



Intelligenza Artificiale e Big Data

Michela Milano

Dipartimento di Informatica – Scienza e Ingegneria

Cesena – 13 Dicembre 2018



Sommario

- Introduzione
 - Intelligenza Artificiale
 - Data analytics
- Casi di studio
- Tecnologie
 - Machine learning
 - Ottimizzazione e decision support systems



Intelligenza Artificiale

- L'essenza dell'intelligenza è ciò che gli scienziati cognitivi, i filosofi e gli psicologi hanno sempre provato a comprendere.
- L'Intelligenza Artificiale ha come scopo quello di ingegnerizzare sistemi che esibiscono una forma di intelligenza attraverso la visione, il linguaggio, l'emozione, il movimento, il ragionamento per spingere ciò che i computer possono fare autonomamente.
- General AI vs Narrow AI



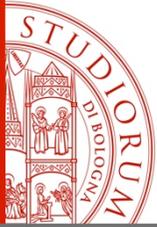
Approcci simbolici e sub-simbolici

AI Simbolica: sistemi basati sulla conoscenza

- Conoscenza comprensibile
 - Regole If-then, teorie logiche
- Il ragionamento e l'apprendimento producono output comprensibile

AI sub-simbolica: reti neurali

- L'intelligenza viene raggiunta tramite strutture "simili" a quelle del cervello umano.
- Il ragionamento e l'apprendimento modificano pesi
- Ciò che viene appreso è incomprensibile

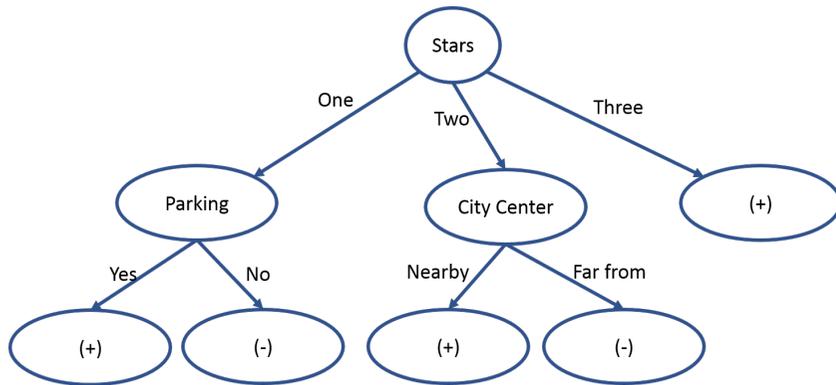


Approcci simbolici e sub-simbolici

PROLOG

descendant(D,A) :- parent(A,D) .
 descendant(D,A) :- parent(P,D) , descendant(P,A) .

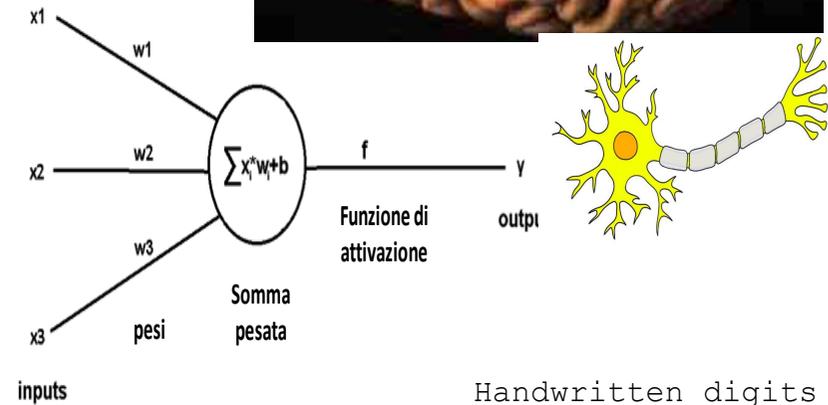
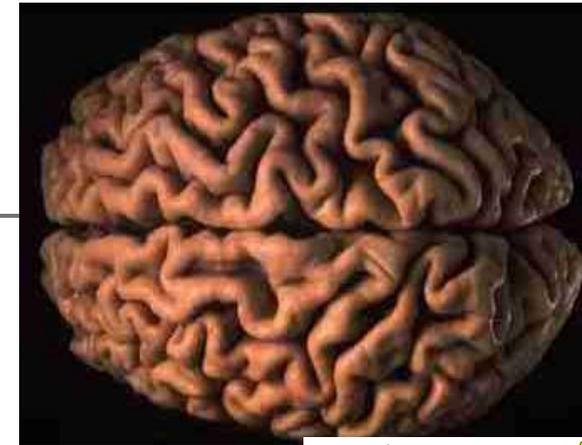
Alberi decisionali



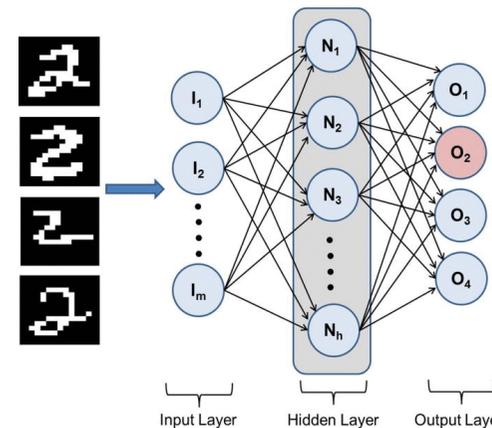
Regole di classificazione

if the Hotel is "three-star"
then it is to be recommended.

If the Hotel is "two-stars"
and is nearby the "City Center" **then** it is to be recommended.



Handwritten digits



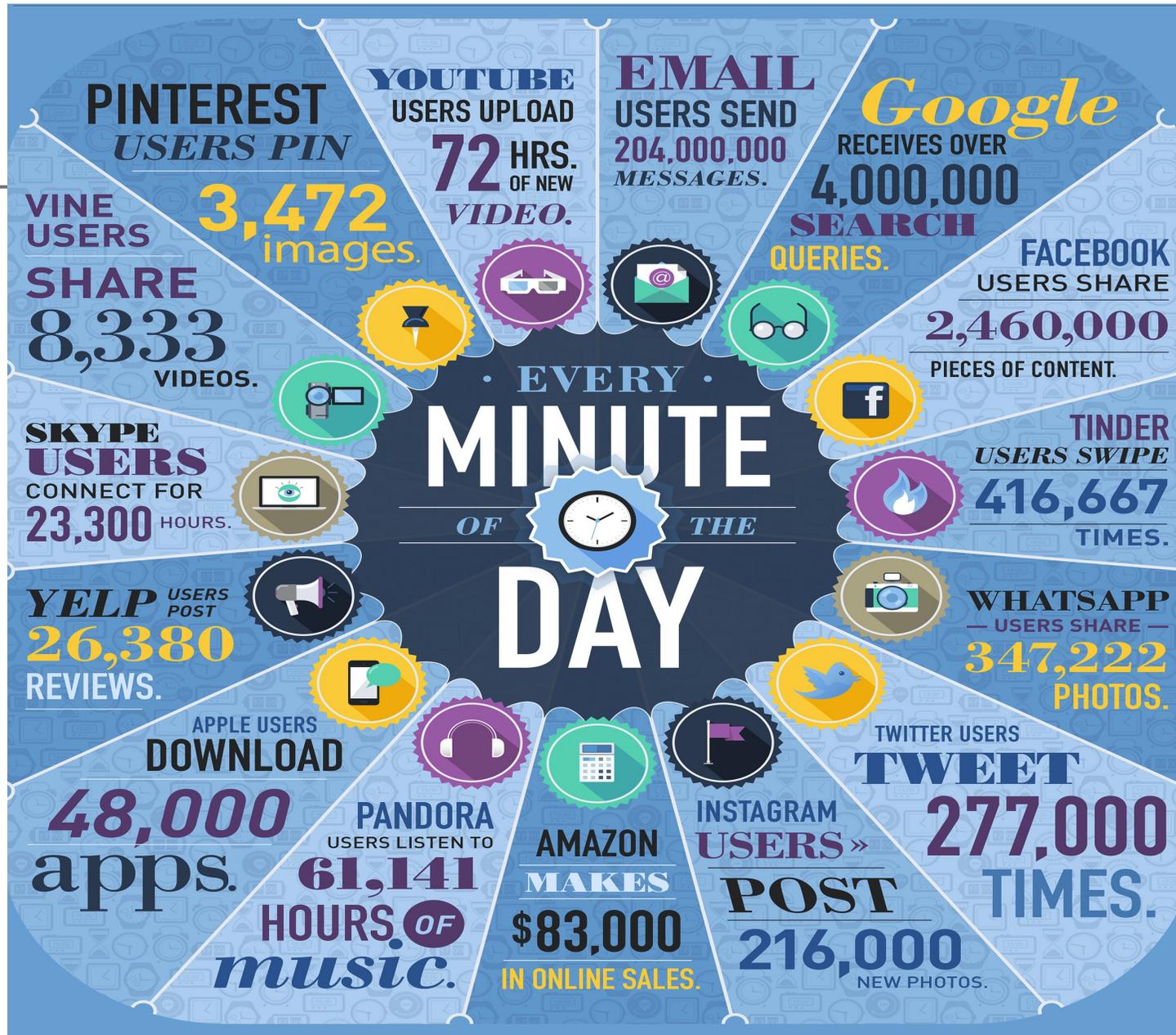
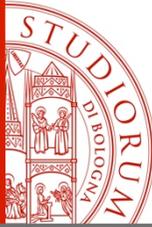


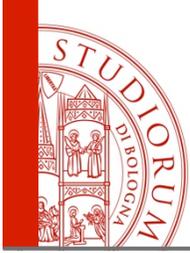
Tecnologie di Machine Learning

- Il Machine learning è uno degli **hot topic** in Intelligenza Artificiale.
- Tutte le più importanti compagnie di ICT al mondo stanno investendo denaro e stanno assumendo persone con un forte background in ML: Google, Facebook, IBM, Baidu, Disney, . . .

PERCHE ?

- Perche' hanno **moltissimi dati**. . .
 - . . . E hanno bisogno di **ricavare valore** da questi dati



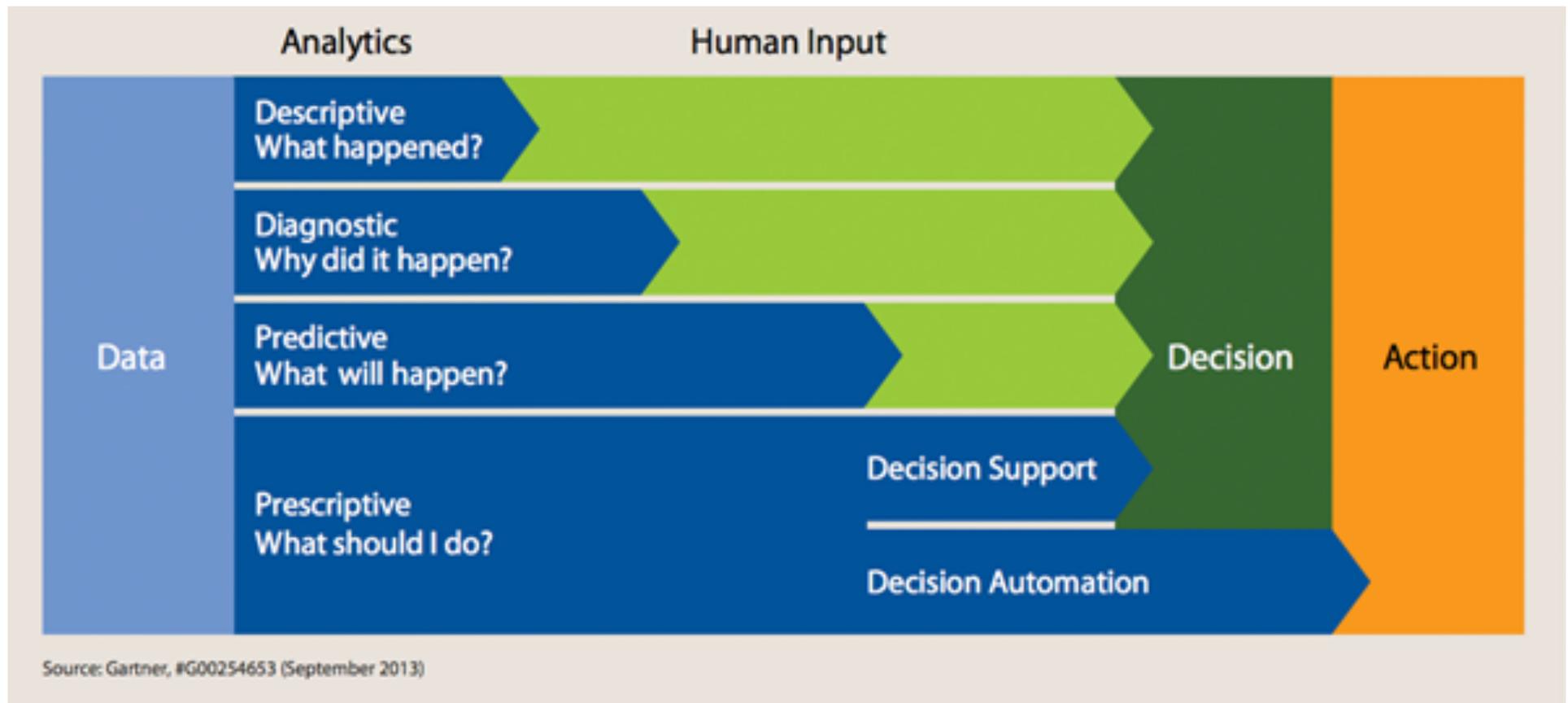


Decision support systems

- L'intelligenza artificiale ha sempre cercato di costruire sistemi in grado di supportare esperti nel prendere decisioni informate e basate sui dati.
- Questi sistemi si basano su modelli che uniscono
 - Conoscenza degli esperti di dominio
 - Conoscenza estratta dai dati



