

# Ex 1. Concept learning – How many Instances?

Domanda 1

Risposta salvata

Punteggio max.:  
2,00

Si consideri un problema di apprendimento in cui la generica istanza sia descrivibile da 3 attributi ciascuno dei quali possa assumere i seguenti valori: A in {0,1,2}, B in {0,1} e C in {0,1,2,3}. Il concetto target T sia infine descritto per mezzo di una funzione in {0,1}.

Quante istanze si possono esprimere in questo problema?

Scegli un'alternativa:

- a.  $(2^2)*(2^3)*(2^4)$
- b. Nessuna alternativa tra quelle proposte risulta corretta
- c.  $3*4*5+1$
- d.  $2^{(2+3+4)}$
- e.  $4*5*6$
- f.  $2^{(2*3)+1}$
- g.  $2*3*4$

[Annulla la scelta](#)

## Ex 1. Concept learning – How many Instances?

$$A \in \{0, 1, 2\}$$

$$B \in \{0, 1\}$$

$$C \in \{0, 1, 2, 3\}$$

$$T \in \{0, 1\}$$

$$\text{count} = |A| * |B| * |C| = 2 * 3 * 4$$

.

## Ex 2. Concept learning – Hypotheses Satisfatibility

- An instance  $x$  satisfies an hypothesis  $h$  iff all the constraints expressed by  $h$  are satisfied by the attribute values in  $x$ .

Domanda 2

Risposta non  
ancora data

Punteggio max.:  
3,00

Contrassegna  
domanda

Modifica  
domanda

Indicare tra le alternative seguenti gli esempi le cui istanze soddisfano l'ipotesi:  $h = < ?, 1, 0, ? >$ . In particolare si consideri la situazione in cui le istanze siano descritte da 4 attributi A,B,C,D che assumono valore in 0,1,2.

Attenzione: vi possono essere 0,1o più alternative corrette. L'inserimento di una risposta errata penalizza una risposta corretta

Scegli una o più alternative:

- a. esempio = (( 1,0,1,2), label= 1)
- b. esempio = (( 1,0,1,2), label= 0)
- c. esempio = (( 1,1,0,1), label= 0)
- d. esempio = (( 2,1,0,2), label= 1)
- e. esempio = (( 2,1,0,1), label= 1)
- f. nessuna corretta tra le alternative

## Ex 2. Concept learning – Hypotheses Satisfiability

- An instance  $x$  satisfies an hypothesis  $h$  iff all the constraints expressed by  $h$  are satisfied by the attribute values in  $x$ .

Domanda 2

Risposta non  
ancora data

Punteggio max.:  
3,00

Contrassegna  
domanda

Modifica  
domanda

Indicare tra le alternative seguenti gli esempi le cui istanze soddisfano l'ipotesi:  $h = < ?, 1, 0, ? >$ . In particolare si consideri la situazione in cui le istanze siano descritte da 4 attributi A,B,C,D che assumono valore in 0,1,2.

Attenzione: vi possono essere 0,1 o più alternative corrette. L'inserimento di una risposta errata penalizza una risposta corretta

Scegli una o più alternative:

- a. esempio = (( 1,0,1,2), label= 1)
- b. esempio = (( 1,0,1,2), label= 0)
- c. esempio = (( 1,1,0,1), label= 0)
- d. esempio = (( 2,1,0,2), label= 1)
- e. esempio = (( 2,1,0,1), label= 1)
- f. nessuna corretta tra le alternative

# Ex 3. Concept learning – Consistent Hypotheses

$$\text{Consistent}(h, D) \equiv (\forall \langle x, c(x) \rangle \in D) \quad h(x) = c(x)$$

Indicare quali tra le seguenti affermazioni risultano corrette in relazione al training set in tabella

Esempi A B C Label

x1 0 1 2 -1

x2 1 1 2 -1

x3 0 0 2 1

(Label negativa: -1; positiva 1)

Scegli una o più alternative:

