decentrati



$P_{n+1}(x_b, y(x_b)):$

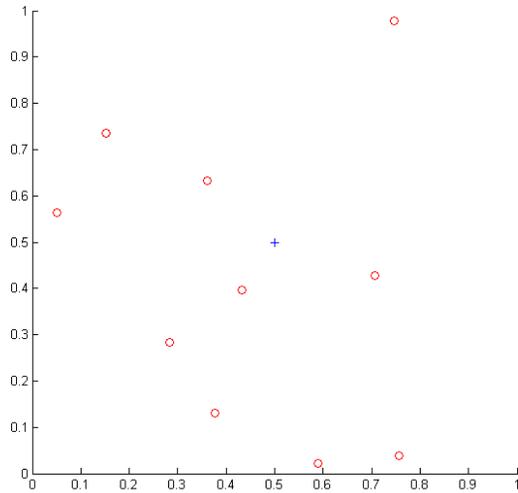
$$\min_{k \geq 0, B \in \mathbb{R}^{1 \times n}} k$$

$$\text{con vincoli: } k \|x_j - x_b\|^2 + B(x_j - x_b) \geq 1, \\ \forall x_j \in L, y(x_j) \neq y(x_b)$$

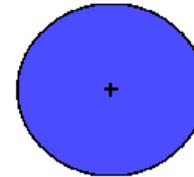


Tipi di regioni costruibili (III)

Cerchi



1
0.9
0.8
0.7
0.6
0.5
0.4



$$P_{\text{CMax}}(x_b, y(x_b)) :$$

$$\min_{k \geq 0}$$

con vincoli:

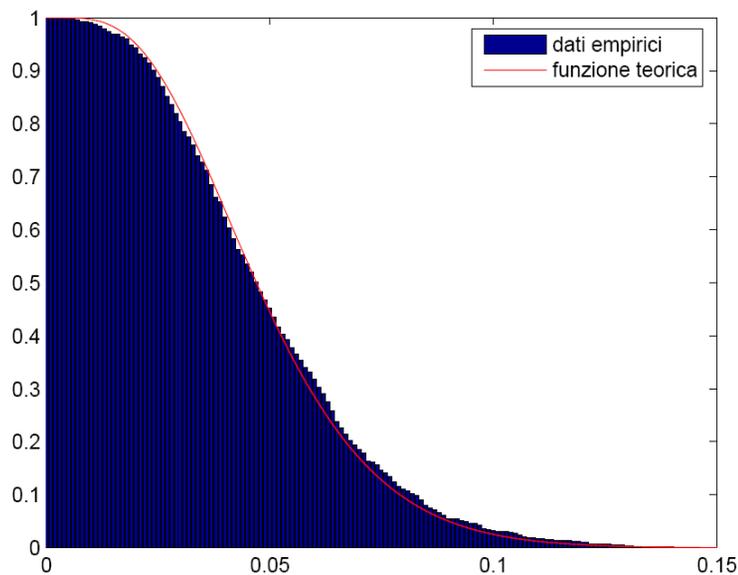
$$k \|x_j - x_b\|^2 \geq 1,$$

$$\forall x_j \in L, y(x_j) \neq y(x_b)$$

Risultati sperimentali

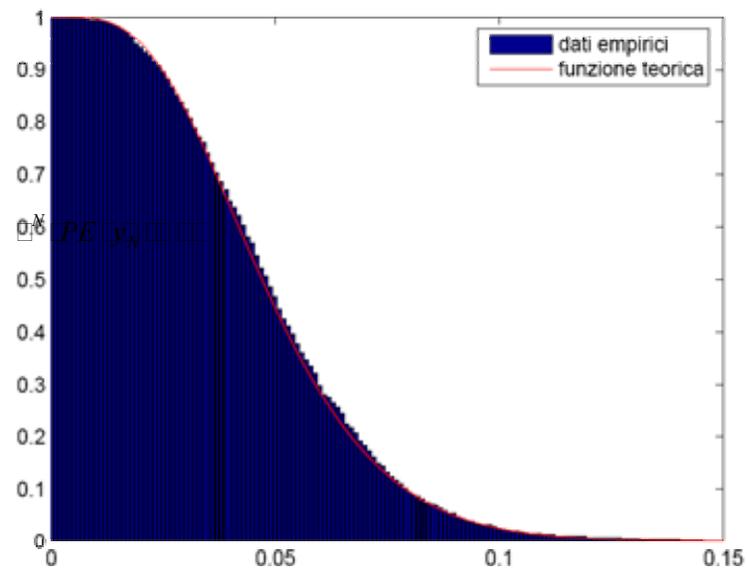
d=5

RealGPE



N=100, 1000Multiestrazioni

IdealGPE

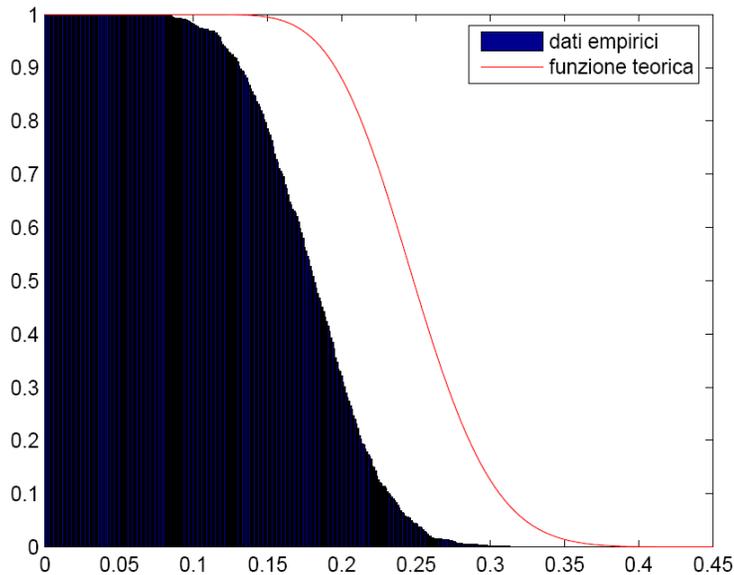


N=100, 1000Multiestrazioni

Risultati sperimentali

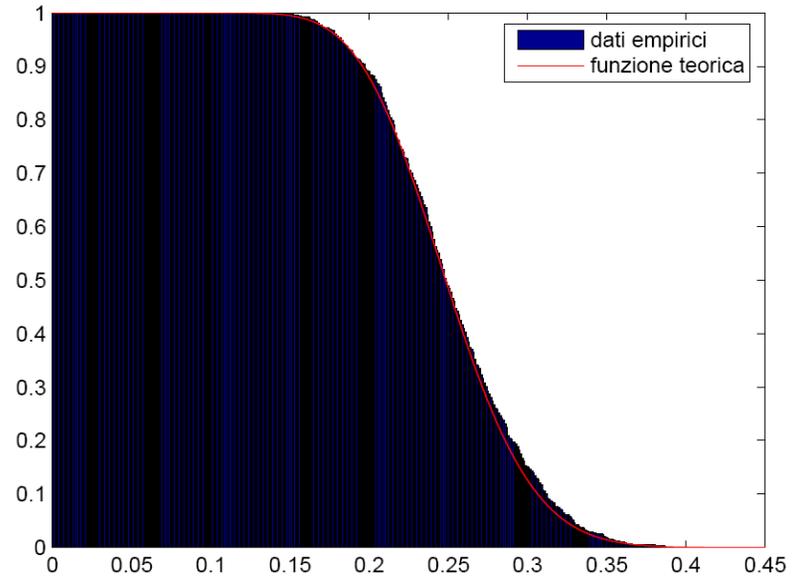
d=25

RealGPE



N=100, 1000Multiestrazioni

IdealGPE



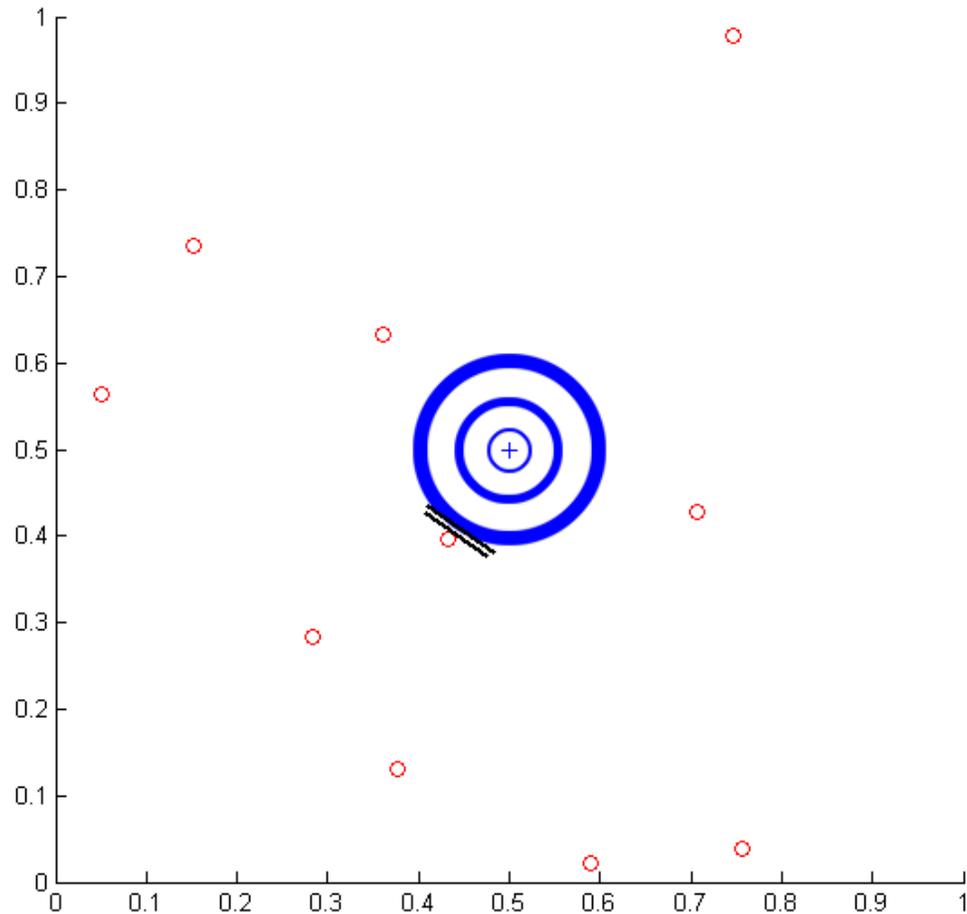
N=100, 1000Multiestrazioni

Risultati sperimentali (errori in alcuni test)