Intervento umano nelle decisioni



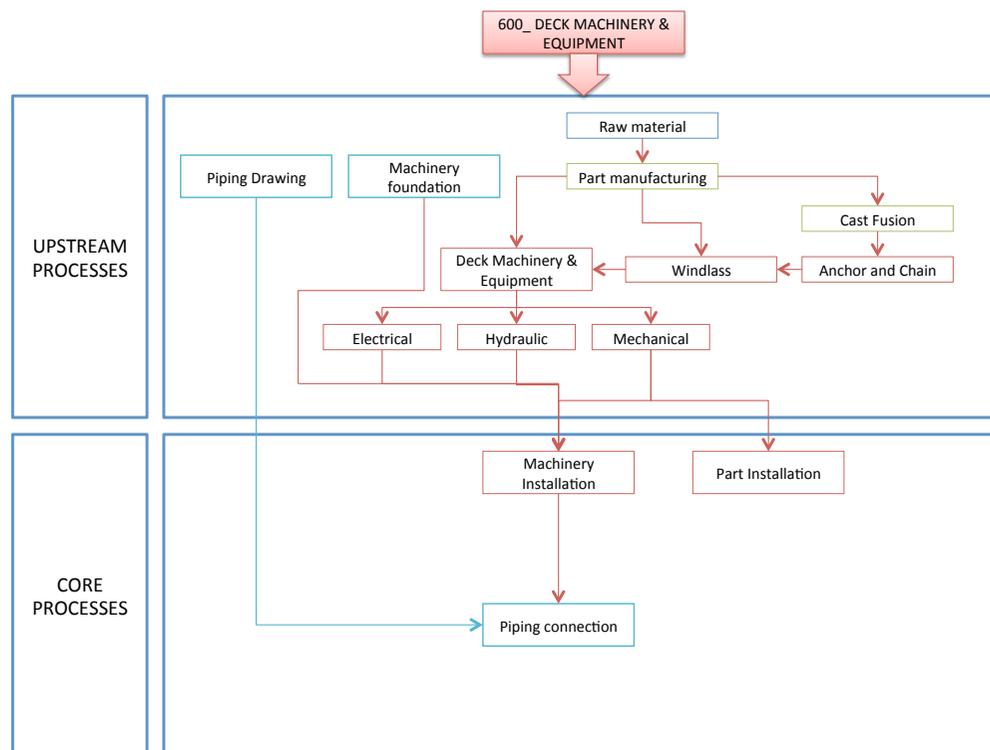


Sistemi di supporto alle decisioni

- **Descriptive analytics**
 - I dati vengono usati per descrivere il sistema
 - Cruscotti di visualizzazione per comprendere i dati
 - Business intelligence
 - Report su statistiche di vendita
 - Dati geo-referenziati
- Per arrivare alle decisioni intervento umano largamente necessario

Esempio di descriptive analytics

- Sistema per configurare un processo produttivo con vincoli ambientali



Per ogni componente del processo estrarre per ogni attività da dati pregressi statistiche su:

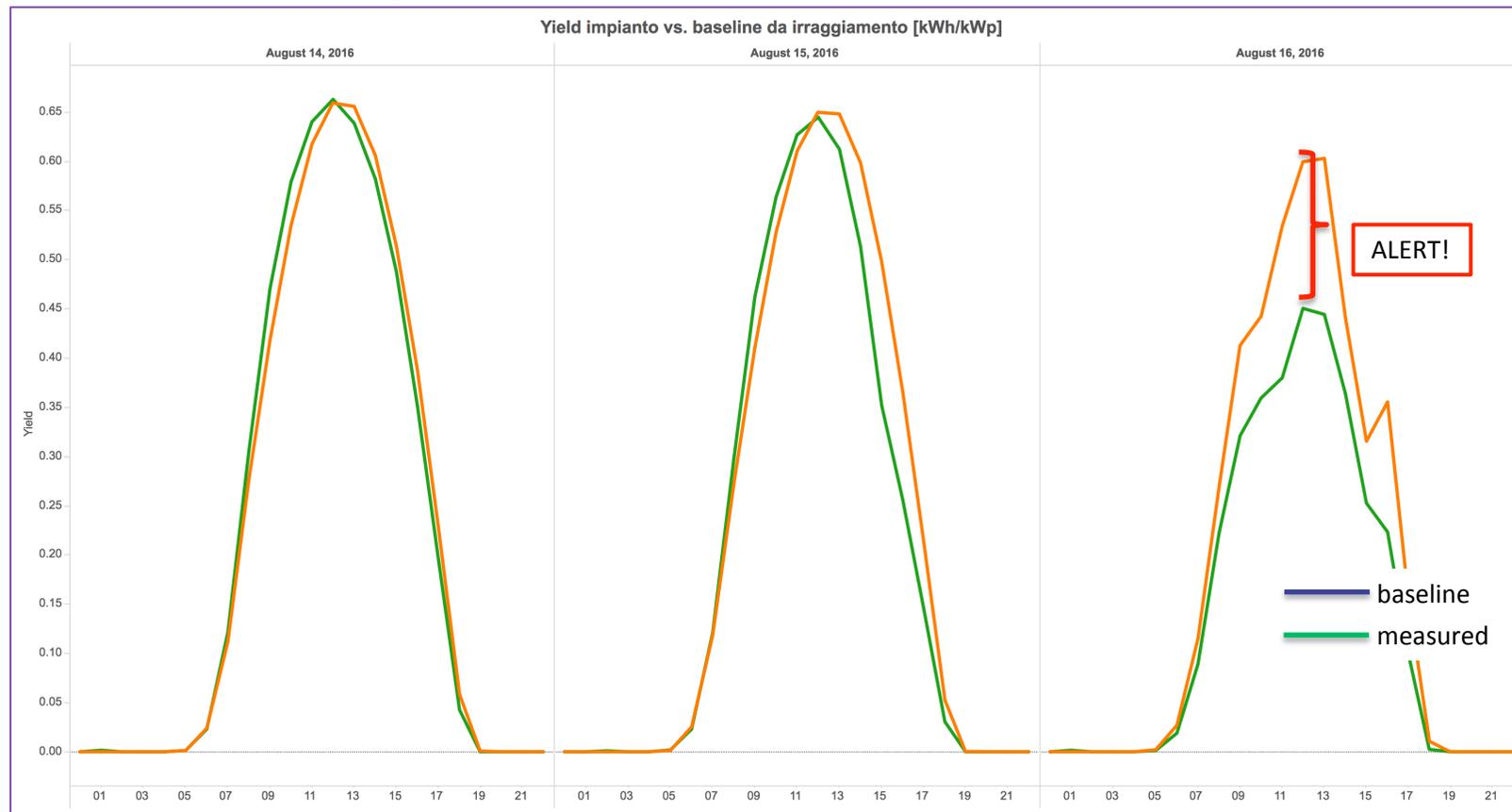
- Consumi energia
- Consumo acqua
- Emissioni in aria
- Emissioni in acqua
- Tempo di lavorazione
- Costo di lavorazione

CARATTERIZZAZIONE DEL SISTEMA: dati statici



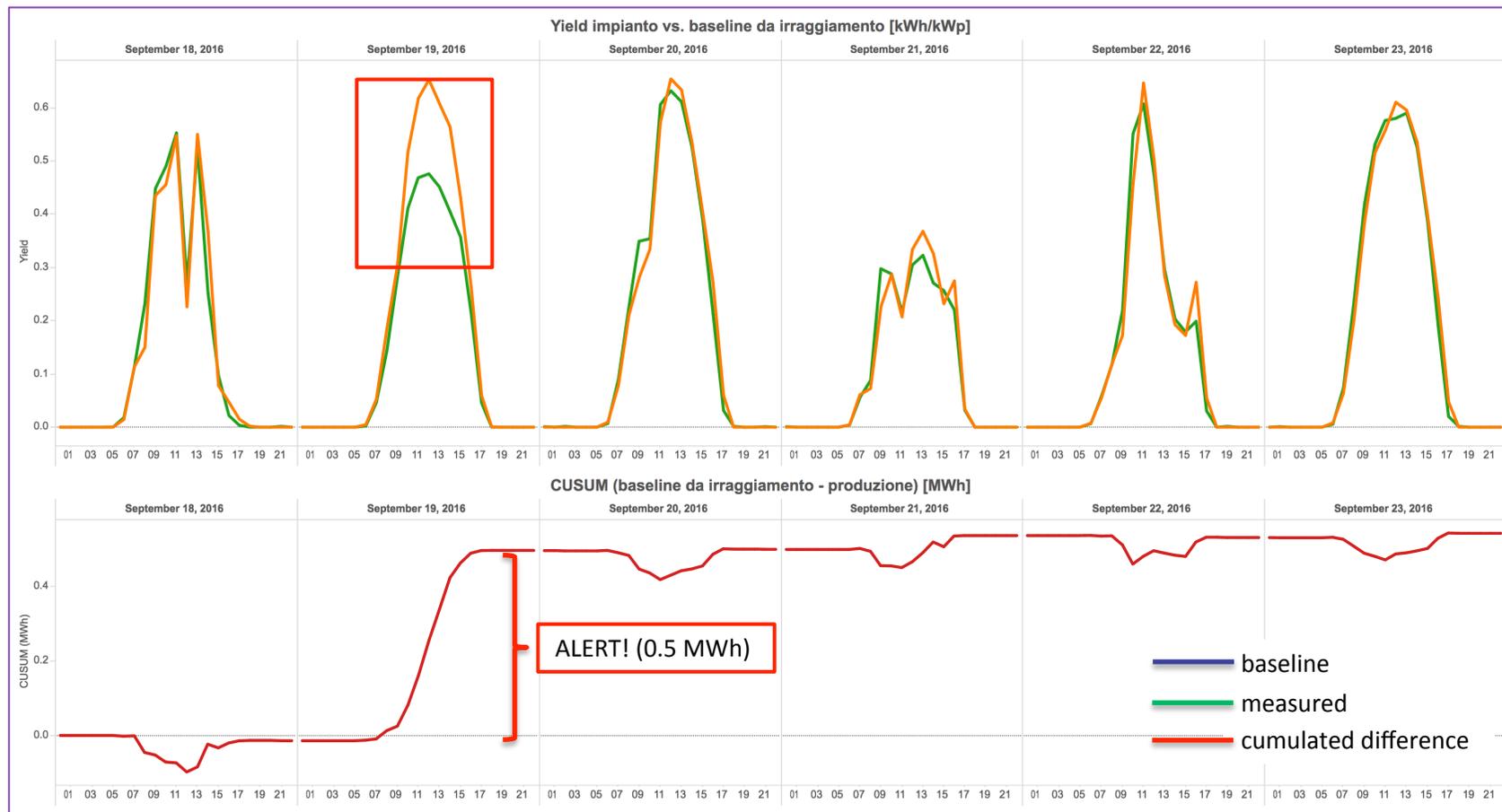
Esempio di descriptive analytics

- **Baseline energetiche**



Esempio di descriptive analytics

- Baseline energetiche





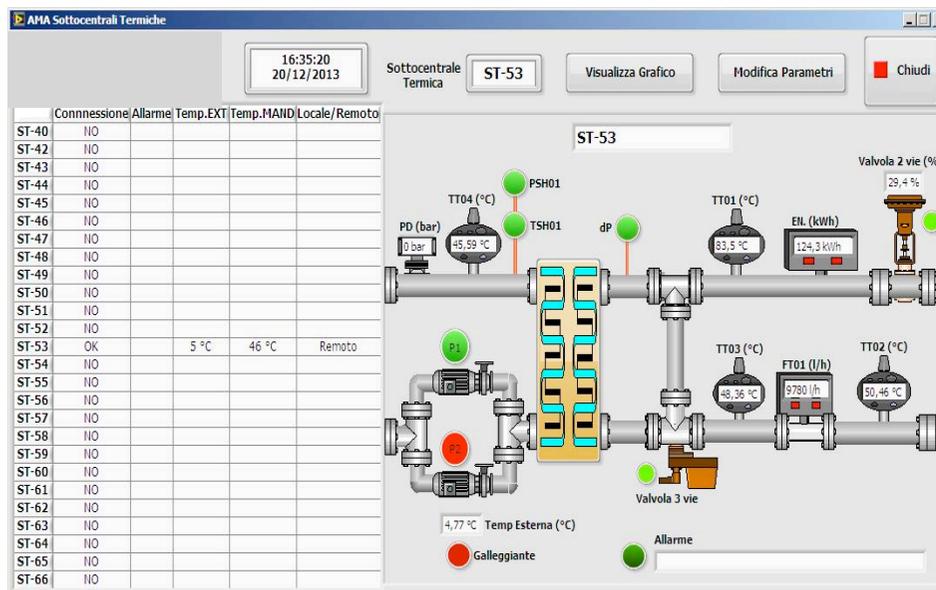
Sistemi di supporto alle decisioni

- **Diagnostic analytics**
 - Utilizza i dati per comprendere le cause
 - Diagnosi di guasti
 - Sistemi di alerting preventiva