- a.  $\langle 0,0,? \rangle$  è un ipotesi consistente
- b. Non vi sono affermazioni corrette tra quelle presentate
- c.  $\langle ?,0,0 \rangle$  è un ipotesi consistente
- d.  $\langle 0,?,? \rangle$  è un ipotesi consistente
- e. Non è possibile trovare ipotesi consistenti
- f.  $\langle ?,0,? \rangle$  è un ipotesi consistente
- g.  $\langle 1,1,1 \rangle$  è un ipotesi consistente

# Ex 3. Concept learning – Consistent Hypotheses

$$\text{Consistent}(h, D) \equiv (\forall \langle x, c(x) \rangle \in D) \quad h(x) = c(x)$$

Indicare quali tra le seguenti affermazioni risultano corrette in relazione al training set in tabella

Esempi A B C Label

x1 0 1 2 -1

x2 1 1 2 -1

x3 0 0 2 1

(Label negativa: -1; positiva 1)

Scegli una o più alternative:



a. <0,0,?> è un'ipotesi consistente

b. Non vi sono affermazioni corrette tra quelle presentate

c. <?,0,0> è un'ipotesi consistente

d. <0,?,?> è un'ipotesi consistente

e. Non è possibile trovare ipotesi consistenti



f. <?,0,?> è un'ipotesi consistente

g. <1,1,1> è un'ipotesi consistente

# Ex 4. Concept Learning

## General to specific relationship

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

Domanda 1

Risposta salvata

Punteggio max.:  
2,00

Si individuino, tra le alternative che seguono, quelle in cui sono elencate ipotesi più generali di:

$$h = < ?, 0, 1, 2 >$$

In questo problema gli attributi possono assumere valore in 0,1,2. Attenzione: vi possono essere 0,1 o più alternative corrette. L'inserimento di una risposta errata penalizza una risposta corretta

Scegli una o più alternative:

- a.  $h = < ?, 1, ?, ? >$
- b. nessuna ipotesi indicata è più generale
- c.  $h = < ?, 0, 1, ? >$
- d.  $h = < ?, ?, 1, 2 >$
- e.  $h = < ?, 1, 0, ? >$
- f.  $h = < ?, ?, ?, 0 >$

## Sol. 4) General to specific relationships

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

Domanda 1

Risposta salvata

Punteggio max.:  
2,00

Si individuino, tra le alternative che seguono, quelle in cui sono elencate ipotesi più generali di:

$$h = < ?, 0, 1, 2 >$$

In questo problema gli attributi possono assumere valore in 0,1,2. Attenzione: vi possono essere 0,1 o più alternative corrette. L'inserimento di una risposta errata penalizza una risposta corretta



Scegli una o più alternative:

- a.  $h = < ?, 1, ?, ? >$
- b. nessuna ipotesi indicata è più generale
- c.  $h = < ?, 0, 1, ? >$
- d.  $h = < ?, ?, 1, 2 >$
- e.  $h = < ?, 1, 0, ? >$
- f.  $h = < ?, ?, ?, 0 >$

# Find S

1. Initialise  $h \in H$  to the most specific hypothesis:  $h \leftarrow \langle a_1, \dots, a_n \rangle$ ,  $(\forall i) a_i = 0$ .
2. FOR each positive training instance  $d \in D$ , do:  
    FOR each attribute  $a_i, i = [1..n]$ , in  $h$ , do:  
        IF  $a_i$  is satisfied by  $d$   
        THEN do nothing  
        ELSE replace  $a_i$  in  $h$  so that the resulting  $h' >_g h$ ,  $h \leftarrow h'$ .
3. Output hypothesis  $h$ .

## Ex. 5 - Find-S

Si consideri il problema di apprendere un concetto (Target) attraverso l'algoritmo Find-S. Ogni istanza del problema è caratterizzabile per mezzo di 4 attributi ( $A_1, A_2, A_3, A_4$ ) ciascuno dei quali può assumere un valore pari a 1 oppure a 0. Si tenga presente inoltre che anche il concetto da apprendere è esprimibile per mezzo di una funzione che assume valori in  $\{0,1\}$ . Si evidenzi tra le alternative quella ottenibile dopo la presentazione degli esempi in tabella.

