

# Classificazione e regressione

Alberto Borghese

Università degli Studi di Milano  
Laboratorio di Sistemi Intelligenti Applicati (AIS-Lab)  
Dipartimento di Informatica  
[alberto.borghese@unimi.it](mailto:alberto.borghese@unimi.it)



## Riassunto



- **Classificazione**
- Analisi errore



## Features



- Globali: livello di luminosità medio, varianza, contenuto in frequenza.....
- Feature locali
- Valutazione della distanza



## Dal clustering alla classificazione



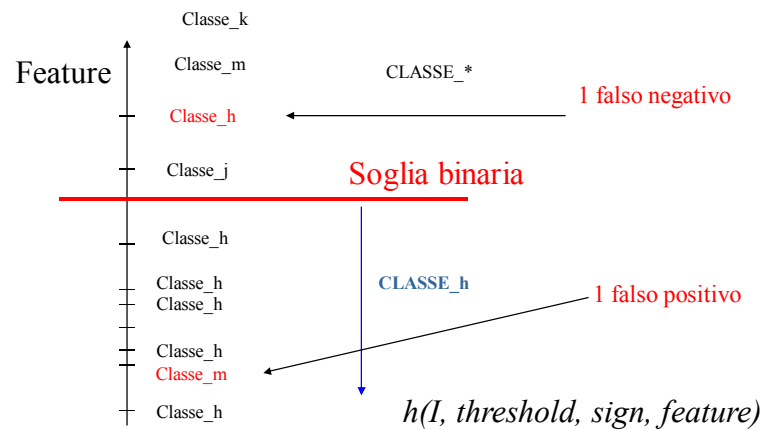
- Clustering (non supervisionato):
  - ◆ Input:  $\{D_0, D_1, D_2, \dots, D_n\}$
  - ◆ Output:  $\{D_0|C_j, D_1|C_k, \dots\}$                       C cluster di appartenenzaPartizionamento dello spazio di input mediante misura di dissimilarità
- Classificazione (supervisionato):
  - ◆ Input:  $\{D_0|C_j, D_1|C_k, \dots\}$                       C classi (etichette)
  - ◆ Output: Modello  $f(w)$                               w parametri.Partizionamento dello spazio di input sfruttando le etichette date dal "teacher"  
Il modello deve essere in grado di fare predizioni corrette.



## Classificatore binario

Classificatore binario. Si seleziona una feature e si sceglie la soglia ottimale.

Un classificatore binario è costituito da: feature, soglia, verso.



A.A. 2015-2016

5/48

<http://borghese.di.unimi.it>



## Gli approcci

Approccio incrementale mediante boosting: aggiungo una feature di classificazione dopo l'altra, ma le considero tutte assieme quando faccio la valutazione.

Approccio incrementale mediante alberi di decisione: aggiungo un feature di classificazione dopo l'altra, e le considero in sequenza.