- Per arrivare alle decisioni intervento umano ancora largamente necessario

Esempio di diagnostic analytics

- Sistema per analizzare i dati di impianto e diagnosticare il tipo di guasto e la sua causa



© 2016 National Instruments Corporation

Dai dati di impianto (da sensori) + dati guasti:

- Classificare le serie temporali sintomatiche di guasto
- Classificare le serie sintomatiche di guasto per tipo di guasto
- Individuare i sensori più informativi
- Creazione modelli per individuarne le cause.



Predictive analytics

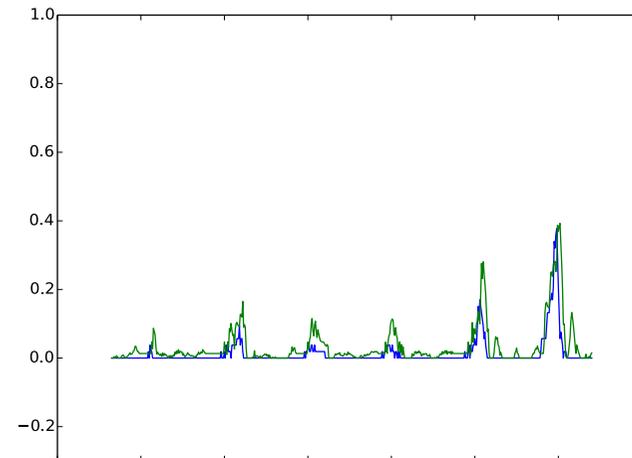
- **Predictive analytics**
 - Utilizza i dati per predire le evoluzioni future
 - Sistemi di simulazione
 - Predizione di serie temporali
 - Predizione di guasti
 - Predizioni di vendite

- Per arrivare alle decisioni what-if analysis



Esempio di predictive analytics

- **Sistema per predire la produzione energetica da un impianto fotovoltaico**
 - Ingresso
 - Serie temporale passata dei dati di produzione
 - Previsione meteo
 - Giorno/mese/anno
 - Uscita
 - Dati di produzione futura





Sistemi di supporto alle decisioni

- **Prescriptive analytics**
 - Sistemi decisionali
 - Sistemi di ottimizzazione (mono e multi obiettivo)
 - Visualizzazione di scenari alternativi

- Per arrivare alle decisioni non serve intervento umano, se non nella selezione

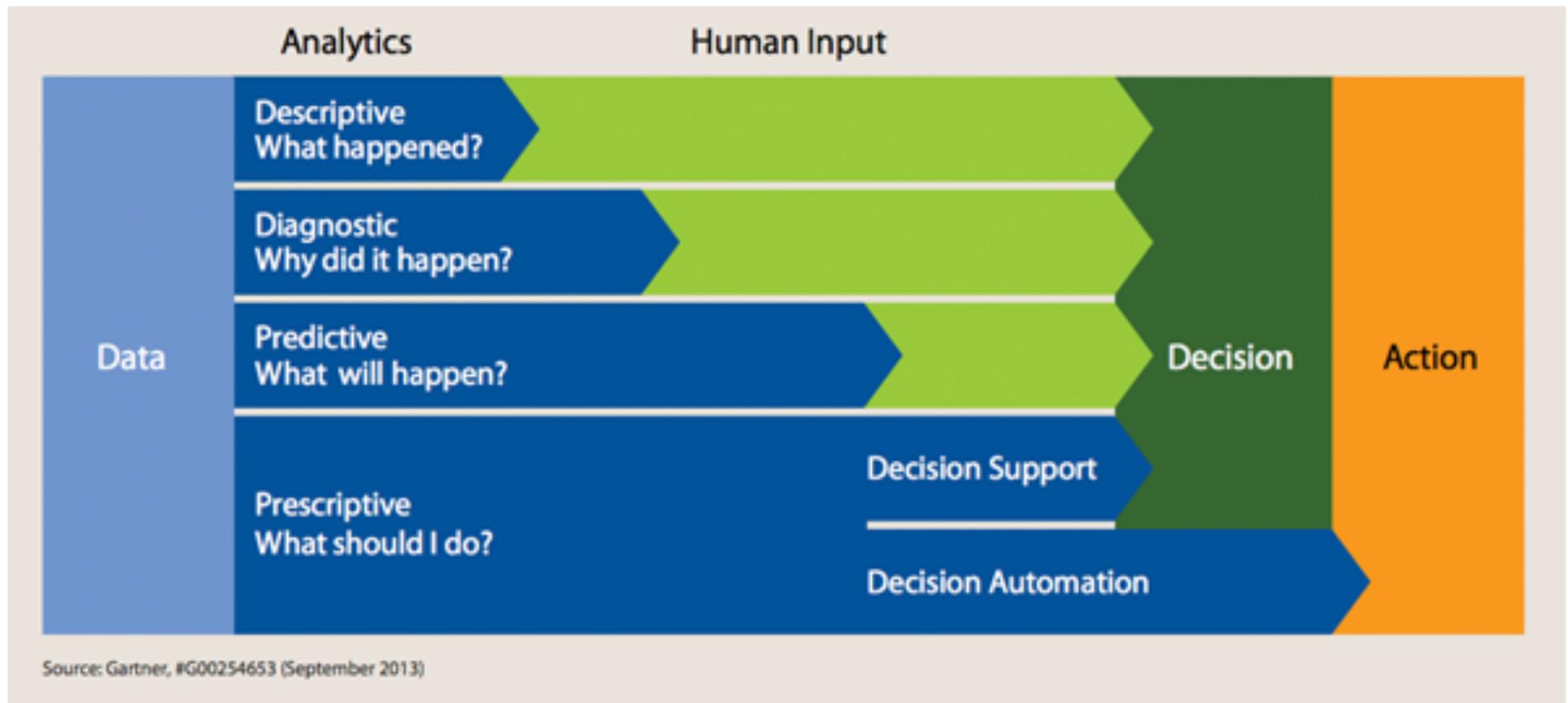


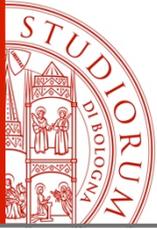
Esempio di Prescriptive Analytics

- **Piani manutentivi ottimi**
 - Ottimizzazione combinatoria stocastica che prende in input
 - Probabilità di guasto/tempo al guasto
 - Modelli Costo/tempo per interventi manutentivi
 - Funzioni obiettivo
 - Output
 - Piani manutentivi
 - Gestione di team di manutentori

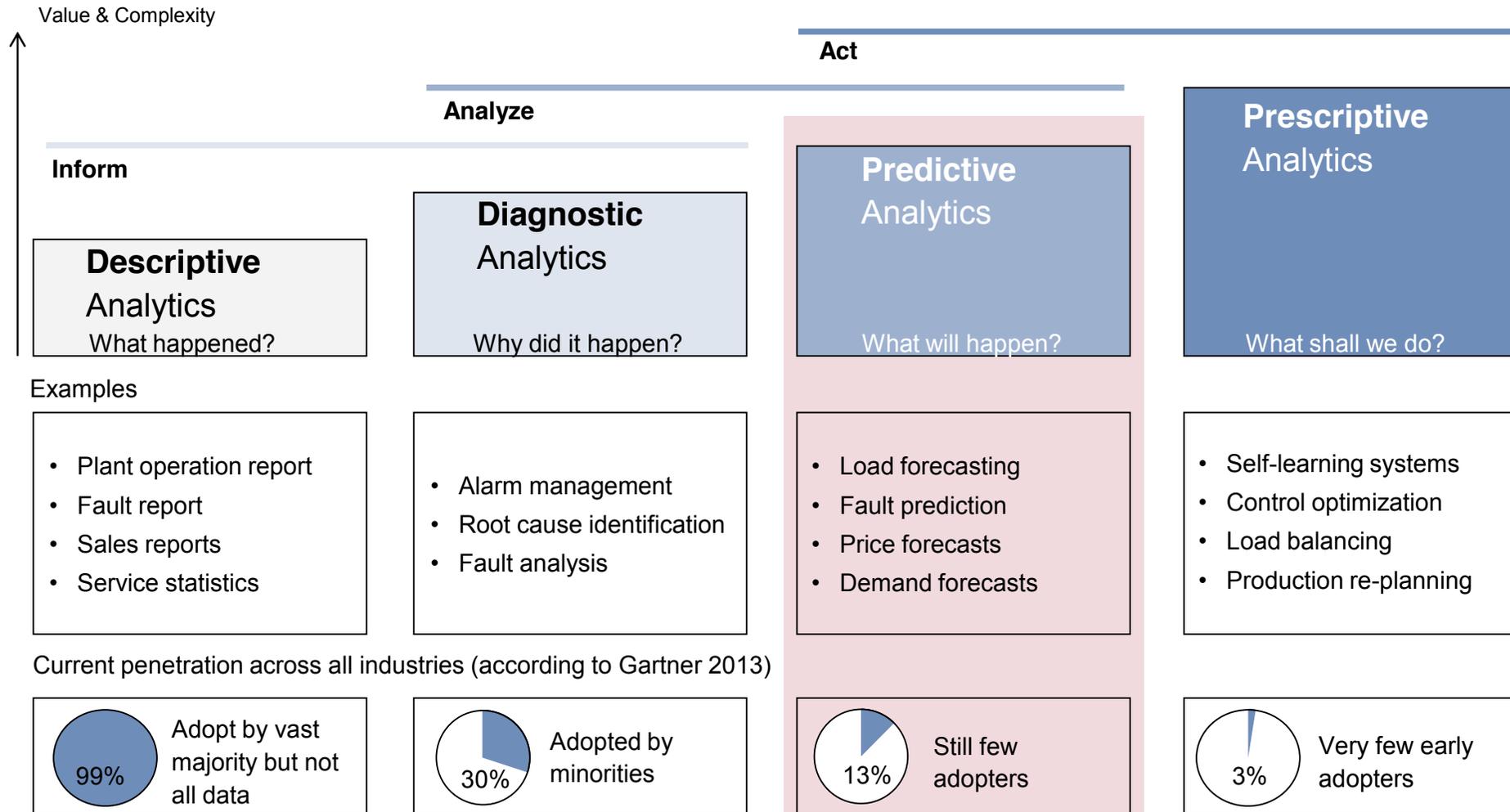


Intervento umano nelle decisioni





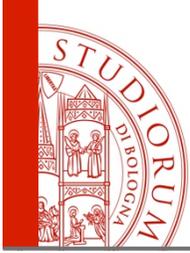
Adozione industriale a livello EU





Limitazione sistemi predittivi

- Predictive analytics: simulatori e predittori
 - What if analysis per il supporto alle decisioni
 - Esamino uno scenario e lo simulo o predico l'effetto/l'impatto
 - Quando il numero di scenari è alto (esponenziale) l'approccio è inutilizzabile



Limitazione sistemi prescrittivi

- Prescriptive analytics: sistemi di supporto alle decisioni
 - Tradizionalmente i modelli decisionali si basano su programmazione matematica/programmazione a vincoli
 - Modello definito dall'interazione con l'esperto
 - Approssimazione del modello non nota