Ex.Numb.	$A_1$	$A_2$	$A_3$	$A_4$	Label
$x_1$	1	0	1	0	1
$x_2$	0	1	0	0	0
$x_3$	1	0	1	1	0
$x_4$	0	0	0	0	1
$x_5$	0	0	1	0	1

Table 1: Training Set

1. Initialise  $h \in H$  to the most specific hypothesis:  $h \leftarrow \langle a_1, \dots, a_n \rangle$ ,  $(\forall i) a_i = 0$ .
2. FOR each positive training instance  $d \in D$ , do:

FOR each attribute  $a_i$ ,  $i = [1..n]$ , in  $h$ , do:

IF  $a_i$  is satisfied by  $d$

THEN do nothing

ELSE replace  $a_i$  in  $h$  so that the resulting  $h' >_g h$ ,  $h \leftarrow h'$ .

3. Output hypothesis  $h$ .

Scegli un'alternativa:

a.  $S = \{?, 0, 0, ?\}$

b.  $S = \{?, ?, ?, ?\}$

c.  $S = \{?, 1, 0, ?\}$

d. Nessuna delle precedenti

e.  $S = \{?, 0, 1, ?\}$

# Ex. 5 - Find S

1. Initialise  $h \in H$  to the most specific hypothesis:  $h \leftarrow \langle a_1, \dots, a_n \rangle$ ,  $(\forall i) a_i = 0$ .
2. FOR each positive training instance  $d \in D$ , do:
  - FOR each attribute  $a_i, i = [1..n]$ , in  $h$ , do:
    - IF  $a_i$  is satisfied by  $d$
    - THEN do nothing
    - ELSE replace  $a_i$  in  $h$  so that the resulting  $h' >_g h$ ,  $h \leftarrow h'$ .
3. Output hypothesis  $h$ .

Ex.Numb.	$A_1$	$A_2$	$A_3$	$A_4$	Label
$x_1$	1	0	1	0	1
$x_2$	0	1	0	0	0
$x_3$	1	0	1	1	0
$x_4$	0	0	0	0	1
$x_5$	0	0	1	0	1

Table 1: Training Set

$$S = \{ \langle \emptyset, \emptyset, \emptyset, \emptyset \rangle \}$$

$$x_1 = (1, 0, 1, 0), t(x_1) = 1 \rightarrow S = \{ \langle 1, 0, 1, 0 \rangle \}$$

$$x_2 = (0, 1, 0, 0), t(x_1) = 0 \rightarrow S = \{ \langle 1, 0, 1, 0 \rangle \}$$

$$x_3 = (1, 0, 1, 1), t(x_1) = 0 \rightarrow S = \{ \langle 1, 0, 1, 0 \rangle \}$$

$$x_4 = (0, 0, 0, 0), t(x_1) = 1 \rightarrow S = \{ \langle ?, 0, ?, 0 \rangle \}$$

$$\rightarrow x_5 = (0, 0, 1, 0), t(x_1) = 1 \rightarrow S = \{ \langle ?, 0, ?, 0 \rangle \}$$