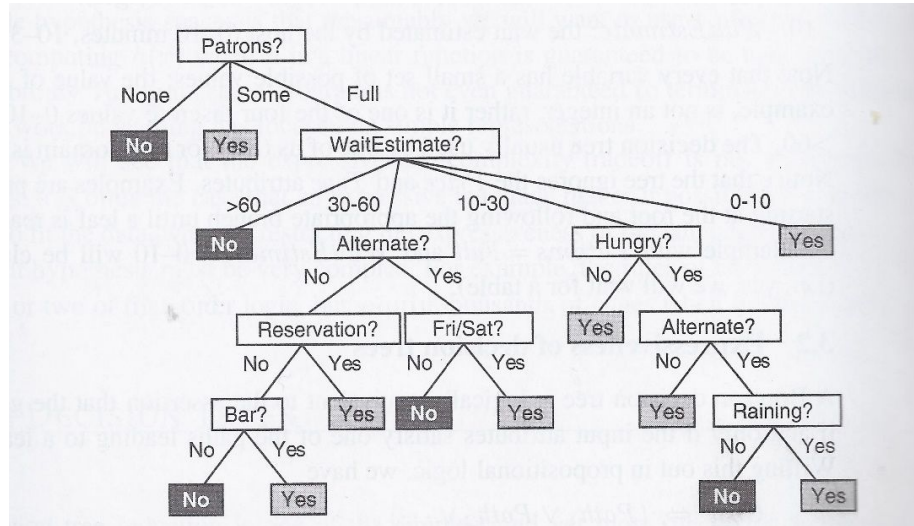
A.A. 2015-2016

6/48

<http://borghese.di.unimi.it>



# Albero di decisione



A.A. 2015-2016

7/48

<http://borghese.di.unimi.it>



# I dati misurati sul campo



Examp	Input attributes										Goal (wait)
	Alt	Bar	Fri	Hun	Pat	Price	Rain	Res	Type	Est	
x <sub>1</sub>	Yes	No	No	Yes	Some	\$\$\$	No	Yes	Fr	0-10	y <sub>1</sub> = Y
x <sub>2</sub>	Yes	No	No	Yes	Full	\$	No	No	Thai	30-60	y <sub>2</sub> = N
x <sub>3</sub>	No	Yes	No	No	Some	\$	No	No	Burg	0-10	y <sub>3</sub> = Y
x <sub>4</sub>	Yes	No	Yes	Yes	Full	\$	Yes	No	Thai	10-30	y <sub>4</sub> = Y
x <sub>5</sub>	Yes	No	Yes	No	Full	\$\$\$	No	Yes	Fr	>60	y <sub>5</sub> = N
x <sub>6</sub>	No	Yes	No	Yes	Some	\$\$	Yes	Yes	I	0-10	y <sub>6</sub> = Y
x <sub>7</sub>	No	Yes	No	No	None	\$	Yes	No	Burg	0-10	y <sub>7</sub> = N
x <sub>8</sub>	No	No	No	Yes	Some	\$\$	Yes	Yes	Thai	0-10	y <sub>8</sub> = Y
x <sub>9</sub>	No	Yes	Yes	No	Full	\$	Yes	No	Burg	>60	y <sub>9</sub> = N
x <sub>10</sub>	Yes	Yes	Yes	Yes	Full	\$\$\$	No	Yes	I	10-30	y <sub>10</sub> = N
x <sub>11</sub>	No	No	No	No	None	\$	No	No	Thai	0-10	y <sub>11</sub> = N
x <sub>12</sub>	Yes	Yes	Yes	Yes	Full	\$	No	No	Burg	30-60	y <sub>12</sub> = Y

A.A. 2015-2016

8/48

<http://borghese.di.unimi.it>



## Scelta della feature

Feature più importante selezionata per prima.

Come si fa a definire l'importanza? Ad esempio % dei casi in cui la classificazione è corretta.

Se la classificazione corretta è del 100%?

