

Esercitazione 1 09/10/2023

Notazione e ripasso

Avendo una serie di input booleani, con una funzione $f(x)$ nascosta che porta al risultato booleano, il machine learning prova a trovare f , ma in realtà trova $h(x)$ che è l'ipotesi.

Se $f(x) = h(x)$ allora l'ipotesi è consistente.

H è lo **spazio delle ipotesi**, all'interno del quale è presente l'ipotesi che viene trovata dall'algoritmo utilizzato. Bisogna trovare l'ipotesi che approssima la funzione al meglio.

Se per ogni input ci sono 2 parametri booleani, A e B , il numero delle istanze è 4, ovvero tutte le possibili combinazioni.

Se il numero dell'insieme delle istanze è minore del massimo numero di combinazioni, esso rappresenta un **concetto** (per esempio 3 input quando il massimo è 4 input).

Il problema nel trovare l'ipotesi ottimale è il **bias** perché c'è sempre una predilezione delle informazioni, anche se involontaria dello sviluppatore.

Dati due campi dell'input A e B , lo spazio delle istanze è X e la sua cardinalità $|X| = |A| \times |B|$, ed è un sottoinsieme dello spazio delle istanze $|C| = 2^x$. (forse x è il numero di campi dell'input).

Esercizio loans

Formulazione:

input: x (customer application)

output: y (good/bad customer?)

target function: $f : X \rightarrow Y$ (ideal credit approval formula)

data: $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ (historical records)

hypotesis: $g : X \rightarrow Y$ (formula to be used)

Regression: Credit line (dollar amount)

Linear regression output: $h(x)$

=

$$\text{Linear regression output: } h(\mathbf{x}) = \sum_{i=0}^d w_i x_i = \mathbf{w}^T \mathbf{x}$$

dove w sono i pesi.

Example of hypothesis space for categorical data

(screenshots slides)

(screenhots most general hypothesis most specific hypothesis)

Nella prima vanno tutti bene, nella seconda nessuno.

Example: consistency

Esempio che può essere in esame

Hypothesis $h = \langle \text{Sunny, Warm, ?, Strong, ?, ?} \rangle$

Data	Example	Sky	AirTemp	Humidity	Wind	Water	Forecast	EnjoySport
	1	Sunny	Warm	Normal	Strong	Warm	Same	Yes
	2	Sunny	Warm	High	Strong	Warm	Same	Yes
	3	Rainy	Cold	High	Strong	Warm	Change	No
	4	Sunny	Warm	High	Strong	Cool	Change	Yes

Task for each example $(x, c(x))$ in D , we will evaluate $h(x)$ equals $c(x)$.

Si chiede se l'ipotesi sia consistente con il training set.

Applico la prima istanza all'ipotesi, e vado a vedere attributo per attributo cosa succede.

Quindi quest'ipotesi è consistente con gli esempi della tabella.

L'ultima colonna della tabella rappresenta il "target" del confronto con l'ipotesi. Il target è Yes o No, ovvero 1 o 0.

Relazione tra ipotesi

Si può dare un ordine alle ipotesi, da quelle più generiche a quelle più specifiche. Non si riesce però ad ordinarle tutte, ma solo alcune. è un'ordinamento parziale.

► Let h_j and h_k be boolean-valued functions defined over X .

► h_j is more-general-than-or-equal-to h_k (written $h_j \geq_g h_k$) iff

$$(\forall x \in X) (h_k(x) = 1 \rightarrow h_j(x) = 1).$$

$h_k(x)$ è più specifica di $h_j(x)$, $h_k(x)$ è un sottoinsieme di $h_j(x)$

Numero di ipotesi

Let's assume the following:

- A collection of attributes, e.g., A_1, A_2, A_3, A_4
- Hypotheses (i.e., Hypotheses space \mathcal{H}) = Conjunction of constraints (on attributes)

- indicate by a "?" that any value is acceptable for this attribute,
- specify a single required value (e.g., *Warm*) for the attribute, or
- indicate by a " \emptyset " that no value is acceptable.

- 2 hypotheses are syntactically different when they are (in fact) two different combinations of possible symbols for each attribute

- hypothesis space cardinality : number of different combinations of possible symbols for each attribute

Quindi nel caso EnjoySports:

► Possible values:

Sky	Sunny, Cloudy, Rainy
Temp	Warm, Cold
Humid	Normal, High
Wind	Strong, Weak
Water	Warm, Cool
Forecast	Same, Change

Attributes:	Sky	AirTemp	Humidity	Wind	Water	Forecast	Total
# Values:	3	*	2	*	2	*	2 = 96
+ 1 ("?")	4	*	3	*	3	*	3 = 972
						+ 1 for the "never" hypothesis	+ 1
						Total:	973

+1 perchè tutte le ipotesi che contengono il simbolo di rifiuto sono semanticamente equivalenti.

2 hypotheses with some \emptyset are semantically equivalents

$\langle 1, 0, \emptyset, 0 \rangle$

$\langle 1, 0, 1, \emptyset \rangle$