# Ex 6. PERCEPTRON

Domanda 1  
Risposta non ancora data  
Punteggio max.: 6,00

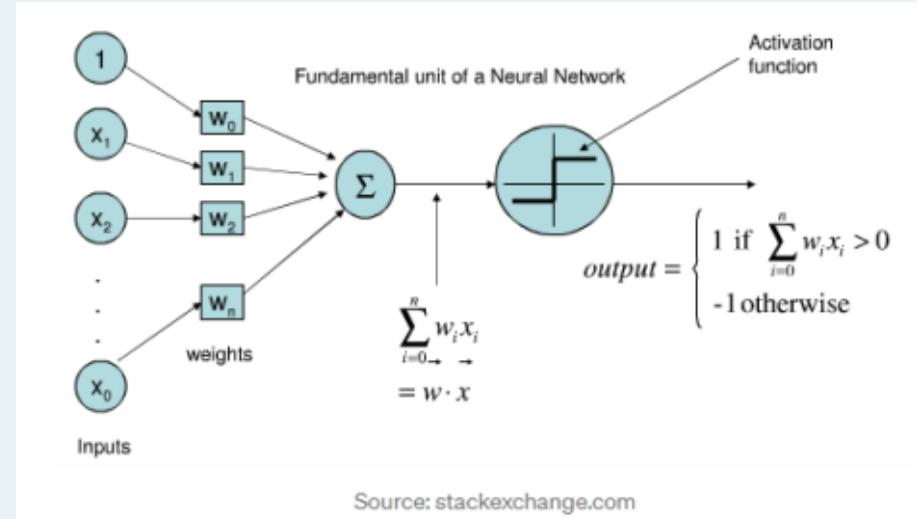
Si riporti il vettore dei pesi ottenibile dopo la presentazione degli esempi in tabella applicando l'algoritmo del Percettrone.

## Training

Esempi	x0	x1	target
1	1	1.5	-1
2	1	0	1

Per quest'esercizio si considerino le seguenti note:

- 1) Il percettrone risponde in (-1 , 1).
- 2) utilizzare la regola di aggiornamento  $w_n \leftarrow w_o + \alpha * (y - \hat{y}) * x$  con:
  - learning rate pari a 0.1
  - L'istanza di ogni esempio del training è:  $x = [x_0, x_1]$ ,
  - target (di ciascuna istanza):  $y$ ,
  - risposta percettrone:  $\hat{y}$
- 3) Pesi iniziali:  $w = [0 , 1 , 1]$
- 4) La prima componente  $w[0]$  dei pesi rappresenta il bias e dovrà essere associata alla prima componente del segnale di ingresso fissata sistematicamente pari a 1.
- 5) Attenzione: non si richiede di trovare una separazione nel caso esista, ma semplicemente di presentare 2 esempi di input e poi di indicare i pesi che si ottengono
- 6) SI PREGA DI RIPORRE ATTENZIONE AI CONTI !!!



## Ex 4. PERCEPTRON

Si riporti il vettore dei pesi ottenibile dopo la presentazione degli esempi in tabella applicando l'algoritmo del Percettrone.

### Training

Esempi	x0	x1	target
1	1.5	1.5	-1
2	1	0	1
3	3	0	-1

Per quest'esercizio si considerino le seguenti note:

- 1) Il percettrone risponde in  $(-1, 1)$ .
- 2) ATTENZIONE: si utilizzi la seguente regola di aggiornamento dei pesi

$$w_i \leftarrow w_i + \alpha * (y - \hat{y}) * x$$

- learning rate pari a 0.1
  - L'istanza di ogni esempio del training è:  $x = [x_0, x_1]$ ,
  - target (di ciascuna istanza):  $y$ ,
  - risposta percettrone:  $\hat{y}$
- 3) Pesi iniziali:  $w = (0, 1, 0)$
  - 4) La prima componente  $w[0]$  dei pesi rappresenta il bias e dovrà essere associata alla prima componente del segnale di ingresso fissata sistematicamente pari a 1.
  - 5) Attenzione: non si richiede di trovare una separazione nel caso esista, ma semplicemente di presentare gli esempi di input (uno ad uno), utilizzare la regola e poi di indicare i pesi finali che si ottengono
  - 6) SI PREGA DI RIPORRE ATTENZIONE AI CONTI !!!