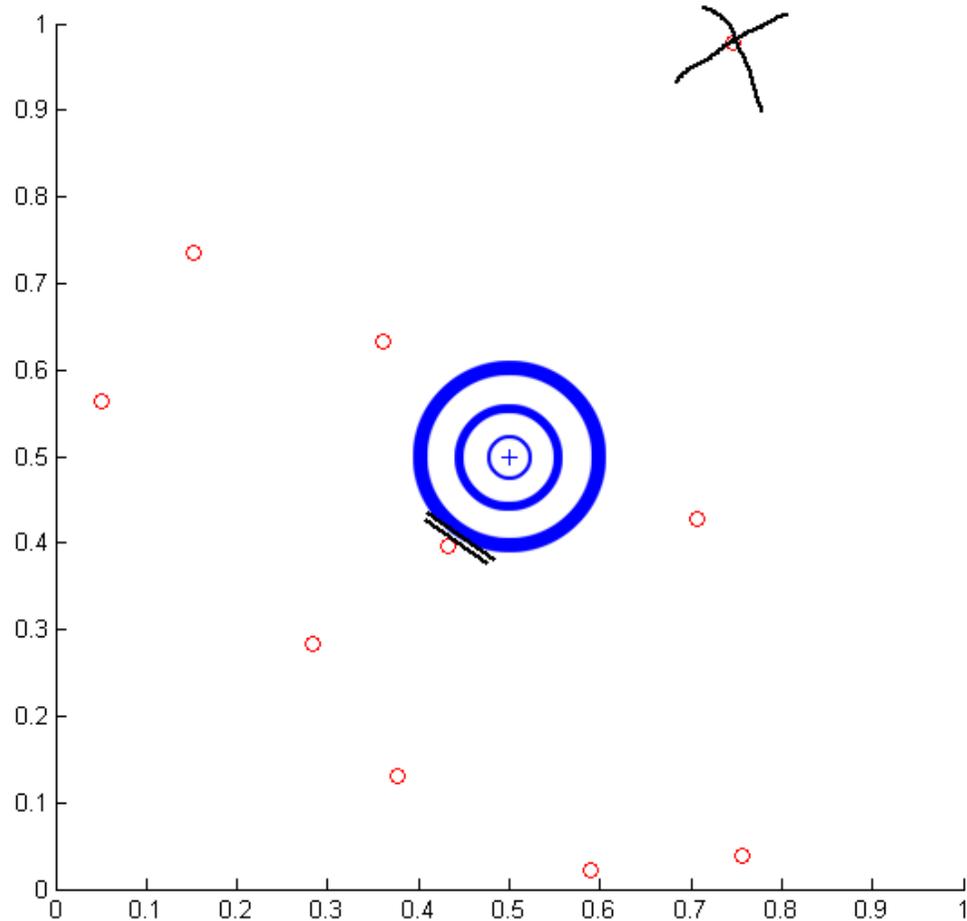
Dataset	RealGPE	NNC	SVM $C = \infty$	SVM $C < \infty$
Breast	28.12	29	27	19
Haberman	84.14	107	111	71
Pima	243.54	247	243	203
Bupa	130.9	124	121	107
Credit	151.7	214	205	190
Glass	40.75	36	42	34

- Competitivo in quanto classificatore senza rigetto
- Garanzie stringenti
- Robustezza a fronte di ipotesi teoriche non soddisfatte

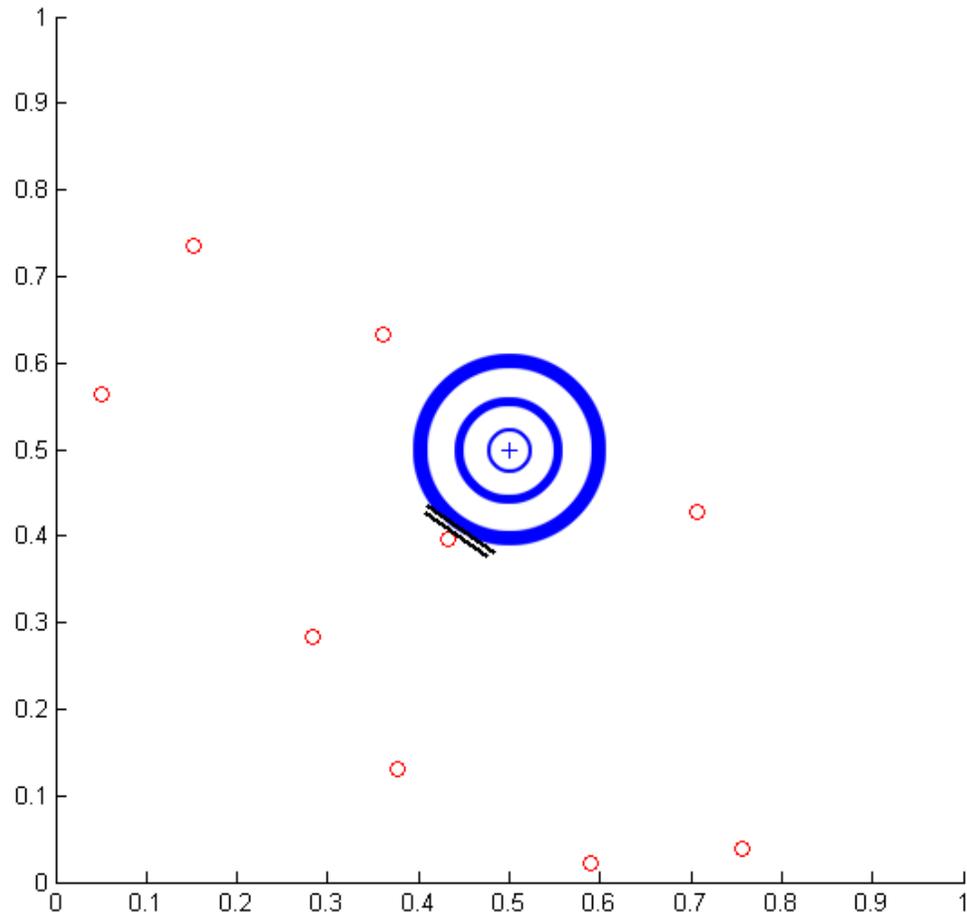
Relazione tra errore e punti sul bordo (spiegazione intuitiva)



