

CLASSIFICAZIONE

MODELLI DI SEPARAZIONE



I modelli di separazione che presenteremo sono i seguenti:

- *Artificial Neural Networks*
- *Support Vector Machines*

Nello specifico per quanto riguarda le *Artificial Neural Networks*, data la ricchezza di tale classe di modelli connessionisti, presenteremo nel dettaglio solamente i seguenti modelli di classificazione:

- *Feedforward Neural Networks*
- *Radial Basis Function Networks*

Per quanto riguarda le *Support Vector Machines*, presenteremo i seguenti modelli:

- *Linear hard margin*
- *Linear soft margin*
- *Non-linear*

Feedforward Neural Networks

Ogni *neurone* ha tipicamente un *insieme* di

- *neuroni di input*
- *neuroni di output*



Neurone "i" è *input* del neurone "j". Neurone "j" è *output* per il neurone "i".

Neuroni collegati in modo *orientato* dalla sinapsi che è *associata* ad un *valore reale* (peso).

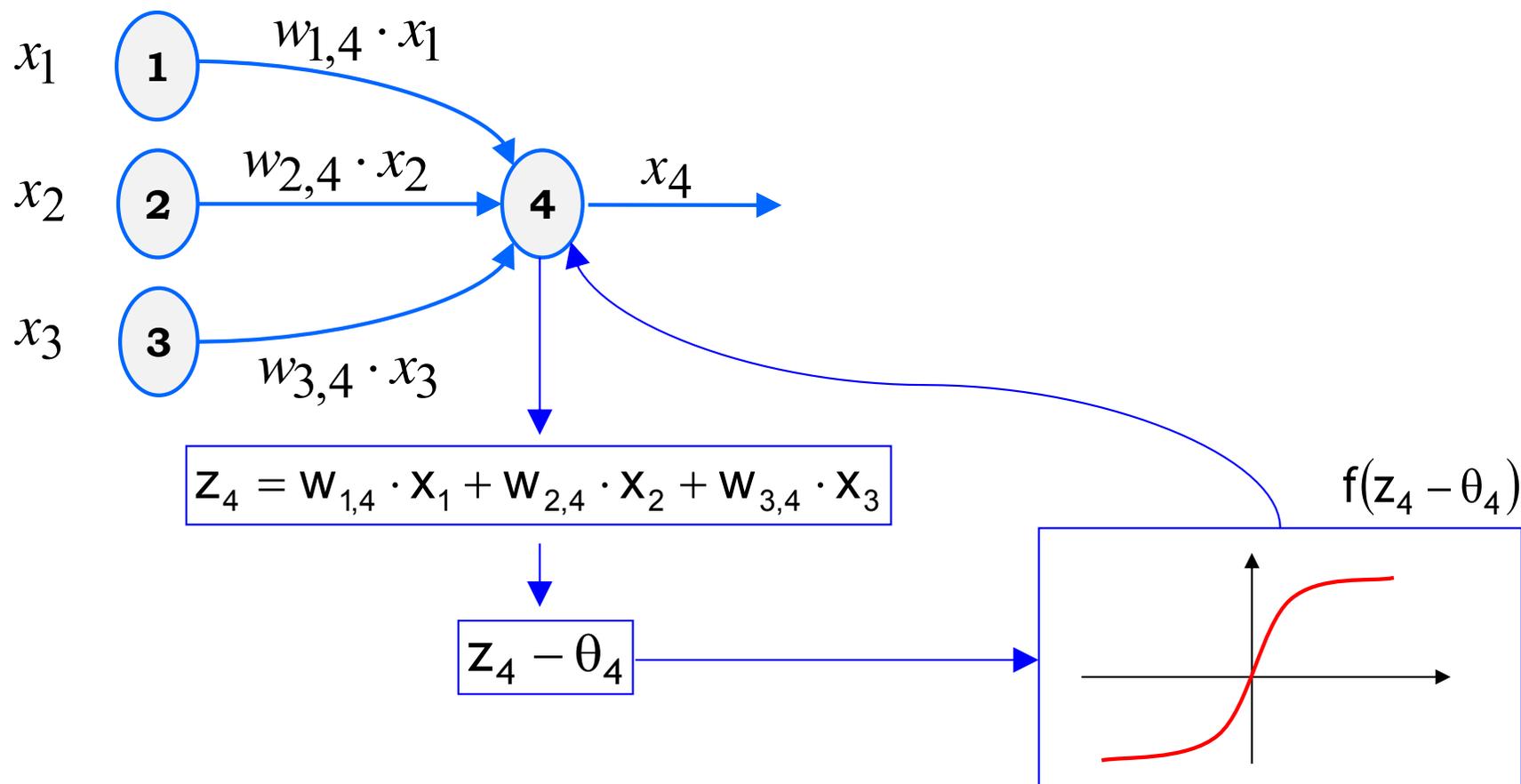


Ogni *neurone* è *caratterizzato* da due elementi:

- *soglia, bias o threshold*
- *funzione di attivazione o di trasferimento*

Ogni neurone:

- riceve segnali da altri neuroni (neuroni di input)
- invia segnali ad altri neuroni (neuroni di output)



Formalmente il *neurone* "j" *calcola* la seguente funzione:

$$y_j = f\left(\sum_{i=1}^n w_{i,j} \cdot x_i - \theta_j\right)$$

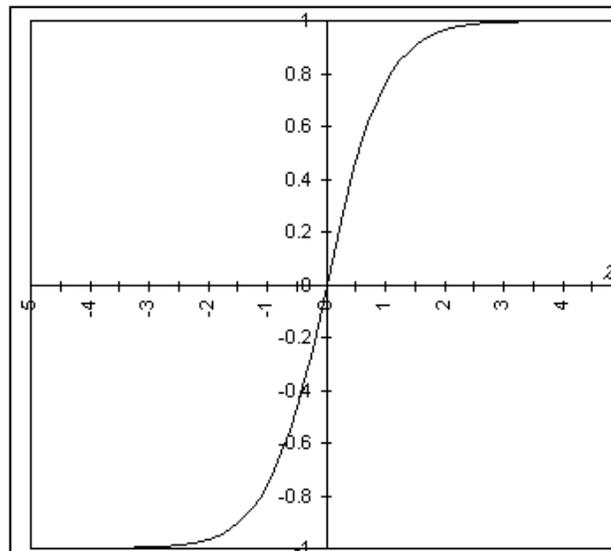
Quali sono le *principali funzioni di attivazione* ?

Tangente Iperbolica

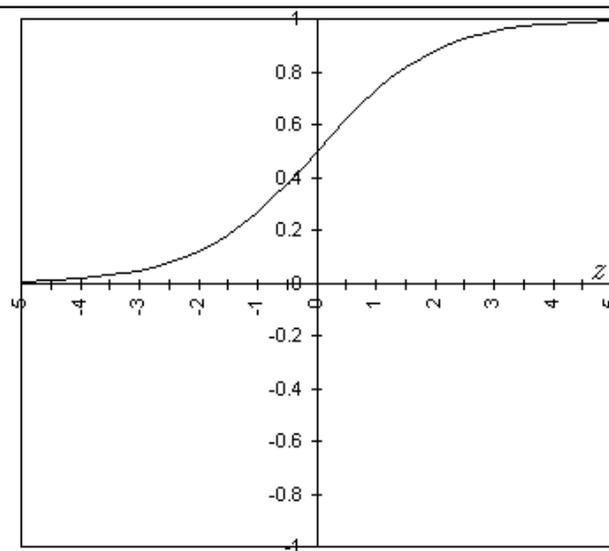
$$f(z) = \frac{\exp(z) - \exp(-z)}{\exp(z) + \exp(-z)}$$