La risposta a questo esercizio (i pesi aggiornati) deve essere data con una stringa di caratteri nell'apposito spazio. In modo particolare si usi il formato seguente come esempio:

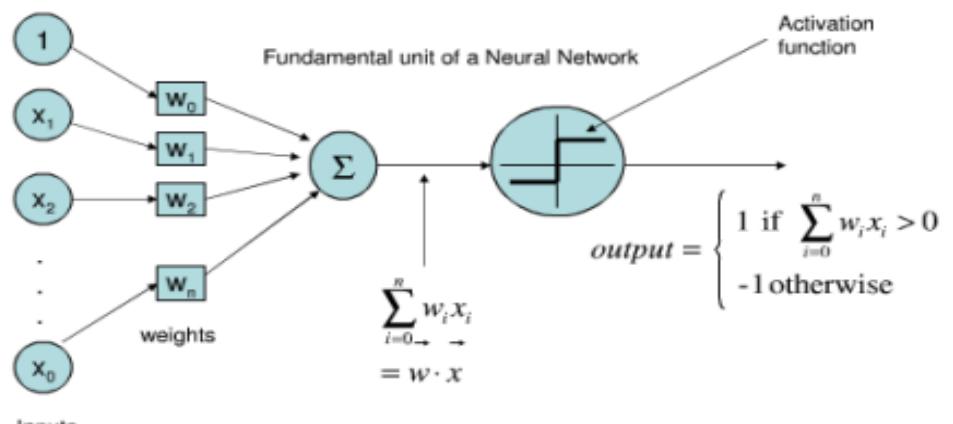
0.1,0.2,-3

Nell'esempio, si noti che la virgola separa le varie componenti del vettore dei pesi, ovvero

La prima componente  $w[0]$  del vettore  $w$  è positiva ed è pari a **0.1**,

La seconda componente  $w[1]$  del vettore  $w$  è positiva ed è pari a **0.2**,

La terza componente  $w[2]$  del vettore  $w$  è negativa ed è pari a **-3**.



Source: stackexchange.com

# Ex 6: perceptron

Esempi	x0	x1	target
1	1.5	1.5	-1
2	1	0	1
3	3	0	-1

Example 1:  $\mathbf{x}_1 = (1.5, 1.5)$ ,  $t(\mathbf{x}_1) = -1$ ;

Input:  $\mathbf{x} = (1, 1.5, 1.5)$

Learning rate:  $\eta = 0.1$

Weights:  $\mathbf{w} = (0, 1, 0)$

$$\langle \mathbf{w}, \mathbf{x} \rangle = (0, 1, 0) \begin{pmatrix} 1 \\ 1.5 \\ 1.5 \end{pmatrix} = 0 + 1.5 + 0 \geq 0$$

$$\Rightarrow \text{sign} (\langle \mathbf{w}, \mathbf{x} \rangle) = +1[\hat{y}] \neq t(\mathbf{x}_1)[y]$$

CHANGE w !!!

# Ex 6: perceptron

Esempi	x0	x1	target
1	1.5	1.5	-1
2	1	0	1
3	3	0	-1

Example 1:  $\mathbf{x}_1 = (1.5, 1.5)$ ,  $t(x_1) = -1$ ;

Input:  $\mathbf{x} = (1, 1.5, 1.5)$

Learning rate:  $\eta = 0.1$

Weights:  $\mathbf{w} = (0, 1, 0)$

Perceptron updating rule:  $\mathbf{w}_{\text{new}} \leftarrow \mathbf{w}_{\text{old}} + \eta(y - \hat{y})\mathbf{x}$

We have,  $\mathbf{w}_{\text{new}} \leftarrow \mathbf{w}_{\text{old}} + 0.1(-2)\mathbf{x}$ , i.e.,

$$\begin{aligned}\mathbf{w}_{\text{new}}[0] &\leftarrow \mathbf{w}_{\text{old}}[0] + 0.1 * (-2) * \mathbf{x}[0] \\ &= 0 + (0.1)(-2)(1) = -0.2\end{aligned}$$

$$\begin{aligned}\mathbf{w}_{\text{new}}[1] &\leftarrow \mathbf{w}_{\text{old}}[1] + 0.1 * (-2) * \mathbf{x}[1] \\ &= 1 + (0.1)(-2)(1.5) = 0.7\end{aligned}$$

$$\begin{aligned}\mathbf{w}_{\text{new}}[2] &\leftarrow \mathbf{w}_{\text{old}}[2] + 0.1 * (-2) * \mathbf{x}[2] \\ &= 0 + (0.1)(-2)(1.5) = -0.3\end{aligned}$$