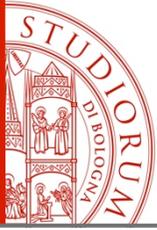
Integrazione livelli predittivo prescrittivo

- Focus su predictive analytics e prescriptive analytics
- Integrando i livelli si possono ottenere modelli decisionali basati sui dati
- Componenti del modello: predittori
- Machine learning per imparare componenti del modello decisionale a partire dai dati
- Accuratezza nota almeno su test set



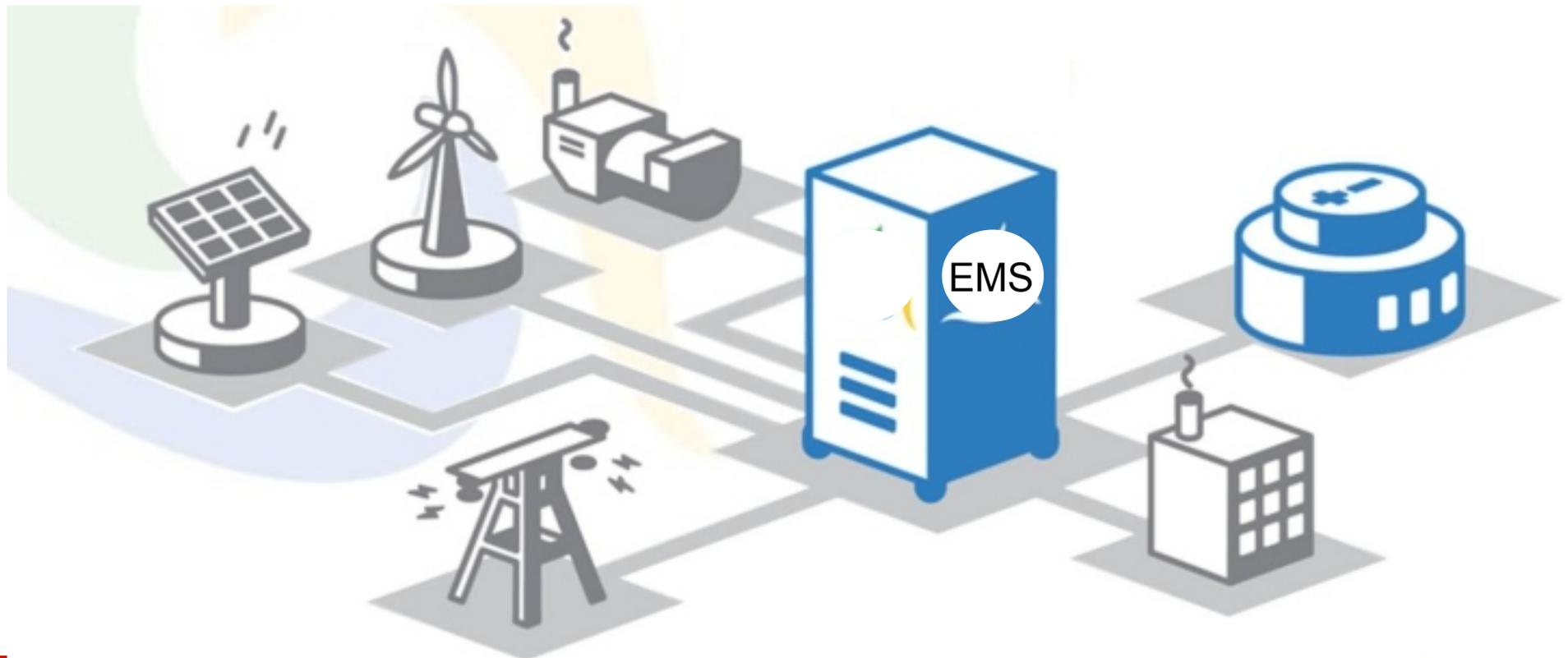
Parte 1: CASI D'USO

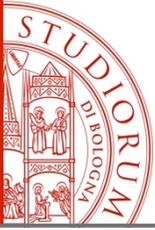
Descrizione



Caso d'uso 1: Virtual power plant

- Virtual Power Plant: Progetto RdS VIRTUS
 - Accumulo/Produzione/Consumo





Caso d'uso 1: Virtual power plant

- Decidere in ogni istante di tempo se acquistare energia, se consumare quella prodotta, o accumulata, se accumulare quella prodotta o mandarla sulla rete per minimizzare il costo totale.





Caso d'uso 1: Virtual power plant

- Modelli predittivi:
 - Predizione prezzi mercato
 - Predizione energia prodotta dalle RES
 - Predizione consumo



Modello prescrittivo

- Orizzonte temporale discretizzato (15 min - 1 h)
- Ad ogni time step t dobbiamo prendere le seguenti decisioni:
 - Load Shift
 - Energia comprata dalla rete Q_t
 - Energia rinnovabile autoconsumata R_t
 - Energia accumulata consumata S_t
 - Energia rinnovabile accumulata A_t
- **Predizioni:**
 - Prezzo energia orario $pr(t)$
 - Produzione da rinnovabili
 - Carico atteso
 - Capacità delle batterie
 - Energia accumulabile/utilizzabili per intervallo temporale

*Predictive
Analytics*

*Static
Data*



Caso d'uso 1: Virtual power plant

- Utilizzando modelli predittivi e prescrittivi integrati

15-25% SAVING SU COSTI ENERGETICI



Caso d'uso 2: Predictive maintenance

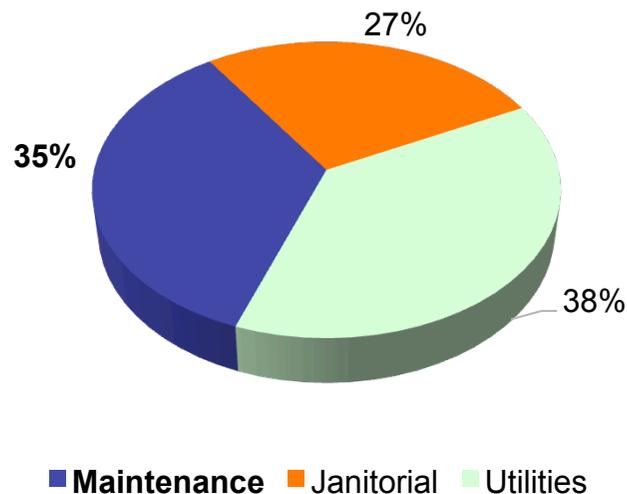
- La manutenzione viene vista sempre più come un asset strategico di business
 - Riduzione downtime
 - Riduzione production losses
 - Riduzione breakdown
- Tre livelli di manutenzione
 - Reactive: avviene un guasto e viene gestito
 - Preventative: manutenzione eseguita regolarmente su un macchinario per diminuire la possibilità di guasto.
 - Predictive: piano di manutenzione ottimizzato sulla probabilità di guasto



Costi manutenzione

Il costo della manutenzione è circa un terzo del costo totale di una facility energetica

Typical Facility Operating Cost Breakdown



Costi

- **Downtime non pianificato**
- Costo di macchinari secondari danneggiati dal guasto principale
- **Uso inefficiente** delle squadre manutentive
- **Manutenzione inutile**

Source: IFMA, October 2009

Despite the importance and expense of maintaining building efficiency most building operators rely on reactive maintenance programs to care for their equipment.



Fault prediction

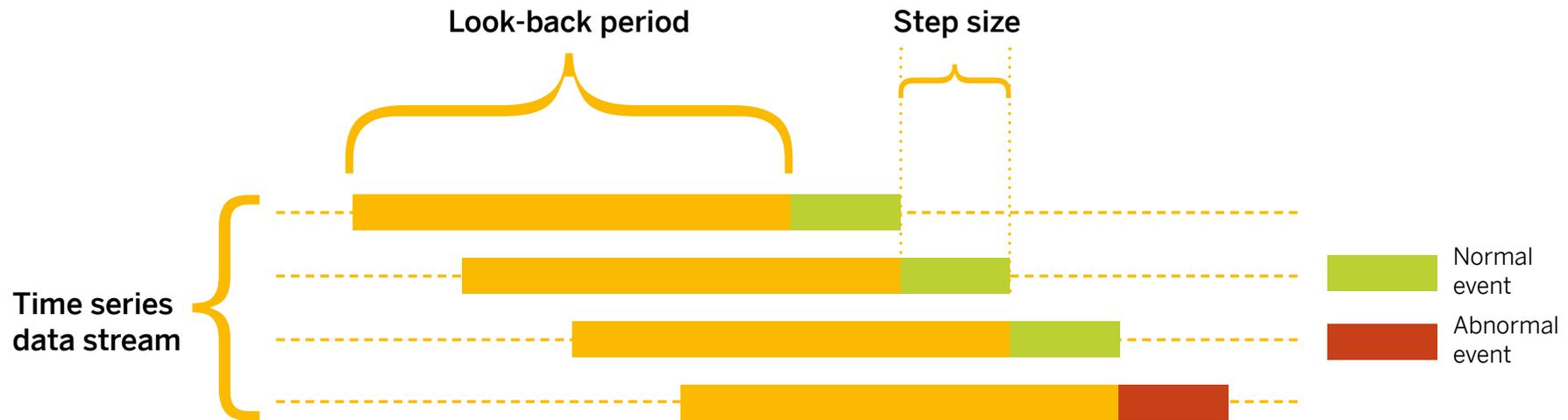
- Data streams provenienti dai sensors
 - 120 sensori campionati ogni 30 seconds

Time stamp	Machine ID	Sensor ID	Sensor value
18/01/2009 10:32:46	1020010	Sensor 1	0.924387
19/01/2009 10:32:46	542	Sensor 4	1.2235
20/01/2009 10:32:46	1020010	Sensor 3	3.397467
21/01/2009 10:32:46	542		123.876
22/01/2009 10:32:46	1020010	Sensor "n"	3.569223
23/01/2009 10:32:46	6482	Sensor 99	12887
24/01/2009 10:32:46	1020010	Sensor 2	3.600801
25/01/2009 10:32:46

- Data preprocessing
 - Divided per machine
 - average/variance per minute/hour



Fault prediction



Survey SAP on predictive maintainance

- Predizione tramite machine learning:
 - Tempo al guasto/Probabilità di guasto
 - Quali sensori sono rappresentativi?



Caso d'uso 2: Predictive maintenance

- Caso di studio di un produttore di energia (oil-gas)

- Eliminazione breakdown: 70%-75%



- Riduzione downtime: 35% - 45%



- Riduzione costi manutenzione: 25%- 30%



- Incremento produzione: 20%- 25%



Roland Berger Strategy Consultants, Survey 2014



Parte 2: LE TECNOLOGIE



Definizione di Machine Learning

- Definizione 1:
 - Learning is constructing or modifying representations of what is being experienced [Michalski 1986]
- Definizione 2:
 - Learning denotes changes in the system that are adaptive in the sense that they enable the system to do the same task or tasks drawn from the same population more efficiently and more effectively the next time [Simon 1984], pag. 28



Il ruolo del Machine Learning

- Impieghi
 - A) estrazione di conoscenza
 - B) miglioramento delle performance di una macchina

Tecniche

- Tecniche simboliche (impieghi A e B)
- Tecniche statistiche (impiego B)
- Tecniche sub-simboliche (impiego B)

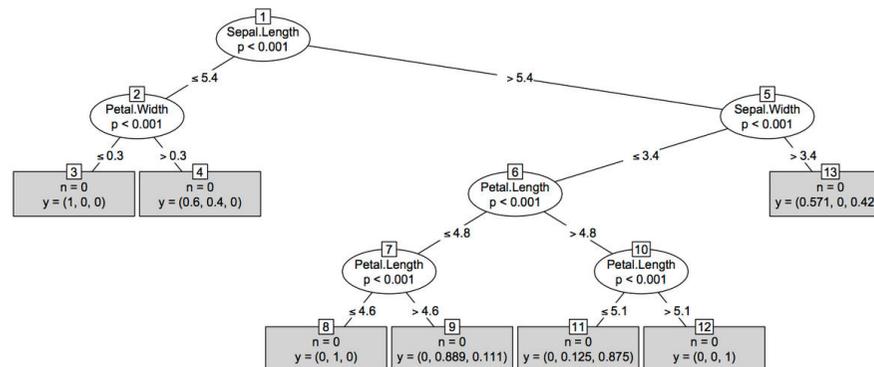
- In ogni caso, è necessario apprendere un modello dai dati e generalizzare