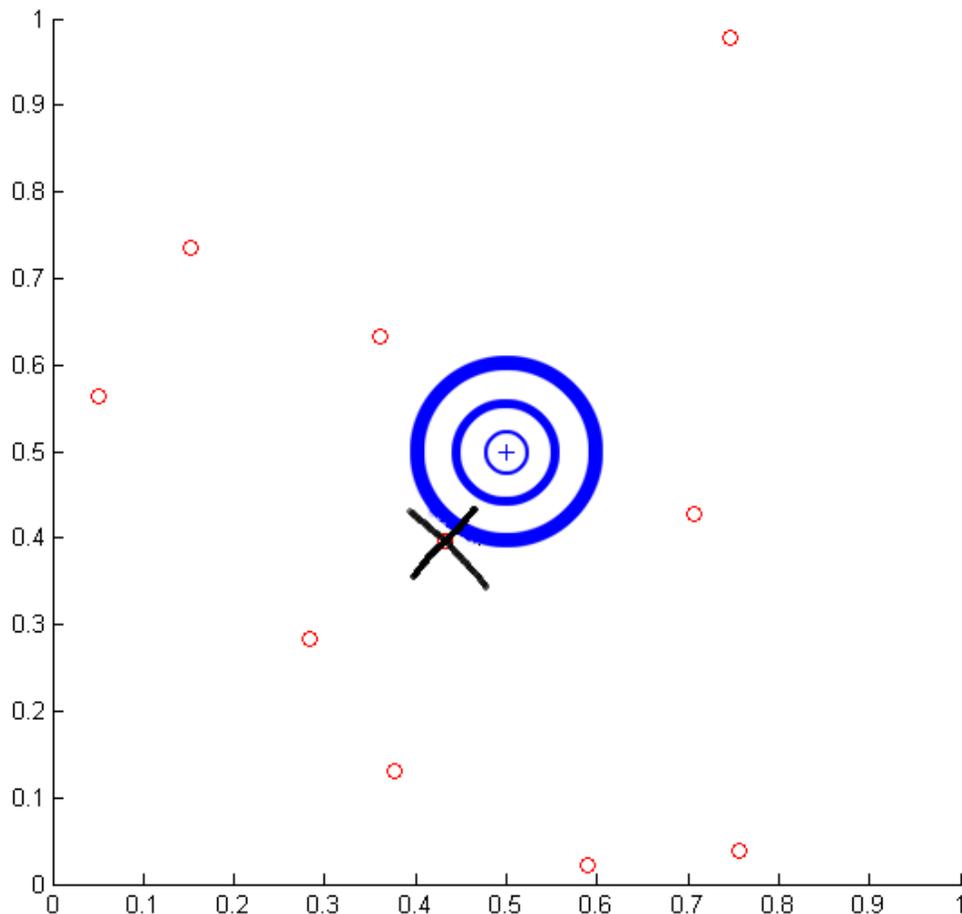
Relazione tra errore e punti sul bordo (spiegazione intuitiva)



Relazione tra errore e punti sul bordo (spiegazione intuitiva)