Then,  $\mathbf{w}_{\text{new}} = (-0.2, 0.7, -0.3)$

# Ex 6: perceptron

Esempi	x0	x1	target
1	1.5	1.5	-1
2	1	0	1
3	3	0	-1

Example 2:  $\mathbf{x}_1 = (1, 0)$ ,  $t(x_2) = 1$ ;

Input:  $\mathbf{x} = (1, 1, 0)$

Learning rate:  $\eta = 0.1$

Weights:  $\mathbf{w} = (-0.2, 0.7, -0.3)$

$$\langle \mathbf{w}, \mathbf{x} \rangle = (-0.2, 0.7, -0.3) \begin{pmatrix} 1 \\ (1) \\ 0 \end{pmatrix} = -0.2 + 0.7 + 0 \geq 0$$

$$\Rightarrow \text{sign}(\langle \mathbf{w}, \mathbf{x} \rangle) = +1[\hat{y}] = t(\mathbf{x}_2)[y]$$

$\mathbf{w}$  unchanged.

# Ex 6: perceptron

Esempi	x0	x1	target
1	1.5	1.5	-1
2	1	0	1
3	3	0	-1

Example 3:  $\mathbf{x}_3 = (3, 0)$ ,  $t(x_3) = -1$ ;

Input:  $\mathbf{x} = (1, 3, 0)$

Learning rate:  $\eta = 0.1$

Weights:  $\mathbf{w} = (-0.2, 0.7, -0.3)$

We have,

$$\langle \mathbf{w}, \mathbf{x} \rangle = (-0.2, 0.7, -0.3) \begin{pmatrix} 1 \\ 3 \\ 0 \end{pmatrix} = -0.2 + 2.1 \geq 0$$

$$\Rightarrow \text{sign}(\langle \mathbf{w}, \mathbf{x} \rangle) = +1 \neq t(\mathbf{x}_2)$$

CHANGE w!

# Ex 6: perceptron

Esempi	x0	x1	target
1	1.5	1.5	-1
2	1	0	1
3	3	0	-1

Example 3:  $\mathbf{x}_3 = (3, 0)$ ,  $t(x_3) = -1$ ;

Input:  $\mathbf{x} = (1, 3, 0)$

Learning rate:  $\eta = 0.1$

Weights:  $\mathbf{w} = (-0.2, 0.7, -0.3)$

Perceptron update rule:  $\mathbf{w}_{\text{new}} \leftarrow \mathbf{w}_{\text{old}} + \eta(y - \hat{y})\mathbf{x}$

We have,  $\mathbf{w}_{\text{new}} \leftarrow \mathbf{w}_{\text{old}} + 0.1(-2)\mathbf{x}$ , i.e.,

$$\begin{aligned}\mathbf{w}_{\text{new}}[0] &\leftarrow \mathbf{w}_{\text{old}}[0] + 0.1 * (-2) * \mathbf{x}[0] \\ &= -0.2 + (0.1)(-2)(1) = -0.4\end{aligned}$$

$$\begin{aligned}\mathbf{w}_{\text{new}}[1] &\leftarrow \mathbf{w}_{\text{old}}[1] + 0.1 * (-2) * \mathbf{x}[1] \\ &= 0.7 + (0.1)(-2)(3) = 0.1\end{aligned}$$

$$\begin{aligned}\mathbf{w}_{\text{new}}[2] &\leftarrow \mathbf{w}_{\text{old}}[2] + 0.1 * (-2) * \mathbf{x}[2] \\ &= -0.3 + (0.1)(-2)(0) = -0.3\end{aligned}$$