Logistica

$$f(z) = \frac{1}{1 + \exp(-z)}$$

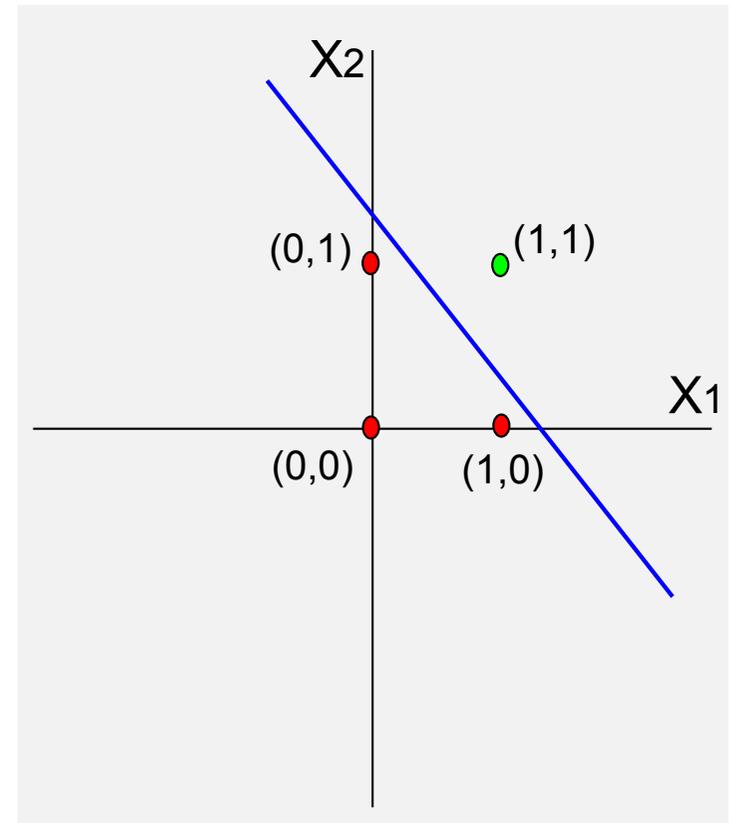
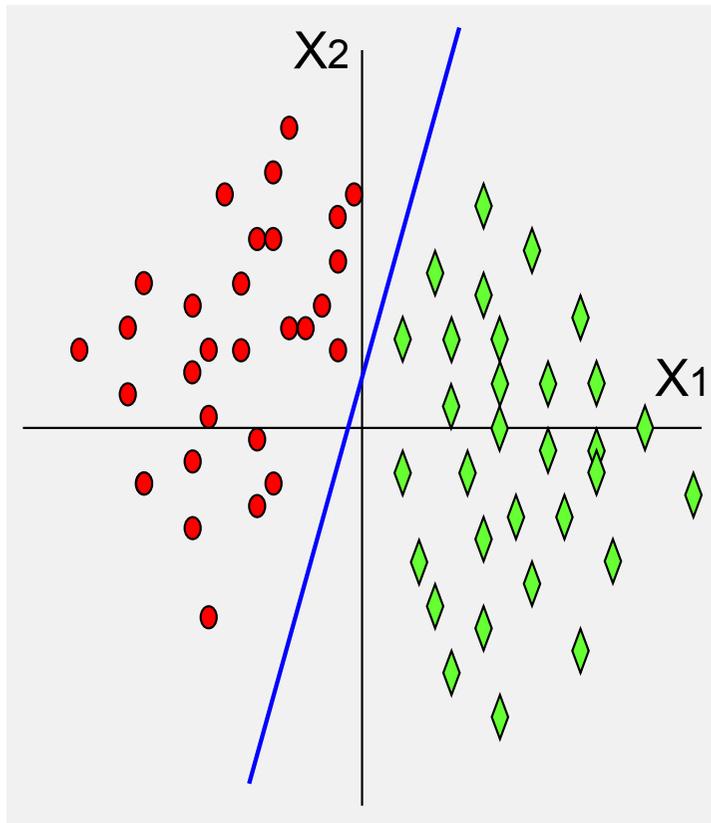