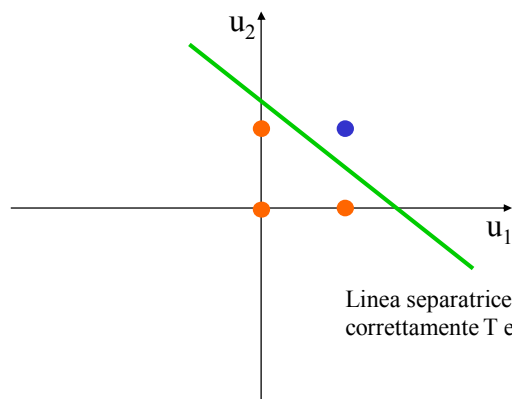


## Classificazione mediante curve separatrici

AND

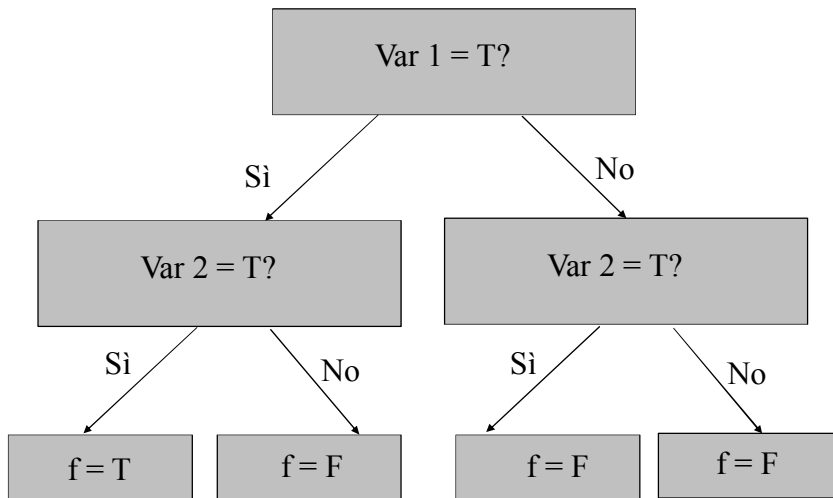
Var1	Var2	Y
F	F	F
F	T	F
T	F	F
T	T	T

$u_1$	$u_2$	$y$
0	0	0
0	1	0
1	0	0
1	1	1





### Classificazione della funzione AND mediante albero



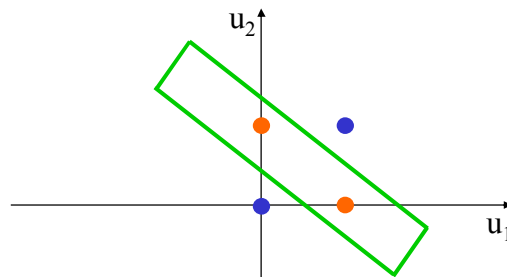
### Classificazione mediante curve separatrici



XOR

Var1	Var2	Y
F	F	F
F	T	T
T	F	T
T	T	F

$U_1$	$u_2$	Y
0	0	0
0	1	1
1	0	1
1	1	0

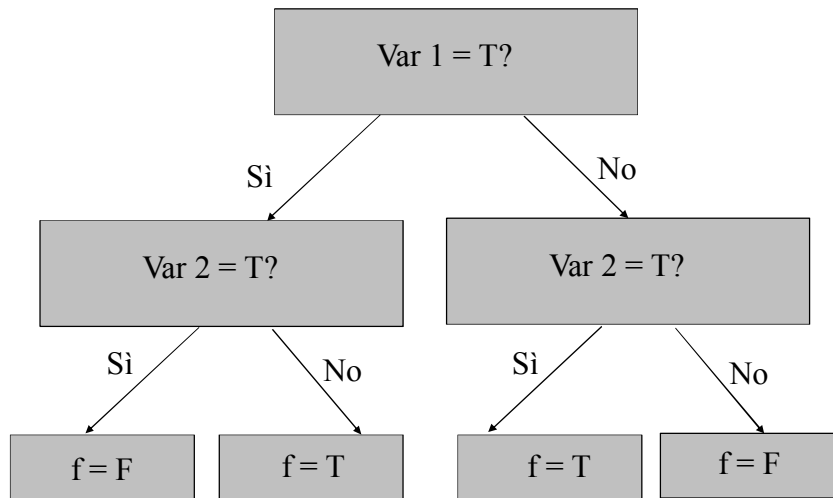


Non esiste una retta separatrice. Occorre utilizzare una spezzata a una curva. Classifica correttamente T e F.

Induce una suddivisione dello spazio.



## Classificazione della funzione XOR mediante albero



## Riassunto



- **Classificazione**
- Analisi errore



## How to classify the error introduced by a model?



Is the model good enough?

Does it have enough parameters?

Does it cover the input domain (in all dimensions)?