Tecniche simboliche di ML

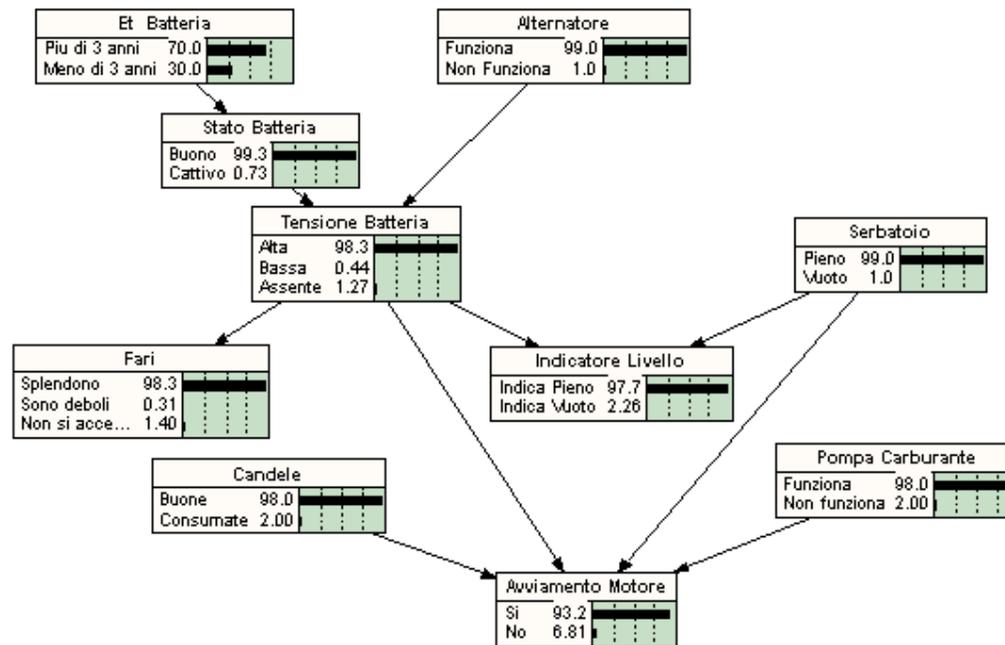
- Estrazione di regole logiche
 $father(X, Y) :- parent(X, Y), male(X)$

- Estrazione di alberi di decisione e varianti



- Alberi di regressione/random forests ecc.....

- Apprendimento bayesiano
 - Estrazione di reti bayesiane



- Regressione lineare, non lineare, multipla



Tecniche sub-simboliche di ML

- Reti neurali
 - Apprendimento supervisionato: a partire da esempi
 - Apprendimento non-supervisionato
 - Apprendimento con rinforzo
- **Reti neurali profonde**
- Support vector machines



Machine learning Tasks: Classificazione

- **La classificazione riguarda la costruzione di una funzione di mapping da variabili di input e a variabili di output che sono discrete o categoriche.**
- Output variables: si chiamano labels
- La funzione di mapping predice la classe o la categoria per una certa osservazione.
- Supervised learning task



Machine learning Tasks: Regressione

La regressione è il task di approssimare una funzione di mapping dalle variabili di input a una variabili di output continua.

Variabili di output sono spesso quantità o probabilità

Supervised learning task

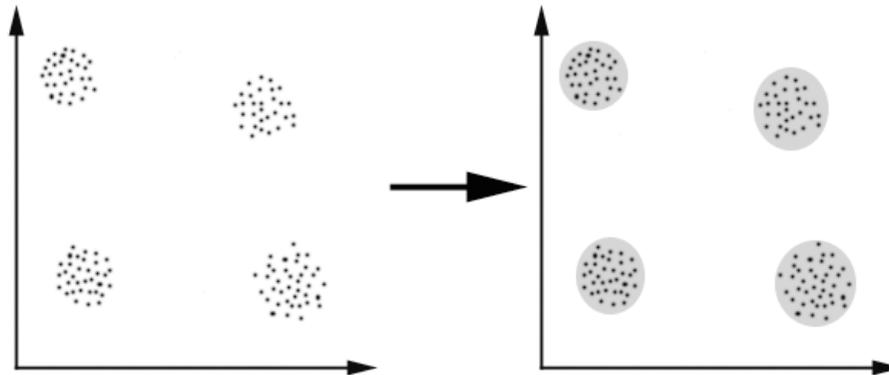


Machine learning Tasks: Clustering

Il Clustering è il problema di organizzare oggetti in gruppi i cui membri sono simili

Il clustering determina un raggruppamento intrinseco a un insieme di dati non etichettati

unsupervised learning task





Cosa si puo' fare con il ML

Classificare le mail come spam o not



Figure source: <http://www.resilientsystems.co.uk/>



Cosa si puo' fare con il ML

Predizione del prezzo delle azioni Apple nelle prossime tre ore





Cosa si puo' fare con il ML

Fare una diagnosi basata sugli esami di un paziente

The average person is likely to generate more than one million gigabytes of health-related data in their lifetime. Equivalent to 300 million books.

IBM Watson Health



Cosa si puo' fare con il ML

Riconoscere le cifre nei codici di avviamento postale

80322-4129 80206

40004 14310

37879 05453

~~33502~~ 75216

35460 44209

Figure source: [LeCun et al., 1989]



Cosa si puo' fare con il ML

Insegnare a un robot ad afferrare una tazza

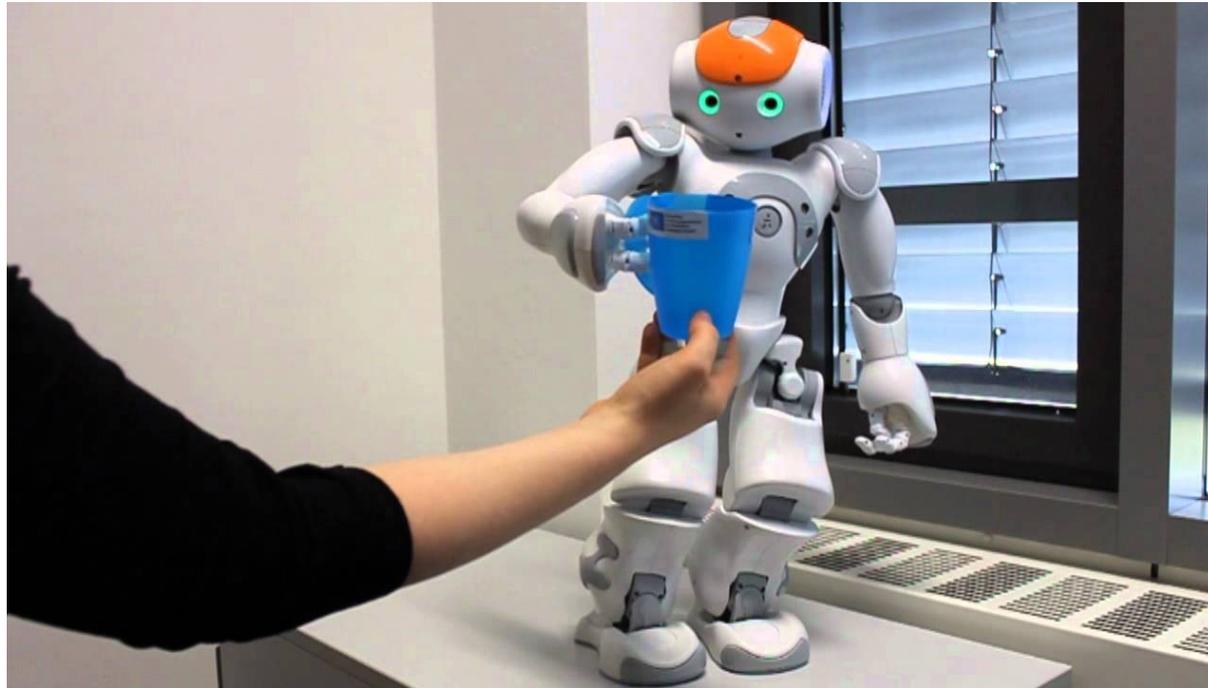


Figure source: <http://www.informatik.uni-bremen.de/>



Cosa si puo' fare con il ML

Progettare una molecola di un farmaco con determinate proprietà

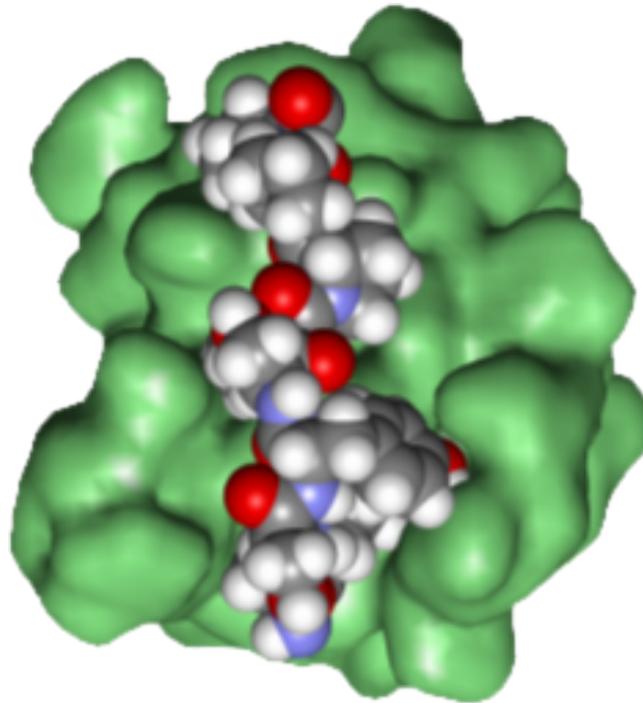


Figure source: <http://pande.stanford.edu/>



Cosa si puo' fare con il ML

Tradurre testo da una lingua ad un'altra





Cosa si puo' fare con il ML

Convertire la voce in un testo



Cosa si puo' fare con il ML

Scrivere automaticamente la didascalia di una figura

Describes without errors	Describes with minor errors	Somewhat related to the image	Unrelated to the image
 <p>A person riding a motorcycle on a dirt road.</p>	 <p>Two dogs play in the grass.</p>	 <p>A skateboarder does a trick on a ramp.</p>	 <p>A dog is jumping to catch a frisbee.</p>
 <p>A group of young people playing a game of frisbee.</p>	 <p>Two hockey players are fighting over the puck.</p>	 <p>A little girl in a pink hat is blowing bubbles.</p>	 <p>A refrigerator filled with lots of food and drinks.</p>
 <p>A herd of elephants walking across a dry grass field.</p>	 <p>A close up of a cat laying on a couch.</p>	 <p>A red motorcycle parked on the side of the road.</p>	 <p>A yellow school bus parked in a parking lot.</p>

Figure source: Google Research



Cosa si puo' fare con il ML

Battere il campione del mondo del gioco del Go (Marzo 2016 !)



Figure source: Google Research



Supervised vs. Unsupervised learning

Supervised learning

- 😊 Le etichette dei punti sono molto utili
- 😊 Funzione da apprendere è ben definita
- 😞 Etichettare tutto il training set è costoso

Unsupervised learning

- 😊 I dati non etichettati sono ovunque
- 😊 Può estrarre l'informazione più utile dai dati
- 😞 Performance peggiori per specifici task

Semi-supervised learning ?



Reinforcement learning

Diverso paradigma: insegnante che da **premi/punizioni**

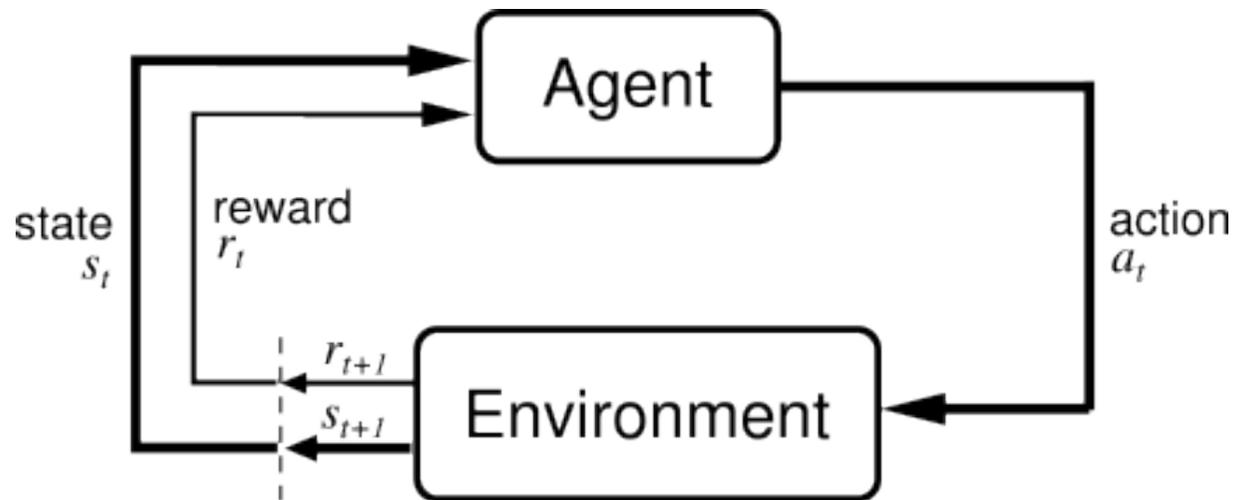


Figure source: web.stanford.edu

Usato in **molti contesti**: robotica, agenti, controllo, . . .