Funzione Tangente Iperbolica



Funzione Logistica

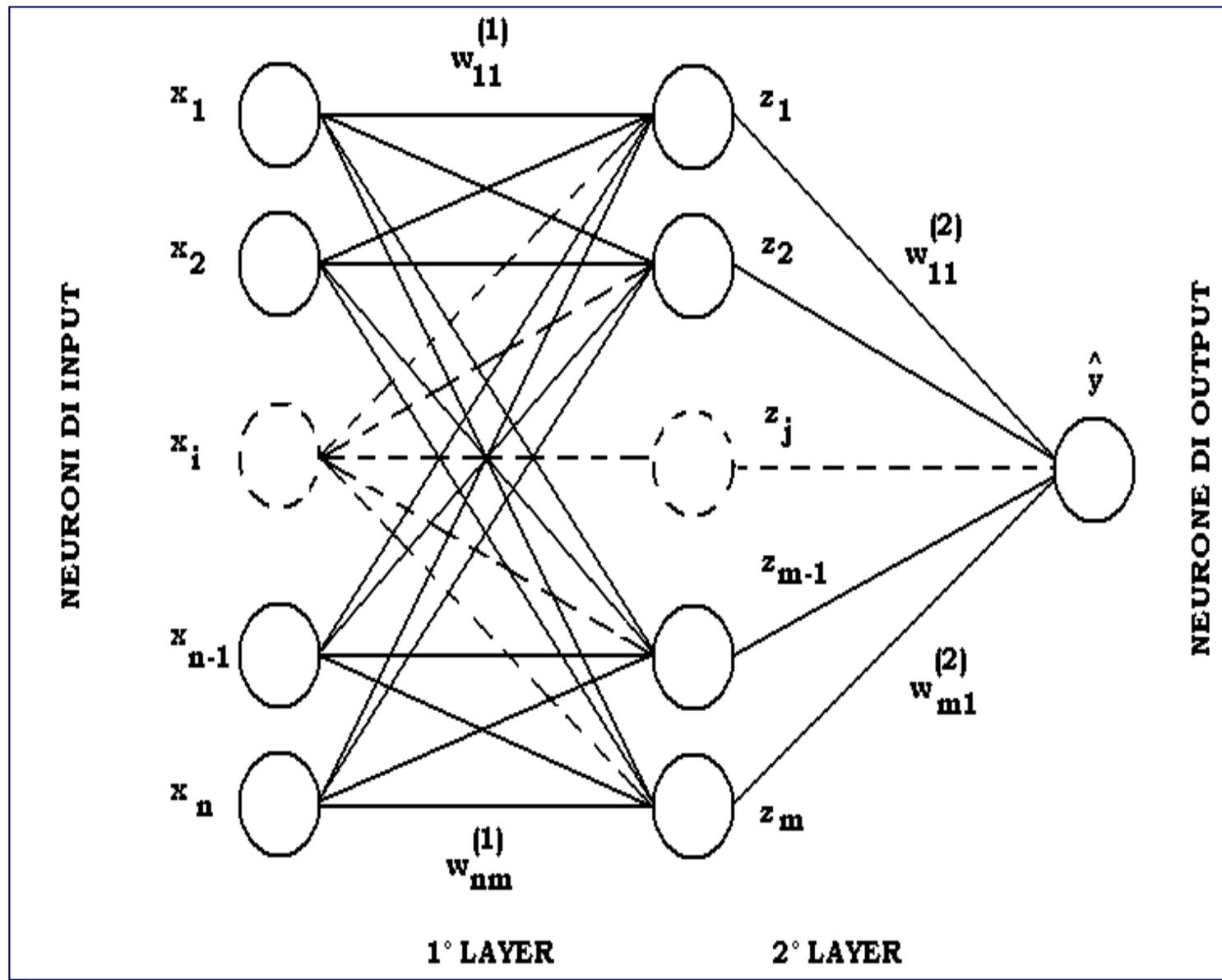
Il perceptrone monostrato *implementa* un *iperpiano* nello *spazio* "n-dimensionale".