**This is not enough to obtain a good model!!**

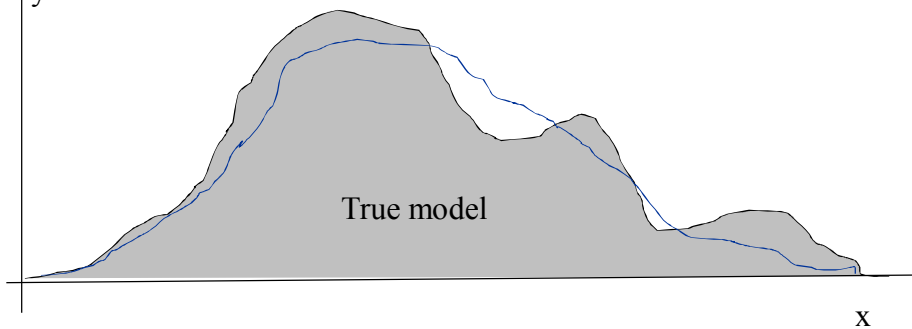
The model should be properly tuned to the data



## How to classify the error introduced by a model?



How is the estimated model related to the true model?



Bias and variability trade-off

Bias is the distance of the model curve from the true unknown curve.  
It is associated to model error.

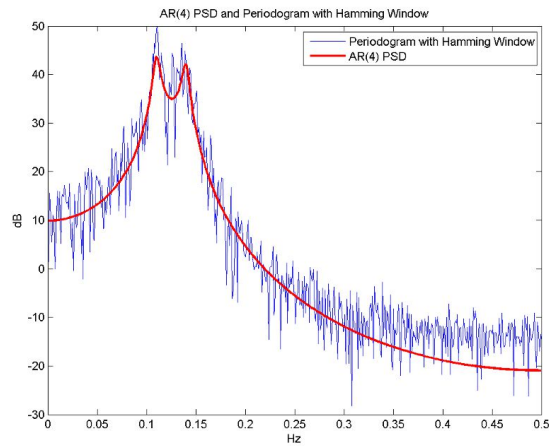




## Variability



How are the measured points related to the estimated model?



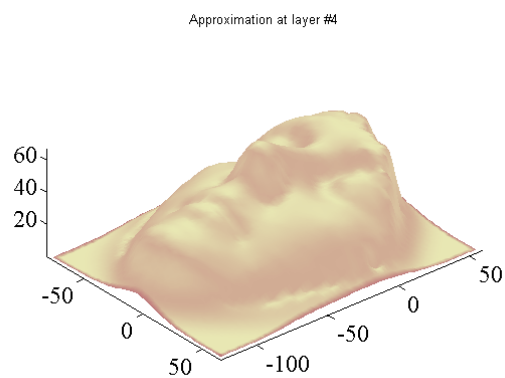
Given  $P_{\text{mes}}(x_{\text{mes}}, y_{\text{mes}})$  and  $y = f(x)$ , the error is measured as:  $\text{dist}(y_{\text{mes}}, f(x_{\text{mes}}))$ , for instance Euclidean distance. It is associated to measurement error.

If variability goes to zero, bias increases and overfitting arises.

In a good model, variability tends to the statistics of the measurement noise.



## Problema dell'overfitting dovuto a sovrapparametrizzazione



Quante unità?



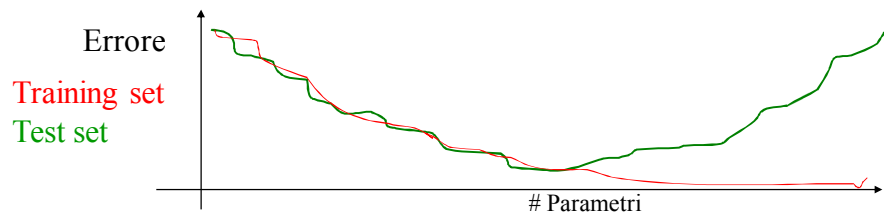
## Scelta empirica - cross-validation



Cross-Validation - Errore sull'insieme di training = Errore sull'insieme di test.

*Si vuole evitare che il modello si specializzi troppo sui pattern di training e non sia in grado di interpolare correttamente.*

*Il numero di parametri viene aumentato fino a quando **entrambi** gli errori diminuiscono.*



## Riassunto



- Classificazione
- Analisi errore