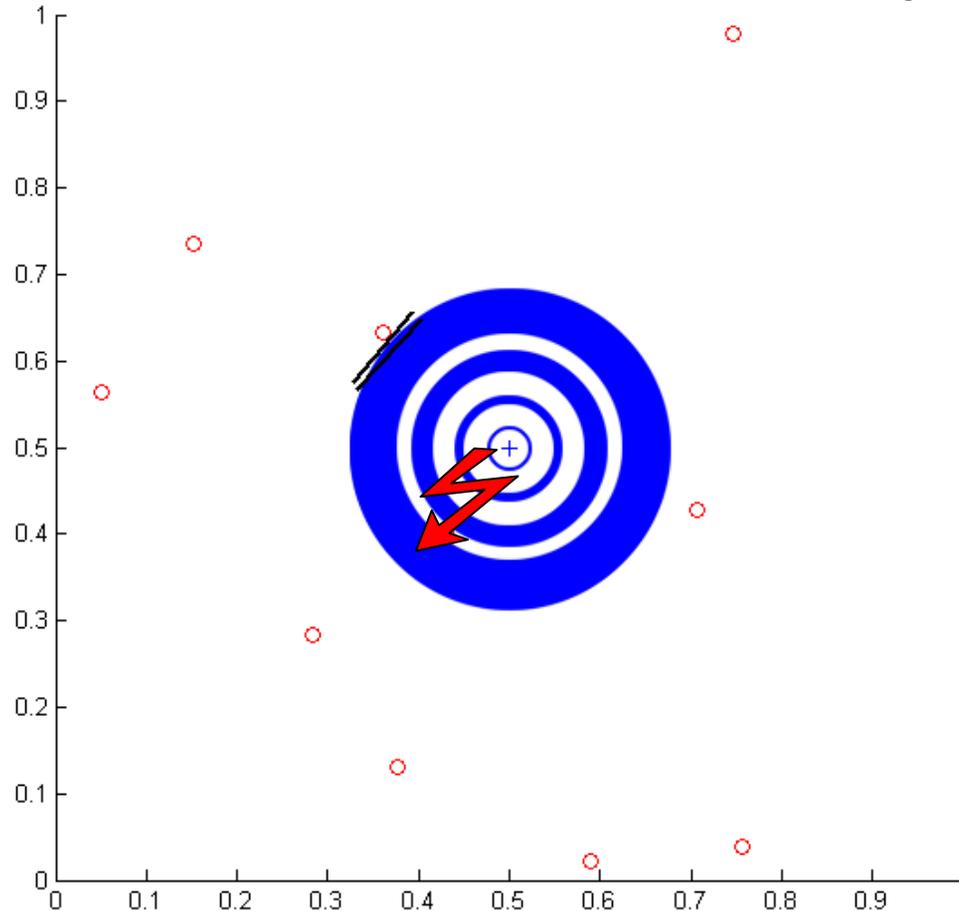


Relazione tra errore e punti sul bordo (spiegazione intuitiva)



Relazione tra errore e punti sul bordo (spiegazione intuitiva)

Un punto che si trovava sul bordo della regione-soluzione, se rimosso, rientra nella nuova regione costruita.



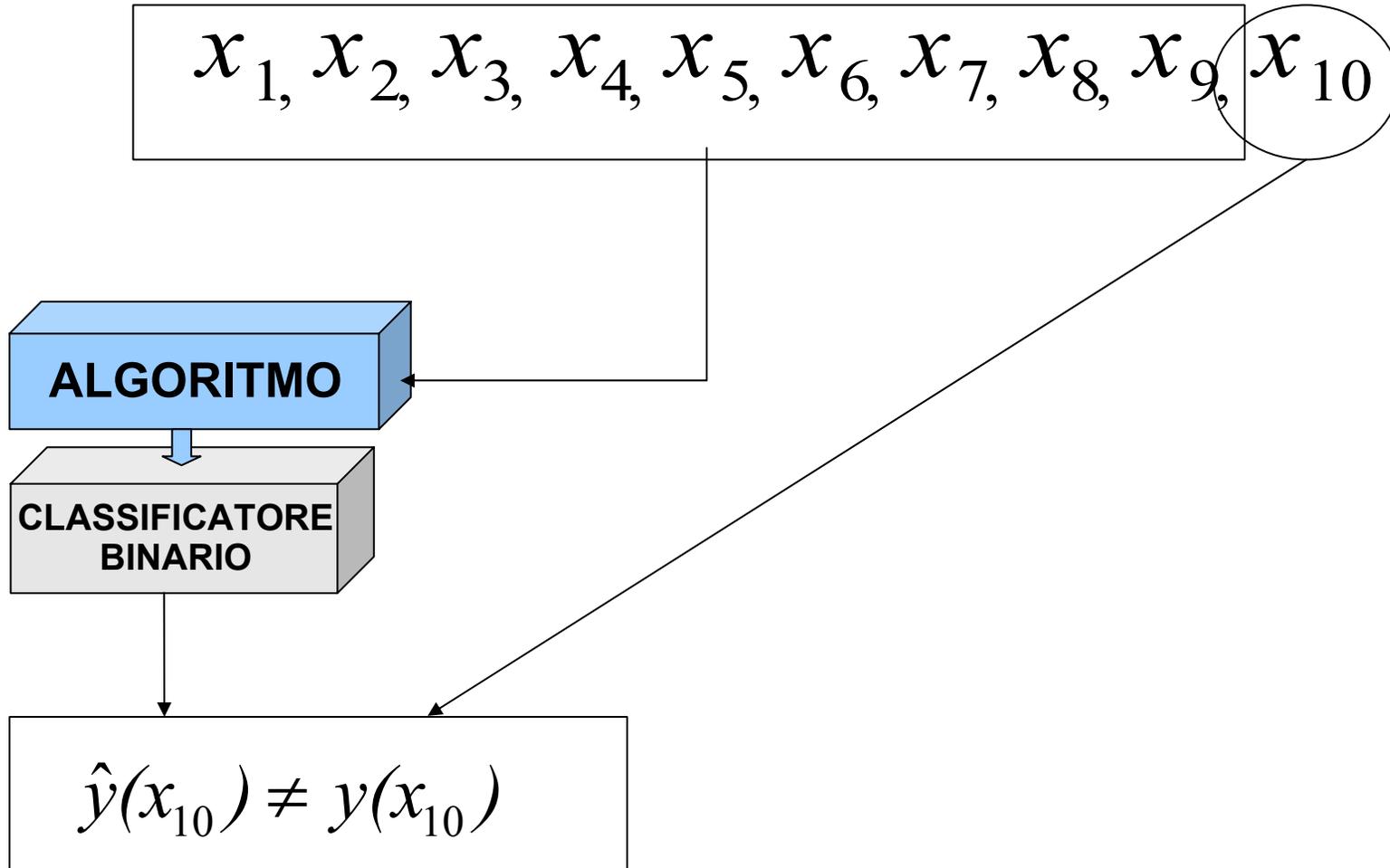
Il leave-one-out (spiegazione intuitiva)