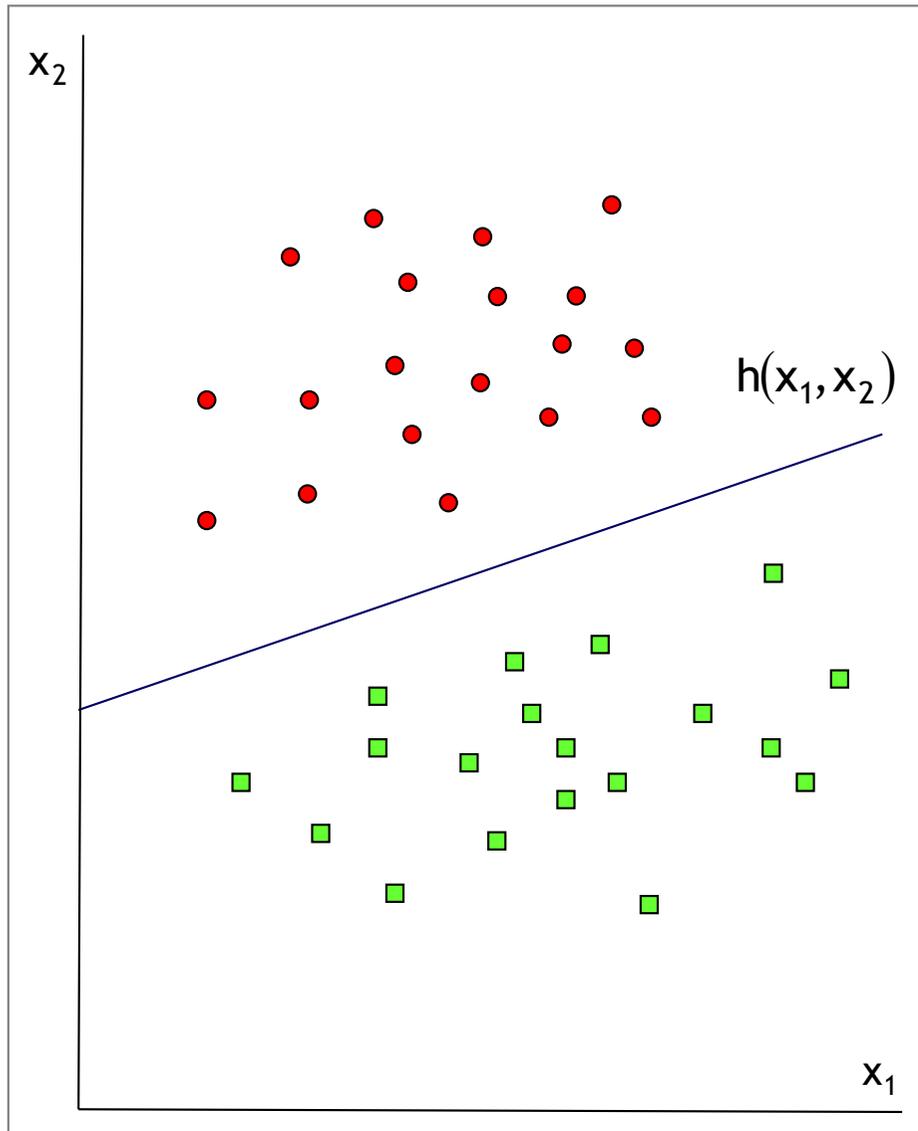


funzione AND

Il *Perceptrone* è in grado di *apprendere qualsiasi funzione* ?

Percettrone Multi-strato o Feedforward Neural Network





Apprendono funzioni lineari con soglia

$$h(\underline{x}) = \text{sign}\{\underline{w} \cdot \underline{x} + b\} = \begin{cases} +1 & \text{se } \underline{w} \cdot \underline{x} + b \geq 0 \\ -1 & \text{altrimenti} \end{cases}$$