Reinforcement learning

A great example: computers playing Atari (by Google DeepMind)



Figure source: Wikipedia



Misure di performance

Dato un problema, come **misurare le performance** di un sistema di ML?

Modo più intuitivo: **contiamo gli errori del nostro sistema. . .**
. . . Ma non è così semplice !

Esempio

Abbiamo costruito un sistema che faccia la diagnosi di una patologia rarissima. Il nostro sistema classifica correttamente il 99.9% degli esempi. Possiamo essere soddisfatti di questo classificatore ?



Misure di performance della classificazione

		True Class	
		0	1
Pred Class	0	TN True Negative	FN False Negative
	1	FP False Positive	TP True Positive

$$\text{Accuracy } A = \frac{TP + TN}{TP + TN + FP + FN}$$

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

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

$$\text{F-Measure } F_1 = \frac{2PR}{P + R}$$

Si estendono facilmente al caso di più classi



Misure di performance della regressione

Mean squared error (MSE)

Root mean squared error (RMSE)

Mean absolute deviation (MAD)

$$MSE = \sum_{i=1}^N \frac{(\hat{y}_i - y_i)^2}{N}$$

$$RMSE = \sqrt{MSE}$$

$$MAD = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i|$$



Selezione del modello

Le performance di un sistema di machine learning dipende dalla fase di training che definisce i parametri del modello.

La procedura per scegliere i parametri migliori è chiamata **selezione del modello** che divide i dati in:

training set (o learning set)

validation set

test set

Training set usato per apprendere la funzione f .

Validation set usato per misurare le performance durante il training

Test set usato per misurare la performance del modello dopo il training su dati non visti



Validazione del modello

Il training in un sistema di ML ha lo scopo di minimizzare una funzione di costo. Spesso ha ottime performance su esempi usati per il training.

Problema

Si può imparare **perfettamente a classificare il training set**, ma fare errori grossolani sul validation set. Generalizzazione **POVERA**

Overfitting



Alberi decisionali

- Sistemi che apprendono alberi di decisione: CLS, IDR, C4, ASSISTANT, ID5, C4.5 etc.
- Problemi appropriati:
 - le istanze sono rappresentate da coppie attributo valore
 - la funzione target ha valori discreti
 - descrizioni disgiuntive di concetti possono essere richieste
 - l'insieme dei dati di training può contenere errori
 - l'insieme dei dati di training può contenere dati mancanti



Algoritmo di determinazione dell'albero

- T: insieme degli esempi;
- $\{C_1, C_2, \dots, C_k\}$: insieme delle classi;
- Considera l'insieme T:
 - T contiene uno o più esempi, tutti appartenenti alla medesima classe => singola foglia con etichetta la classe
 - T non contiene nessun esempio (insieme vuoto) => singola foglia con etichetta la classe più frequente nell'insieme padre
 - T contiene casi che appartengono a più classi => partizionamento di T in più sottoinsiemi secondo un test su un attributo => nodo associato al test, con un sottoalbero per ogni possibile risultato del test stesso. Richiama l'algoritmo su ogni ramo/sottoinsieme



Esempio

- Istanze: condizioni meteo
- Classe: Idoneità di una attività
- Attributi:
 - outlook, con valori {sunny,overcast,rain}
 - temperature, con valori numerici
 - humidity, con valori numerici
 - windy, con valori {true, false}

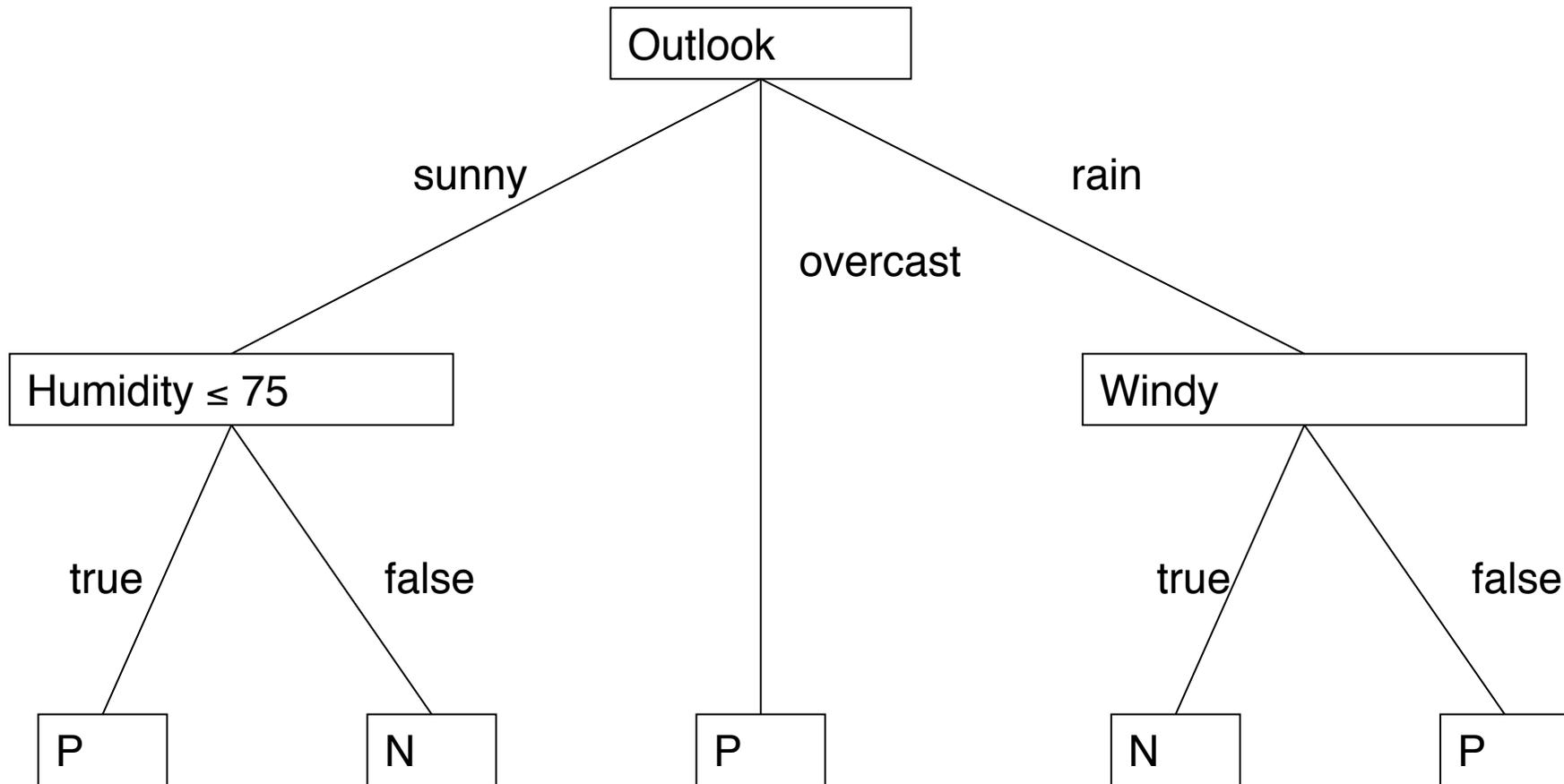


Training set

No	Outlook	Temp (°F)	Humid (%)	Windy	Class
D1	sunny	75	70	T	P
D2	sunny	80	90	T	N
D3	sunny	85	85	F	N
D4	sunny	72	95	F	N
D5	sunny	69	70	F	P
D6	overcast	72	90	T	P
D7	overcast	83	78	F	P
D8	overcast	64	65	T	P
D9	overcast	81	75	F	P
D10	rain	71	80	T	N
D11	rain	65	70	T	N
D12	rain	75	80	F	P
D13	rain	68	80	F	P
D14	rain	70	96	F	P



Albero risultante





Determinazione dell'attributo di split

- Come scegliere il test (attributo) ad ogni passo?
- Generazione di tutti i possibili alberi: non praticabile dal punto di vista computazionale.
- Viene quindi scelto un test sulla base di una euristica con scelta irrevocabile (*nonbacktracking*)
- Varie euristiche basate su:
 - sull'informazione
 - sull'errore
 - sulla significatività
- Vedremo solo quella basata sull'informazione



Entropia di un insieme di esempi

- Entropia di un insieme di esempi: informazione media necessaria ad identificare la classe di un esempio in T . Dato l'insieme di esempi S e la classe C_j

$$info(S) = - \sum_{j=1}^k \frac{freq(C_j, S)}{|S|} \times \log_2 \left(\frac{freq(C_j, S)}{|S|} \right)$$

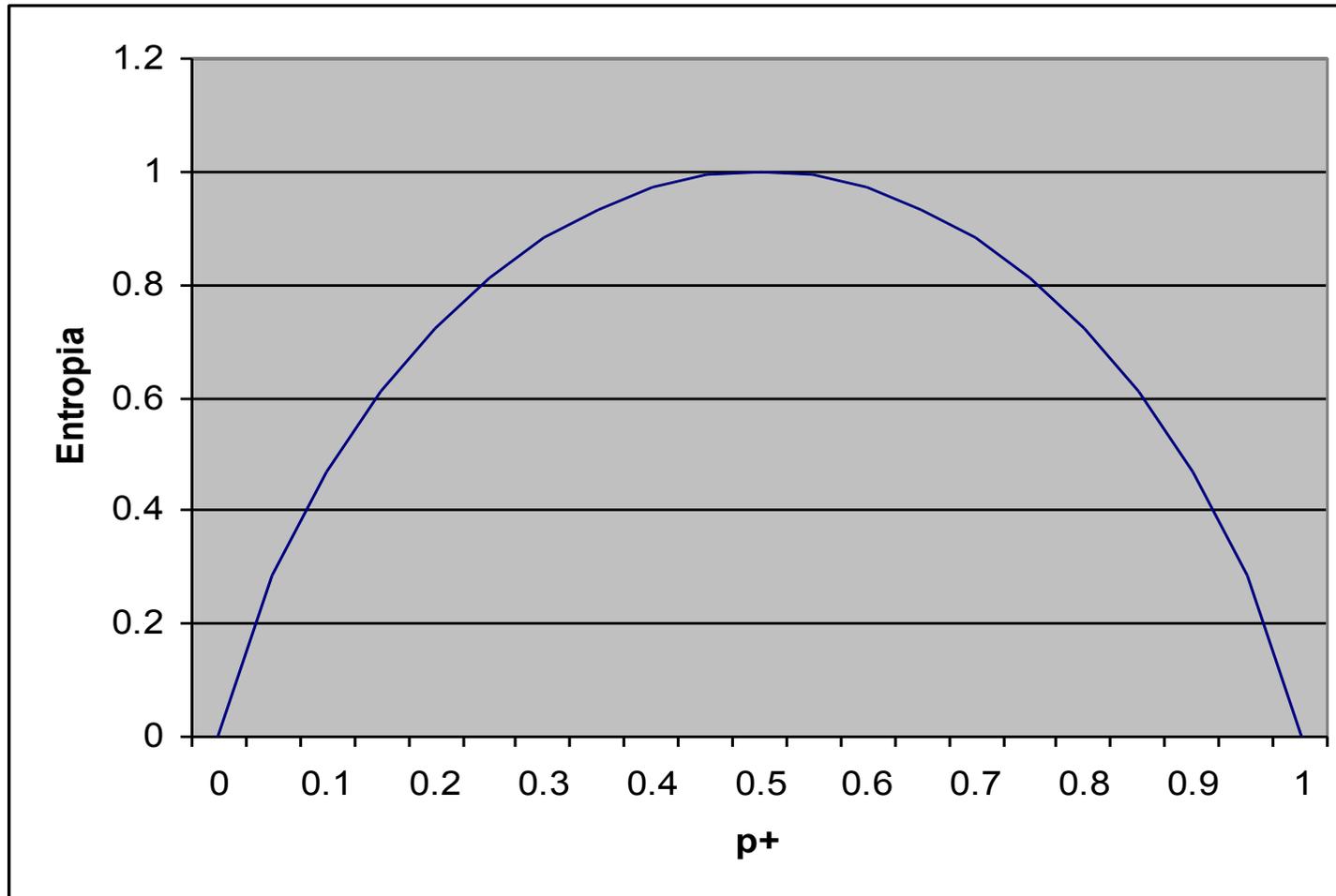
Entropia del training set

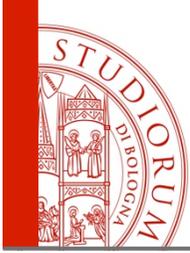
$$info(T)$$

Può anche essere vista come una misura della **disuniformità** di una arbitraria collezione di esempi: più è alta più la collezione è disuniforme