$x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}$

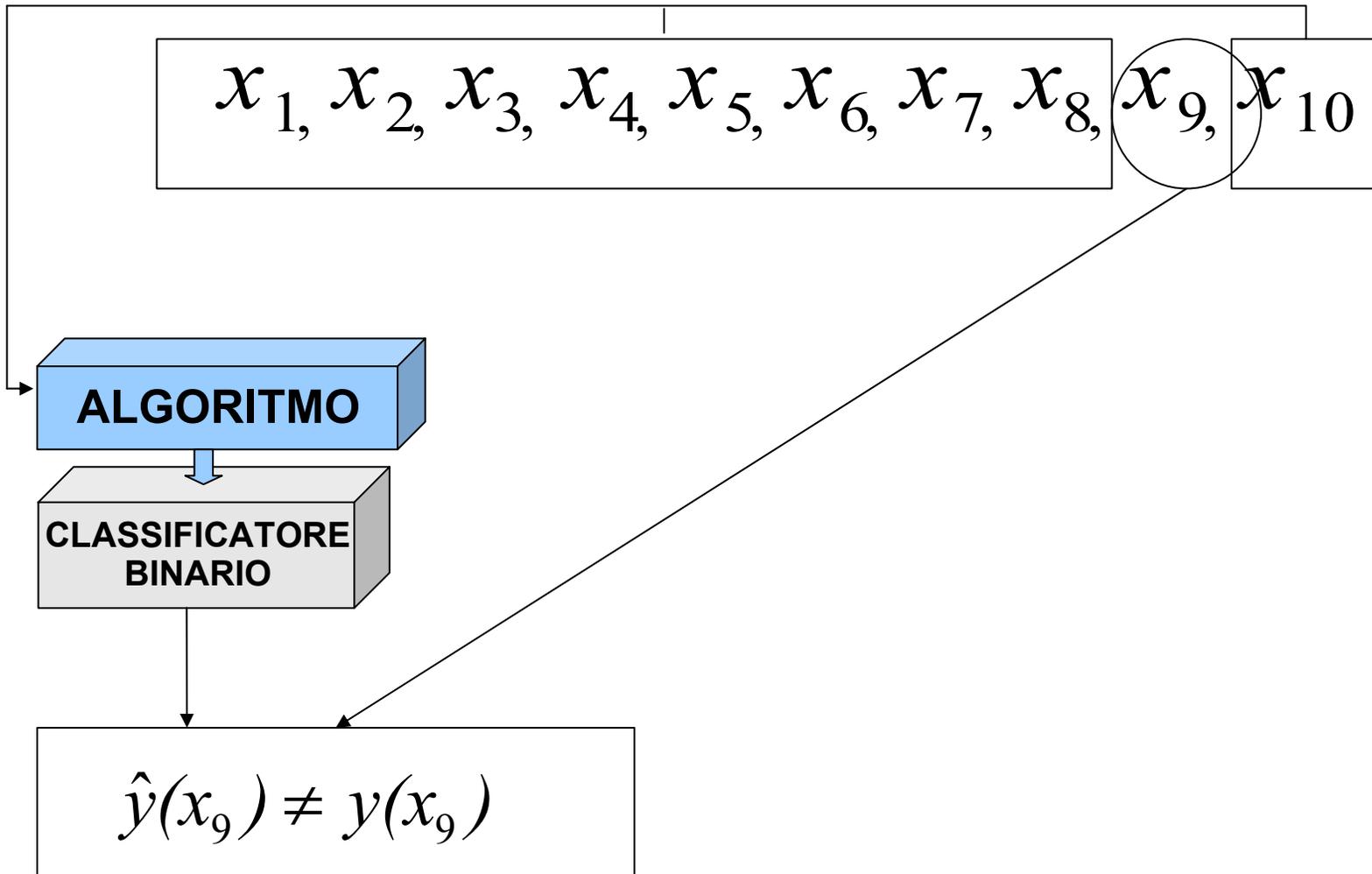
ALGORITMO

CLASSIFICATORE
BINARIO