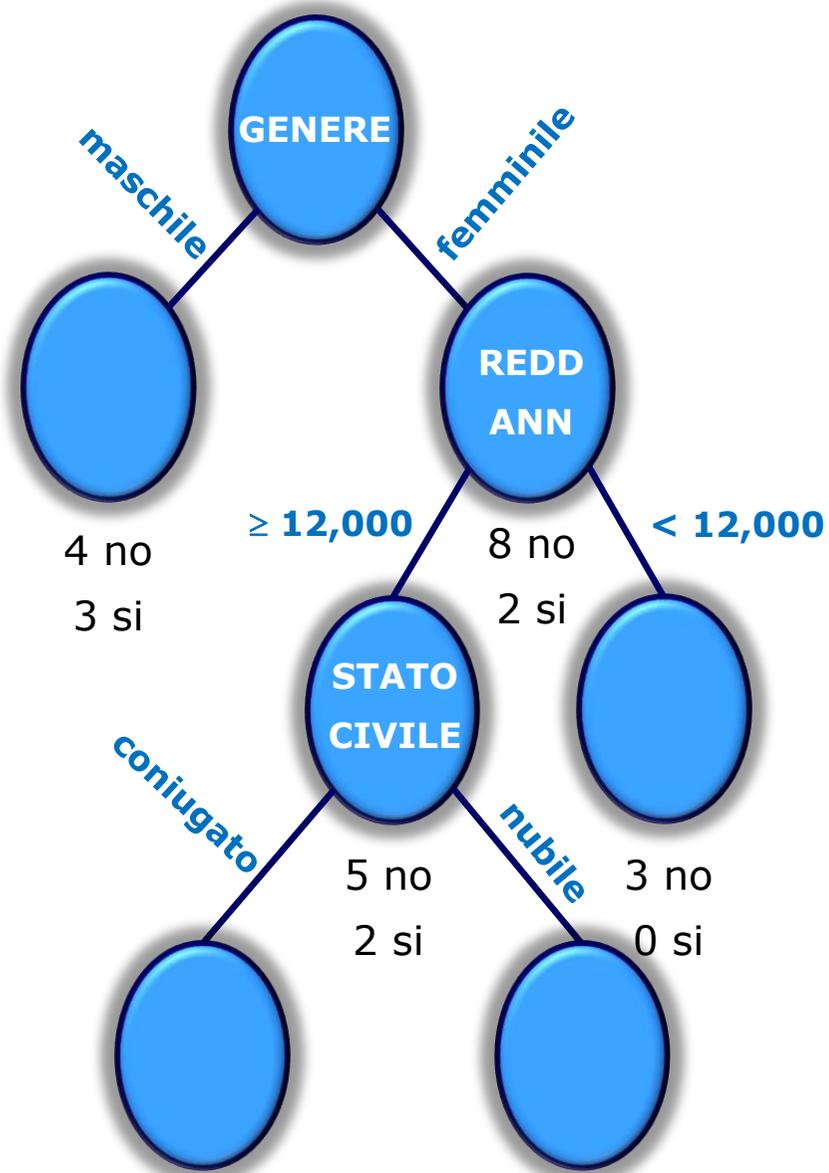
La funzione $h(\bullet)$ ha come argomento un *iperpiano* nello spazio delle variabili esplicative. Ogni istanza viene classificata in base alla parte dell'iperpiano nella quale si trova.

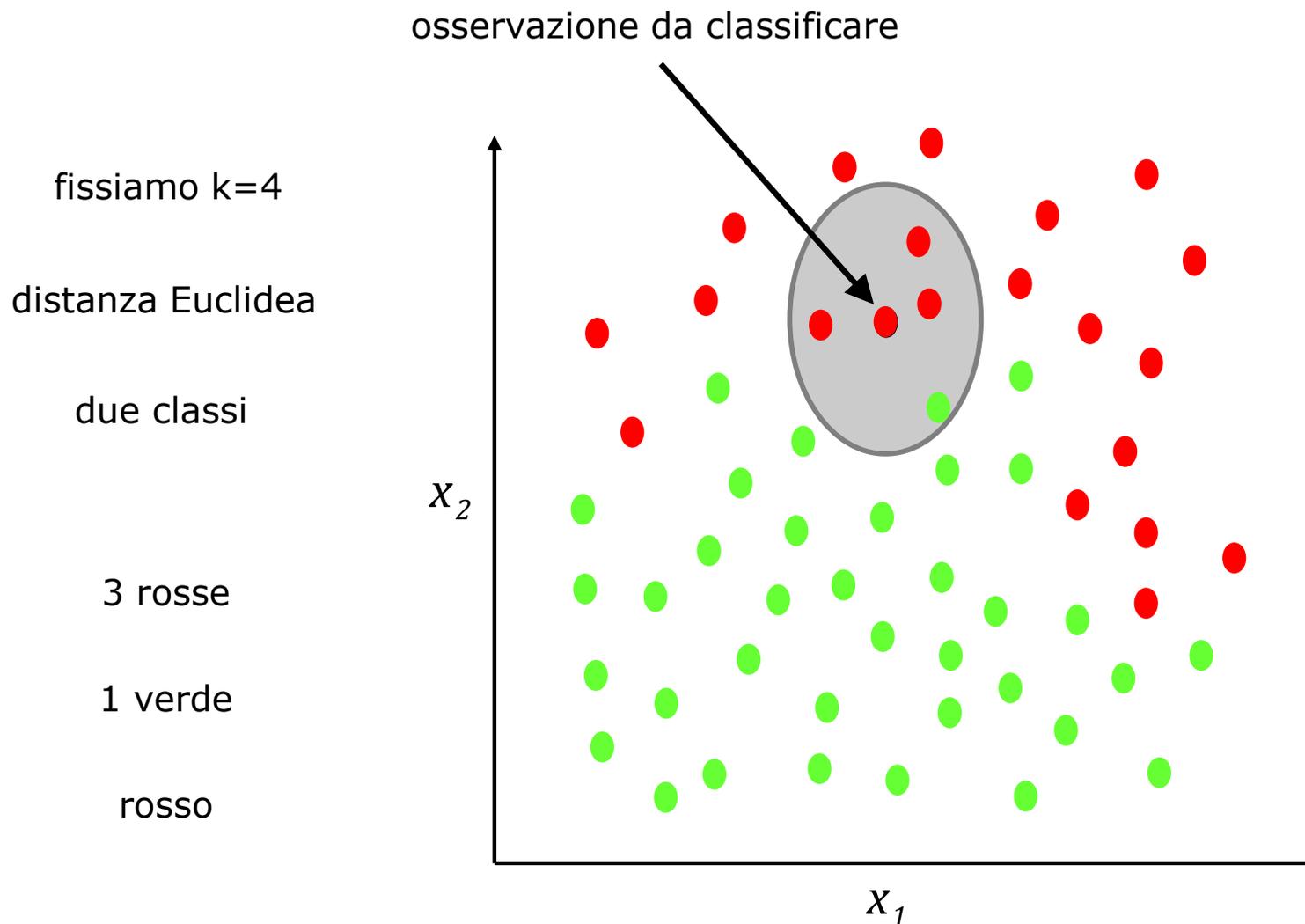
Utilizzano procedure di classificazione basate su schemi algoritmici elementari ed intuitivi.

