Python per il Calcolo Scientifico

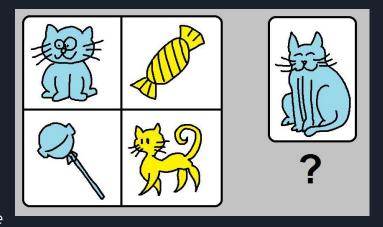
Angelo Cardellicchio

Classificazione

- Cosa è la classificazione?
- Un approccio supervisionato
- Train, test e cross-validazione
- Output, label e classi
- Metriche (alcuni esempi)
- Algoritmi (giusto un paio...)

Cosa è la classificazione?

- L'idea è quella di capire se un oggetto appartiene o meno ad una classe.
- Molto dipende dal dominio.
- Ad esempio: l'immagine è o no un gatto?
 - O anche: l'immagine è blu o gialla?
- Il classificatore è, in pratica, una funzione matematica.
 - Questa funzione non sempre è lineare (anzi, spesso non lo è!)



Un approccio supervisionato



Esperto di dominio

Train, test e cross-validazione

	Sentence structure	Parents	Sister	Friend	Cousins	Grand-	Aunts/	Other	Total
					75.00	parents	uncles	children	
	<pre><person> is your <relationship></relationship></person></pre>	2	1	1	2	4	4	0	14
	rou nave snumbers stelationarilys (a)	-				_	-	•	-
	<pre><person> is a woman/man/girl/boy</person></pre>	2	1	1	2	4	4	6	20
	<pre><person> is a <pre><pre><pre><pre><pre><pre><pre><pre></pre></pre></pre></pre></pre></pre></pre></pre></person></pre>	2	0	0	0	4	4	0	10
	<pre><person> goes to the school/kindergarten</person></pre>	0	1	1	2	0	0	3	7
	CHARGERT 2 THAC IS CHARTON	-							
	<pre><person> does not have a <noun></noun></person></pre>	2	1	1	0	0	0	0	4
	Your <relationship>'s name is <name></name></relationship>	2	1	1	0	0	0	0	4
4	persone lines moure (s)		ž.	Ĉ.	Ü	· ·	-	-	-
	<pre><person> likes to <verb></verb></person></pre>	4	2	2	0	0	0	0	8
	<pre><person> does not like <noun>(s)</noun></person></pre>	4	2	2	0	0	0	0	8
	<pre><person> does not like to <verb></verb></person></pre>	4	2	2	0	0	0	0	8
3	<pre><person> is <number> years old</number></person></pre>	2	1	1	2	4	4	6	20
18	<pre><person>'s favorite <noun> is</noun></person></pre>	4	2	2	4	0	0	0	12
,	The <pre><pre>complement></pre></pre>	-	*	-	-	-	-	-	8
;	You like <noun>(s)</noun>	-	-	-		(=)		-	6
,	You like to <verb></verb>		-			-	-	-	6
	You do not like <noun>(s)</noun>					-	-		6
)	You do not like to <verb></verb>		2	121	4		120	12	6
	Your favorite <noun> is</noun>		œ			(=)	(-)		4
	You have <number> <noun>(s)</noun></number>		4						4
	You do not have a <noun></noun>		-		-	-	-	-	4
3	Women/men/girls/boys like to								7
					^		A-COLO		
	progressive)	SWE			2000				
5	How-to sentences		-					-	28
6	Other sentences	-	4		-		-	-	4
								Total	225

Output, label e classi

Gatto o cane?



 Se le classi sono solo due, o addirittura una sola, e la domanda è se il dato appartiene o meno alla classe, abbiamo la classificazione binaria. Altrimenti, è multiclasse.

Output, label e classi

Passengerld	Survive	ţ	Pclass	Name	Sex	Ī	\ge	SibSp	Parch	Ticket	Fare	Cabin	Embarked
1	0		3	Braund, Mr. Owen Harris	male		2.0	1	0	A/5 21171	7.2500	NaN	s
2	1		1	Cumings, Mrs. John Bradley (Florence Briggs Th	female	9 3	8.0	1	0	PC 17599	71.2833	C85	С
3	1		3	Heikkinen, Miss. Laina	female	9	6.0	0	0	STON/O2. 3101282	7.9250	NaN	s
4	1		1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	9 3	5.0	1	0	113803	53.1000	C123	s
5	0		3	Allen, Mr. William Henry	male		5.0	0	0	373450	8.0500	NaN	s

Più label binarie -> multi-label

Più label non binarie, più classi: multi-task o multi-output

Metriche (alcuni esempi)

• **Accuracy**: quante volte il classificatore riesce ad individuare correttamente la classe di un campione?

$$acc(y, \hat{y}) = \frac{1}{n} \sum_{i=0}^{n-1} 1(\hat{y}_i = y_i)$$

• Recall: quante volte il classificatore riesce ad individuare gli elementi appartenenti ad una classe rispetto al numero totale degli appartenenti alla stessa?

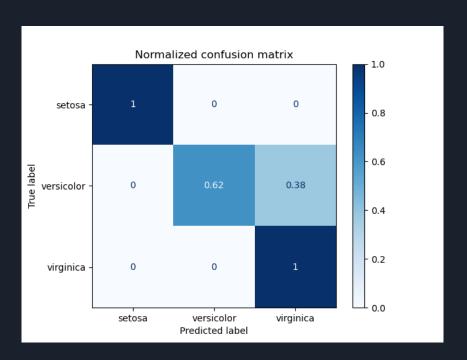
$$R = \frac{TP}{TP + FN}$$

• **Precision**: quante volte il classificatore riesce a non assegnare una classe a campioni appartenenti ad un'altra?

$$P = \frac{TP}{TP + FP}$$

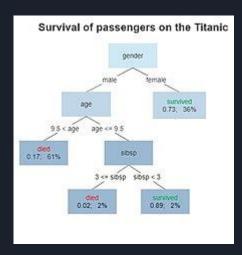
Metriche

Sulla diagonale ci sono i *true positive*, nelle parti triangolari invece gli errori compiuti dal classificatore.



Algoritmi (giusto un paio)

- Alberi decisionali
- Creano un "albero di regole" a partire dai dati.



Algoritmi (giusto un paio)

- Support Vector Machine
- Individuano un iperpiano di separazione tra due classi.