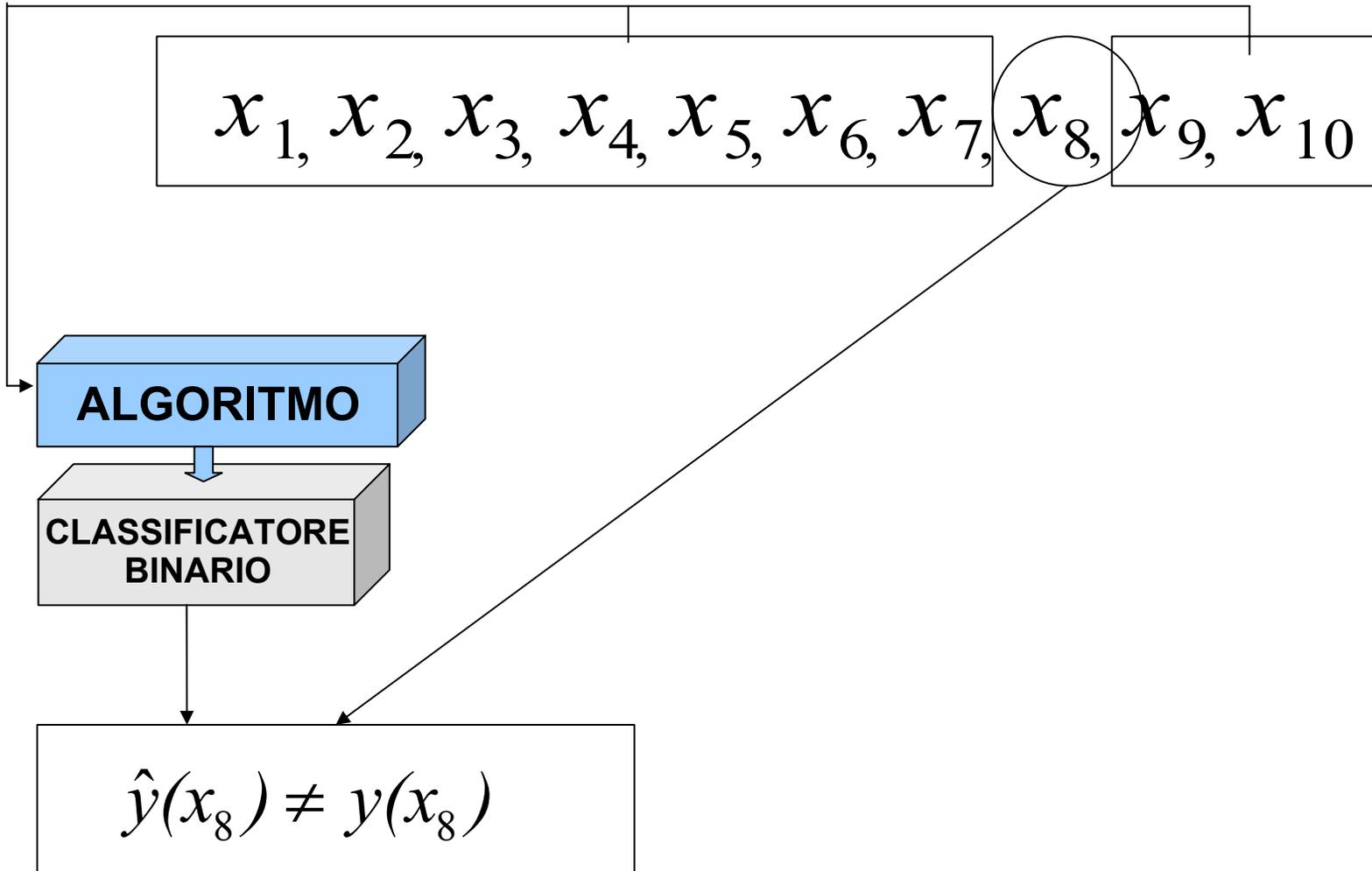
$$\hat{y}(x_{10}) \neq y(x_{10})$$



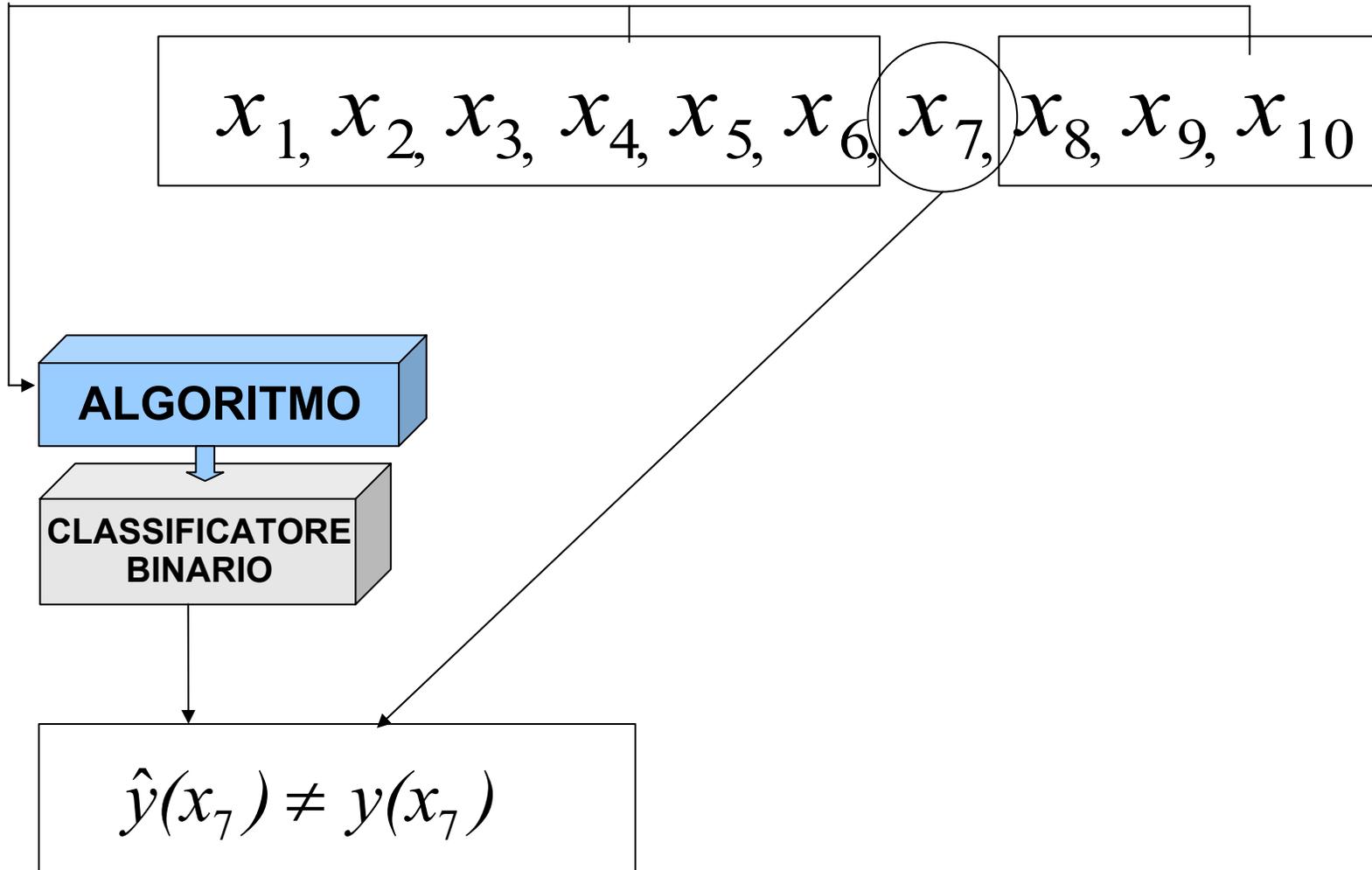
Il leave-one-out (spiegazione intuitiva)