Entropia di un insieme di esempi



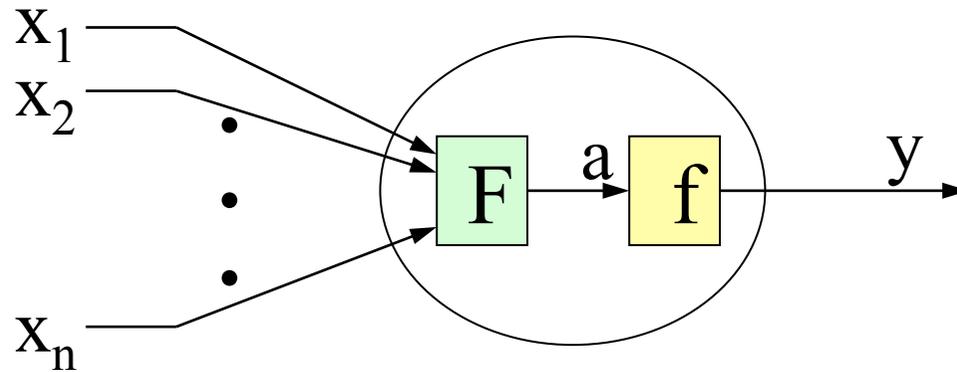


Caratteristiche degli alberi decisionali

- Lavorano bene con attributi discreti
- Discernono gli attributi informativi
 - Alberi piccoli preferibili
 - A volte fanno overfitting sul TS
- Robusti a errori
- Robusti a dati mancanti
- Varianti:
 - Alberi di regressione
 - Random forests
 - Ensemble di alberi decisionali o alberi di regressione che danno in uscita la moda della classe per la classificazione o la predizione media per la regressione degli alberi individuali

Reti Neurali

- Basate sul concetto di neurone artificiale



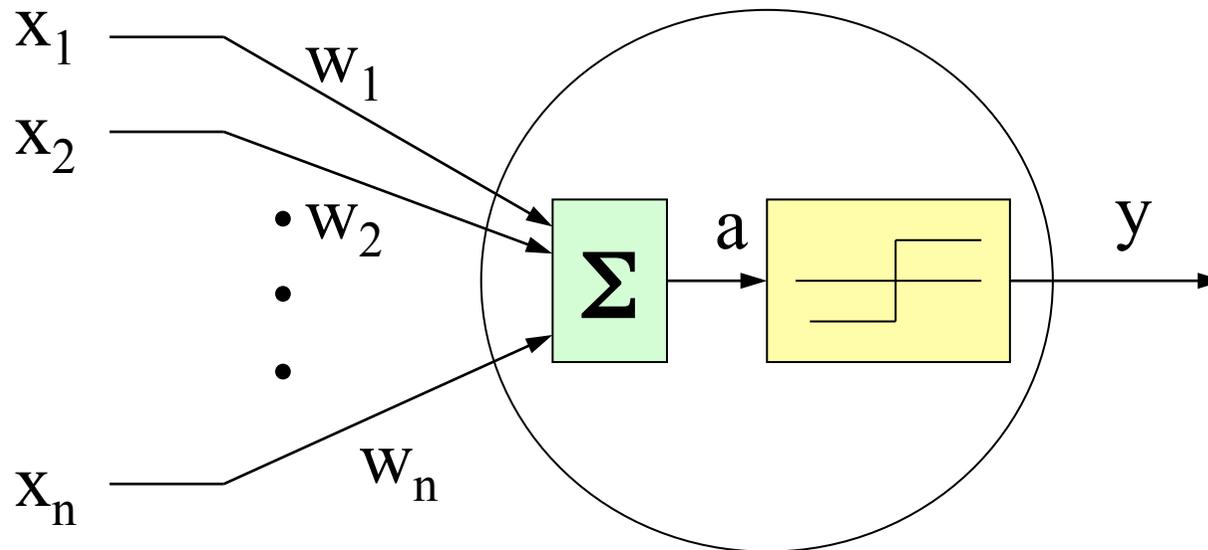
$$a(t) = F(x_1, x_2, \dots, x_n)$$

$$y(t) = f(a)$$

- La funzione F è detta funzione di attivazione e f la funzione di uscita.

Reti Neurali

- Neurone binario a soglia

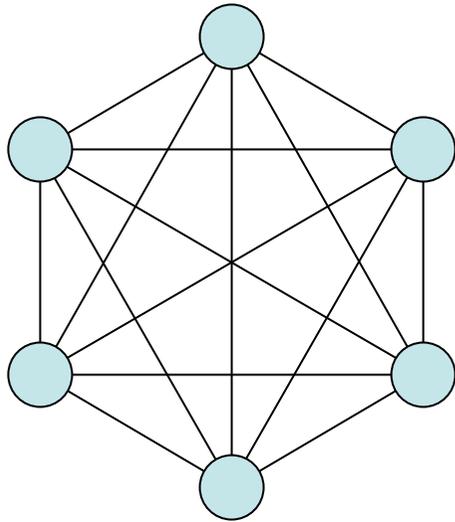


$$a = \sum_i w_i x_i$$
$$y = \text{HS}(a - \theta)$$

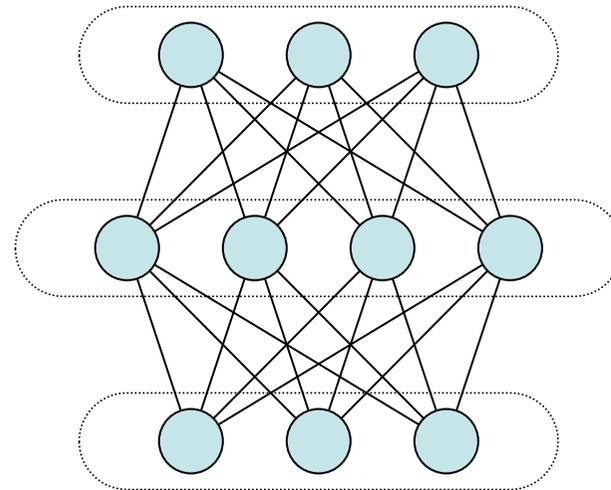
Funzionamento neuronale:
gli impulsi ricevuti dai dendriti
aumentano il potenziale elettrico nel
neurone fino a una certa soglia

Reti Neurali: struttura

- Neurone binario a soglia



Completamente connessa



stratificata

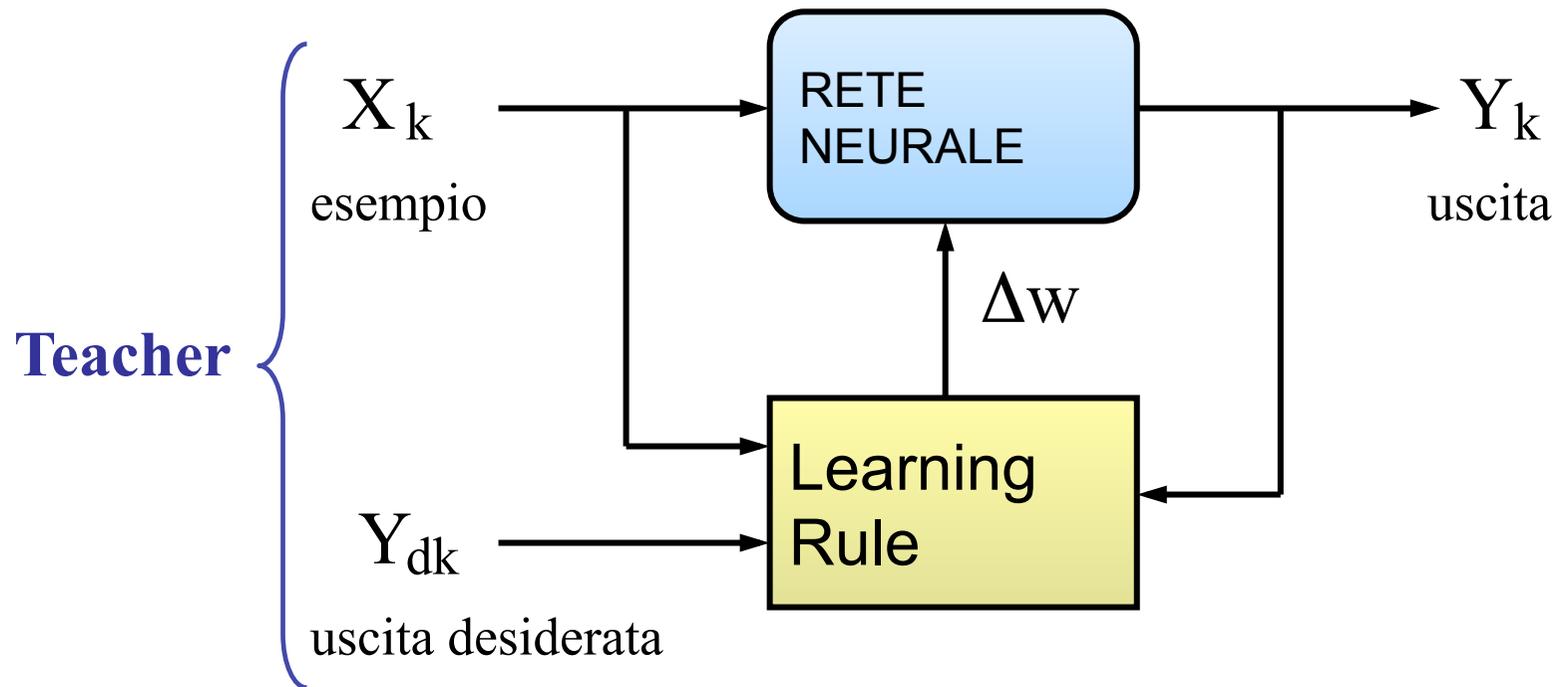


Reti Neurali: apprendimento

- Hebb (neurofisiologo) ha scoperto che l'apprendimento avviene per cambiamento delle sinapsi.
- Capacità della rete di modificare il comportamento in una direzione desiderata al variare delle connessioni sinaptiche (pesi).
- I paradigmi di apprendimento possono essere suddivisi in tre classi fondamentali:
 - supervisionato
 - competitivo
 - con rinforzo



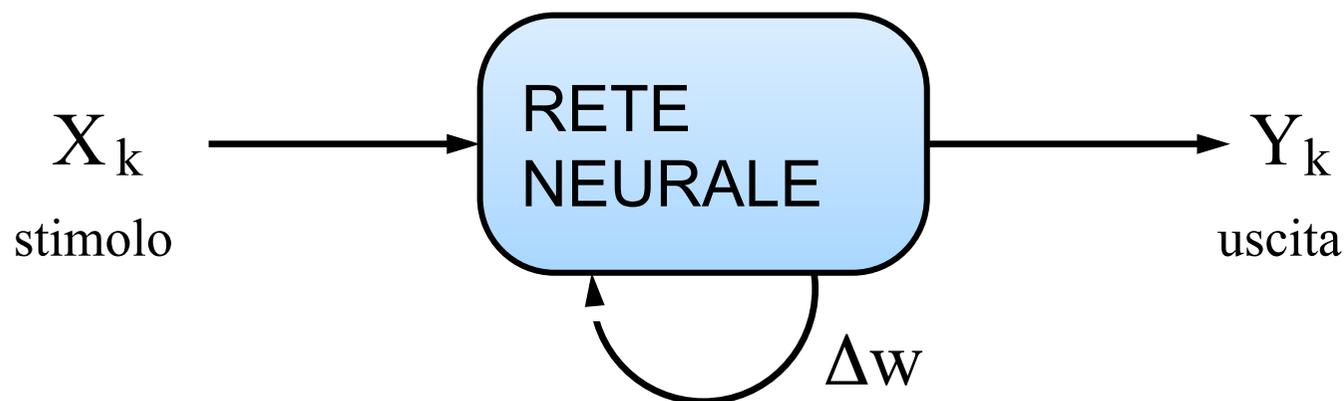
Reti Neurali: apprendimento supervisionato