A questa categoria appartengono:

- **Nearest Neighbor**, basati sulla nozione di distanza tra le osservazioni
- **Alberi di classificazione**, adottano schemi divide and conquer per indurre raggruppamenti di osservazioni quanto più possibile omogenee rispetto alla specifica classe target presa in considerazione
- **Random Forest**, sfruttano lo schema degli alberi di classificazione per sviluppare modelli efficienti ed efficaci combinando tra loro diverse previsioni

GENERE	ETA	PROVINCIA	REGIONE	REDD ANN	STATO CIVILE	EVASORE
maschile	32	VA	Lombardia	20,000 €			celibe	no
femminile	45	AQ	Abruzzo	13,500 €			coniugato	si
femminile	21	RM	Lazio	11,600 €			nubile	no
femminile	62	RM	Lazio	15,350 €			nubile	no
maschile	68	RC	Calabria	10,945 €			divorziato	no
maschile	19	AN	Marche	10,233 €			celibe	no
maschile	24	LT	Lazio	10,450 €			coniugato	si
femminile	22	VI	Veneto	11,567 €			coniugato	no
femminile	29	NO	Piemonte	16,350 €			nubile	no
maschile	52	FI	Toscana	11,245 €			nubile	si
femminile	34	MI	Sicilia	13,450 €			coniugato	no
femminile	33	MI	Basilicata	7,500 €			coniugato	no
femminile	55	TN	Trentino	13,450 €			coniugato	si
maschile	39	FR	Lazio	11,590 €			celibe	si
femminile	55	MI	Lombardia	23,500 €			coniugato	no
maschile	27	MI	Lombardia	35,800 €			celibe	no
femminile	31	AQ	Abruzzo	14,750 €			coniugato	no





È possibile pesare il voto di ogni osservazione in modo inversamente proporzionale alla distanza.

I modelli probabilistici risolvono il problema della classificazione supervisionata tramite l'utilizzo della seguente probabilità condizionata

$$P(Y | \underline{X})$$

dove per il momento assumiamo che Y sia una variabile binaria mentre con \underline{X} indichiamo un vettore binario n -dimensionale. Inoltre, indicheremo con X_i la componente i -ma del vettore \underline{X} .

In accordo alla *formula di Bayes* possiamo scrivere

$$P(Y = y_i | \underline{X} = x_k) = \frac{P(\underline{X} = x_k | Y = y_i) \cdot P(Y = y_i)}{\sum_j P(\underline{X} = x_k | Y = y_j) \cdot P(Y = y_j)}$$

dove y_i indica l' i -mo elemento del supporto di Y , mentre x_k indica la k -ma possibile assegnazione del vettore \underline{X} .