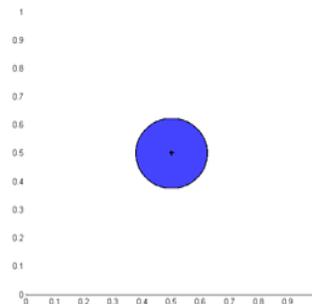
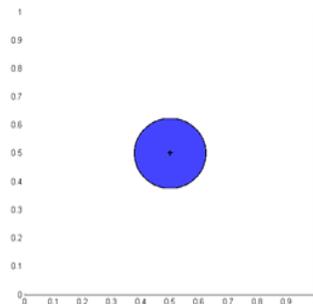
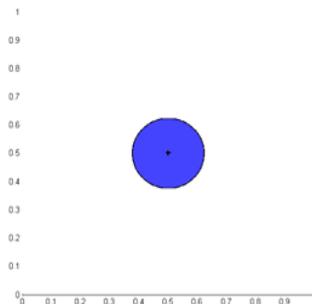
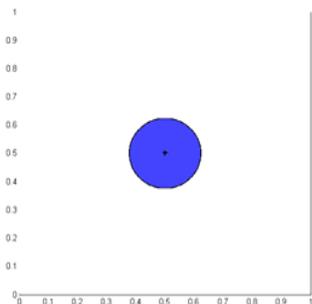
Il leave-one-out (spiegazione intuitiva)



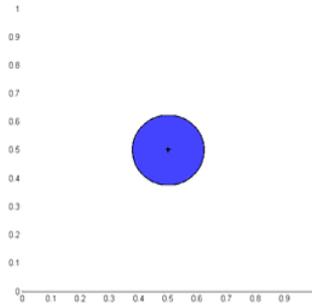
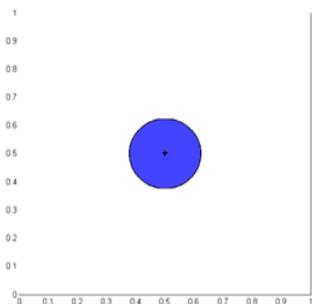
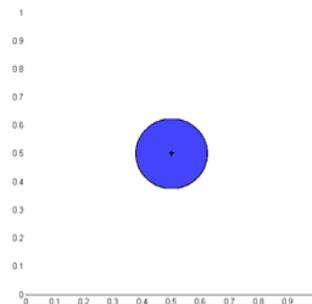
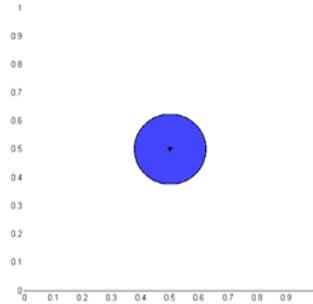
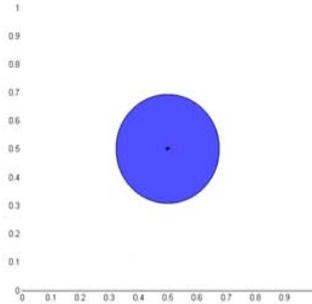
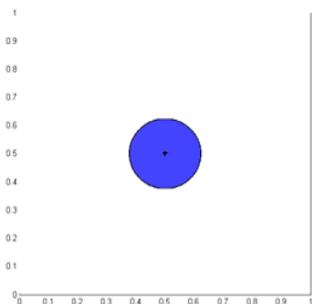
Il leave-one-out (spiegazione intuitiva)



Il leave-one-out (spiegazione intuitiva)



L



$$\frac{1}{10}$$

$$\frac{\textit{punti sul bordo}}{N}$$

Relazione tra errore e punti sul bordo (spiegazione intuitiva)

- L'errore di leave-one-out è molto legato al valore atteso dell'errore di generalizzazione
- Dalla distribuzione dell'errore di generalizzazione si può ricavare il valore atteso dell'errore di generalizzazione:

$$\frac{d}{N}$$

- L'algoritmo è stato costruito con proprietà che permettono di generalizzare quanto visto nel caso banale del cerchio (ottimizzazione convessa)