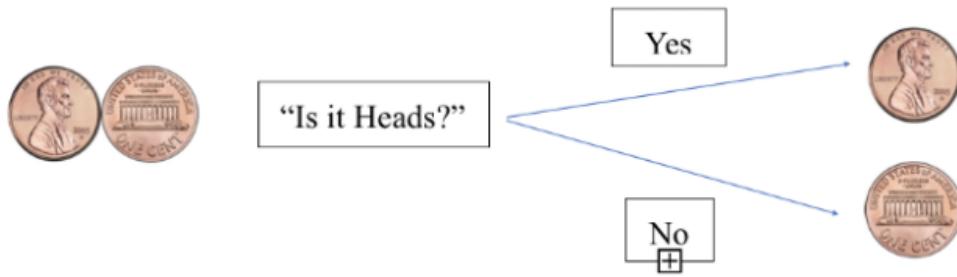
Then,  $\mathbf{w}_{\text{new}} = (-0.4, 0.1, -0.3)$

# Information Gaining & Entropy

Shannon entropy

$$\text{Entropy}(S) = - \sum_i p_i \log_2 p_i$$

# Ex 7. Entropy: A fair coin flip



Originally Published On <https://www.setzeus.com/public-blog-post/intro-to-information-theory>

Shannon entropy

$$\text{Entropy}(S) = - \sum_i p_i \log_2 p_i$$

$P(\text{heads}) = P(\text{tails}) = 1/2$ . (This is called a Bernoulli distribution.)

How much information can we get from one coin flip?

$$H(P) = \sum_x P(x) \cdot \log(1/P(x)).$$

In our case, this is

$$H(P) = P(\text{heads}) \cdot \log(1/P(\text{heads})) + P(\text{tails}) \cdot \log(1/P(\text{tails})).$$

We can substitute these probabilities to find

$$H(P) = 1/2 \cdot \log(2) + 1/2 \cdot \log(2) = 1/2 + 1/2 = 1,$$

at least if we take log to base two.

So the entropy is one bit.

## Ex. 7 Entropy of Bernoulli RVs

- **Bernoulli distribution** is a distribution of a single binary random variable

$$\begin{aligned} p(x = 1 \mid \theta) &= \theta \\ p(x = 0 \mid \theta) &= 1 - \theta \end{aligned}$$

$$\text{Bern}(x \mid \theta) = \theta^x (1 - \theta)^{1-x}$$

- What about the entropy?

$$\begin{aligned} \mathbf{H}[x] &= - \sum_{x \in \{0,1\}} p(x \mid \theta) \ln p(x \mid \theta) \\ &= -p(x = 0 \mid \theta) \ln p(x = 0 \mid \theta) - p(x = 1 \mid \theta) \ln p(x = 1 \mid \theta) \\ &= -(1 - \theta) \ln (1 - \theta) - \theta \ln \theta \end{aligned}$$

# Ex 8. Entropy & IG

Si risponda alle domande che seguono scegliendo l'opzione corretta in riferimento al training in tabella.

## Esempi A B T

x1 0 1 0

x2 1 1 1

x3 1 0 0

x4 0 0 1

Domande:

1)  $H(T | A=0) =$

2)  $H(T | B) =$

5)  $IG(T ; A) =$

6)  $IG(T ; B) =$

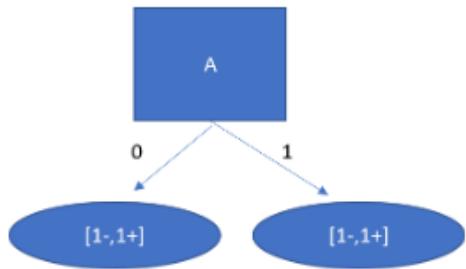
## Ex 8. Entropy & IG

Example	A	B	T
$x_1$	0	1	0
$x_2$	1	1	1
$x_3$	1	0	0
$x_4$	0	0	1

We have:  $P_T = [1/2, 1/2]$ , then:

$$\begin{aligned}H[P_T] &= -P_T(0) \log(P_T(0)) - P_T(1) \log(P_T(1)) \\&= -1/2 \log(1/2) - 1/2 \log(1/2) = 1\end{aligned}$$