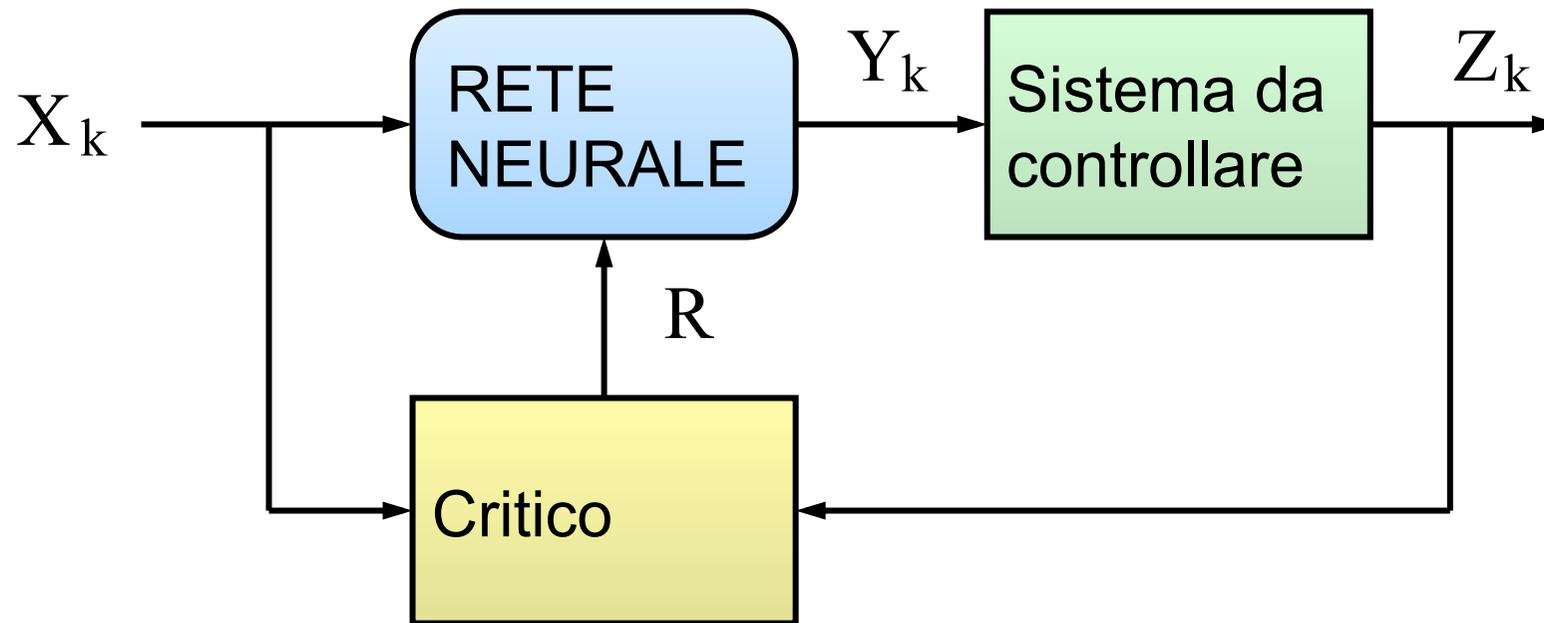
Reti Neurali: apprendimento competitivo

- I neuroni competono per specializzarsi a riconoscere un particolare stimolo. Stimoli simili finiscono nella stessa classe.
- Alla fine, ogni stimolo attiva un particolare neurone (isomorfismo tra stimoli e neuroni di uscita).



Reti Neurali: apprendimento con rinforzo

- Apprendimento basato su premi e punizioni





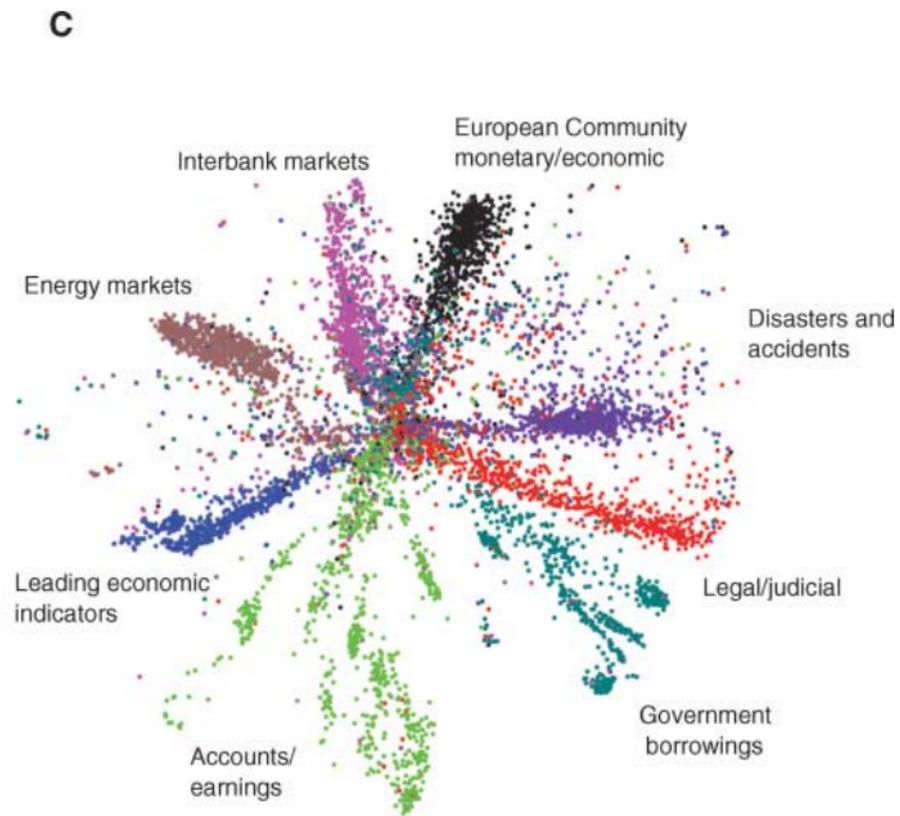
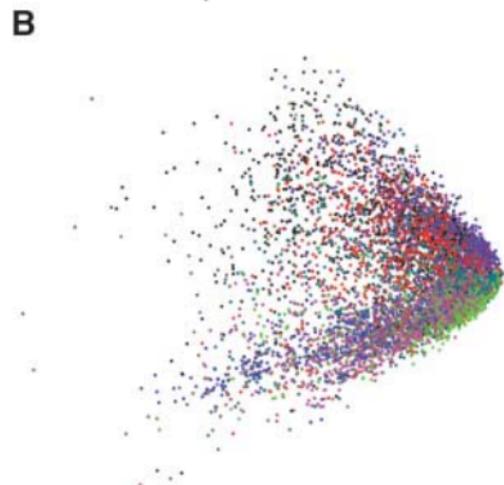
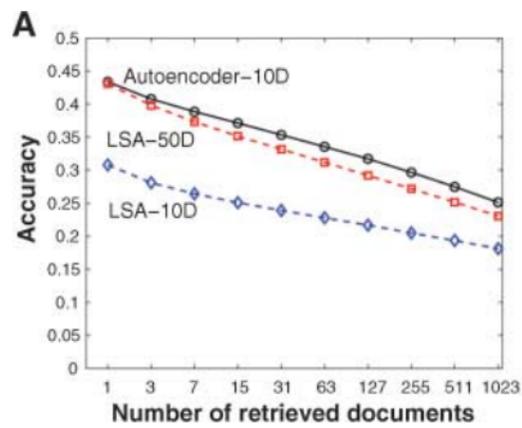
Reti Neurali profonde

- Esplosione della ricerca 10-12 anni fa.
- Risultati straordinari nei seguenti ambiti
 - Speech recognition
 - Machine translation
 - Computer vision
 - Object recognition in images
 - Video understanding
 - Natural language understanding
 - Game playing



Reti Neurali profonde

Document classification on the Reuter corpus





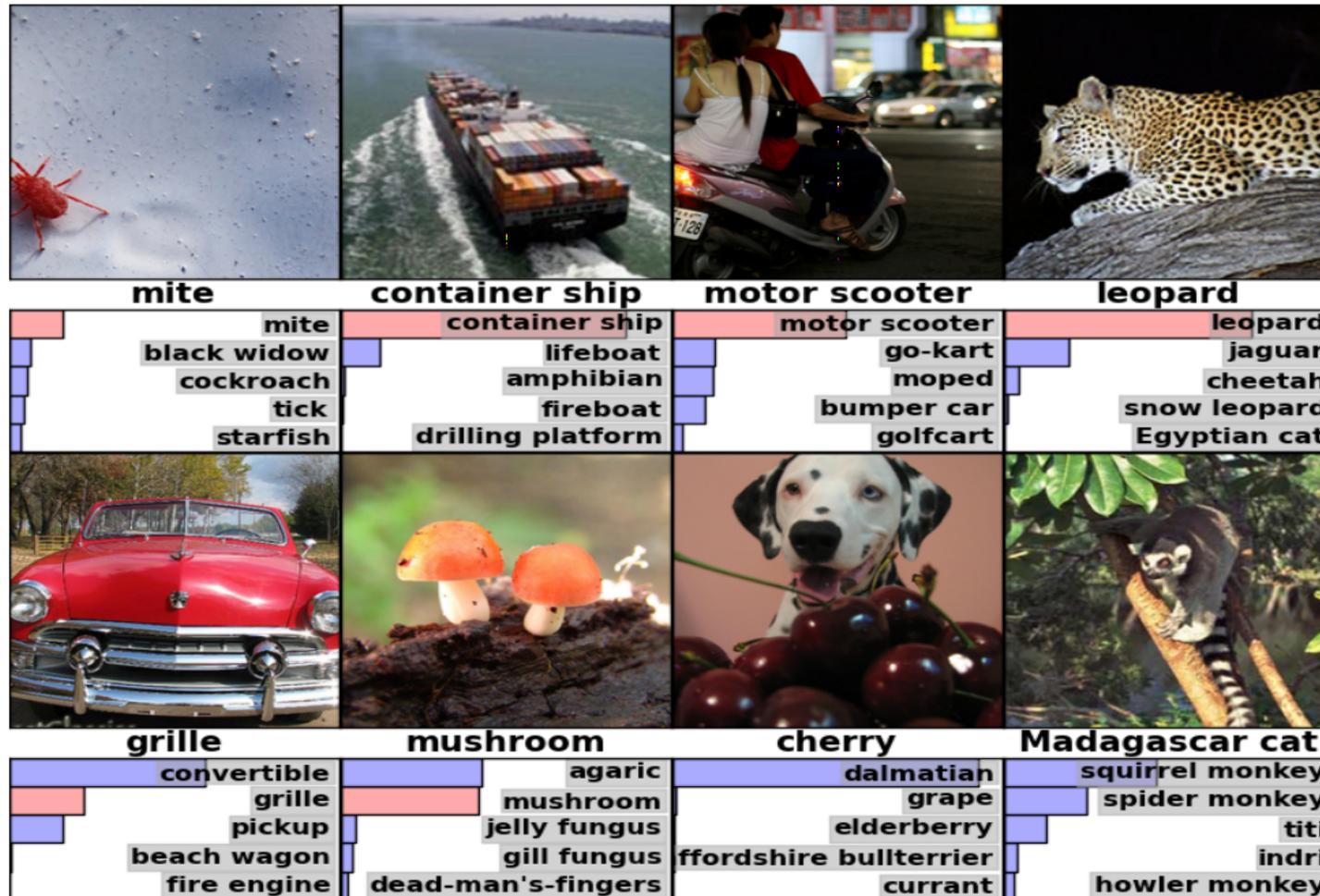
Reti Neurali profonde

- **ImageNet Challenge:** Universal image classifier
 - 14M images
 - 20k categories
 - Tagged via crowdsourcing
 - Labelled with the WordNet hierarchy



[Figure from vision.stanford.edu]

Reti Neurali profonde



[Figure from Krizhevsky et al., 2012]



Reti Neurali profonde





Tecnologie per l'ottimizzazione

- Combinatorial Optimization problems
 - Consideriamo problemi discreti NP-hard
- Minimizzare (Massimizzare) una funzione obiettivo di molte variabili soggette a vincoli
 - Matematici
 - Simbolici
- Aree di applicatione
 - Allocazione di risorse, scheduling, planning, routing, sequenziamento, design, configurazione....



Tecnologie per l'ottimizzazione

- Modello del problema
 - *Variabili decisionali*
 - *Vincoli tra variabili*
 - *Matematici*
 - *Logico-simbolici*
- Risoluzione del problema
 - *Propagazione di vincoli*
 - *Ricerca nello spazio delle soluzioni*
 - *Branch & Bound e varianti per l'ottimizzazione*



Conclusioni

- Sistemi decisionali sono molto diversi tra loro
- Dipendono dal problema da risolvere
- Composizione tra strumenti predittivi e prescrittivi
- Prendere in considerazione tutti gli aspetti dei sistemi socio-tecnici
- Molto ancora da fare: scalabilità, tecnologie stabili e riusabili