# Ex 8. Entropy & IG



Example	A	B	T
$x_1$	0	1	0
$x_2$	1	1	1
$x_3$	1	0	0
$x_4$	0	0	1

$$H[T|A] = P_A(0)H[T|A=0] + P_A(1)H[T|A=1], \text{ where}$$

$$\begin{aligned} H[T|A=0] &= -P_{T|A}(0|0)\log(P_{T|A}(0|0)) - P_{T|A}(1|0)\log(P_{T|A}(1|0)) \\ &= -1/2\log(1/2) - 1/2\log(1/2) = 1 \end{aligned}$$

$$\begin{aligned} H[T|A=1] &= -P_{T|A}(0|1)\log(P_{T|A}(0|1)) - P_{T|A}(1|1)\log(P_{T|A}(1|1)) \\ &= -1/2\log(1/2) - 1/2\log(1/2) = 1 \end{aligned}$$

Moreover, we have:  $P_A(0) = 1/2$ ,  $P_A(1) = 1/2$  then,  $H[T|A] = 1$

Finally,  $IG[T; A] = H[T] - H[T|A] = 1 - 1 = 0$

## Ex 8. Entropy & IG

- $\text{IG}[\text{T}; \text{A}] = H[\text{T}] - H[\text{T}|A] = 1 - 1 = 0$
- $\text{IG}[\text{T}; \text{B}] = H[\text{T}] - H[\text{T}|B] = 1 - 1 = 0$
- $H(\text{T}|A = 0) = 1$
- $H(\text{T}|B) = 1$

# Ex 9. Entropy & IG

Si risponda alle domande che seguono scegliendo l'opzione corretta in riferimento al training in tabella.

Example	D	T
$x_1$	...	0
$x_2$	...	1
$x_3$	...	0
$x_4$	...	1

Nota: L'attributo T esprime il valore di label (es. classe di appartenenza). I valori assunti da D non sono presenti e dovranno essere identificati. In modo particolare:

2) I valori di D dovranno essere scelti in modo che  $IG [ T ; D ]$  sia max

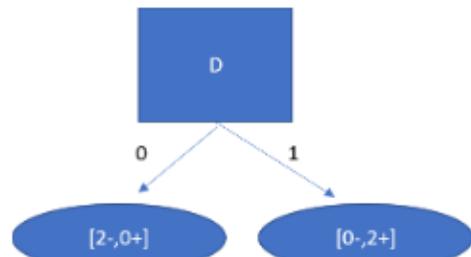
Si consideri inoltre che in tale risposta i valori vengono proposti utilizzando una sequenza come ad esempio (1,0,0,0) dove:

- l'elemento i-esimo della sequenza rappresenta il valore che un esempio di training  $x_i$  assume per il corrispettivo attributo;

Domande: \_\_\_\_\_

4) Valori attributo D = \_\_\_\_\_

# Ex 9. Entropy & IG



Example	D	T	
$x_1$	0	0	-
$x_2$	1	1	-
$x_3$	0	0	-
$x_4$	1	1	-

$$H[T|D] = P_D(0)H[T|D=0] + P_D(1)H[T|D=1], \text{ where}$$

$$\begin{aligned} H[T|D=0] &= -P_{T|D}(0|0)\log(P_{T|D}(0|0)) - P_{T|D}(1|0)\log(P_{T|D}(1|0)) \\ &= -1\log(1) - 0\log(0) = 0 \end{aligned}$$

$$\begin{aligned} H[T|D=1] &= -P_{T|D}(0|1)\log(P_{T|D}(0|1)) - P_{T|D}(1|1)\log(P_{T|D}(1|1)) \\ &= -0\log(0) - 1\log(1) = 0 \end{aligned}$$

Moreover, we have:  $P_D(0) = 1/2$ ,  $P_D(1) = 1/2$  then,  $H[T|D] = 0$

Finally,  $IG[T; D] = H[T] - H[T|D] = 1 - 0 = 1$