Well done AI! Indeed, optimally done!
Parte 3
Il ruolo dell'ottimizzazione in AI

Laura Palagi
Department of Computer, Control, and Management Engineering
Antonio Ruberti

SAPIENZA
UNIVERSITÀ DI ROMA

ALMA-WednesdAIs – 15 luglio 2020

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

ALMA MATER RESEARCH INSTITUTE FOR HUMAN-CENTERED ARTIFICIAL INTELLIGENCE

Recap Machine Learning

Non si tratta di una macchina fisica,

ma di **modelli matematici e algoritmi** in grado di «riprodurre» il comportamento di un sistema «ignoto»

○ Uso Predittivo

Predire i valori di alcune grandezze o per assegnare ad una classe

○ Uso Descrittivo

Individuare le caratteristiche salienti e siano "human-interpretable"

https://cloud.google.com/products/ai/ml-comic-1

SAPIENZA
UNIVERSITÀ DI ROMA

Perché per la costruzione di modelli di ML ho bisogno di algoritmi avanzati di ottimizzazione ?

Arthur Samuel (1901-1990)

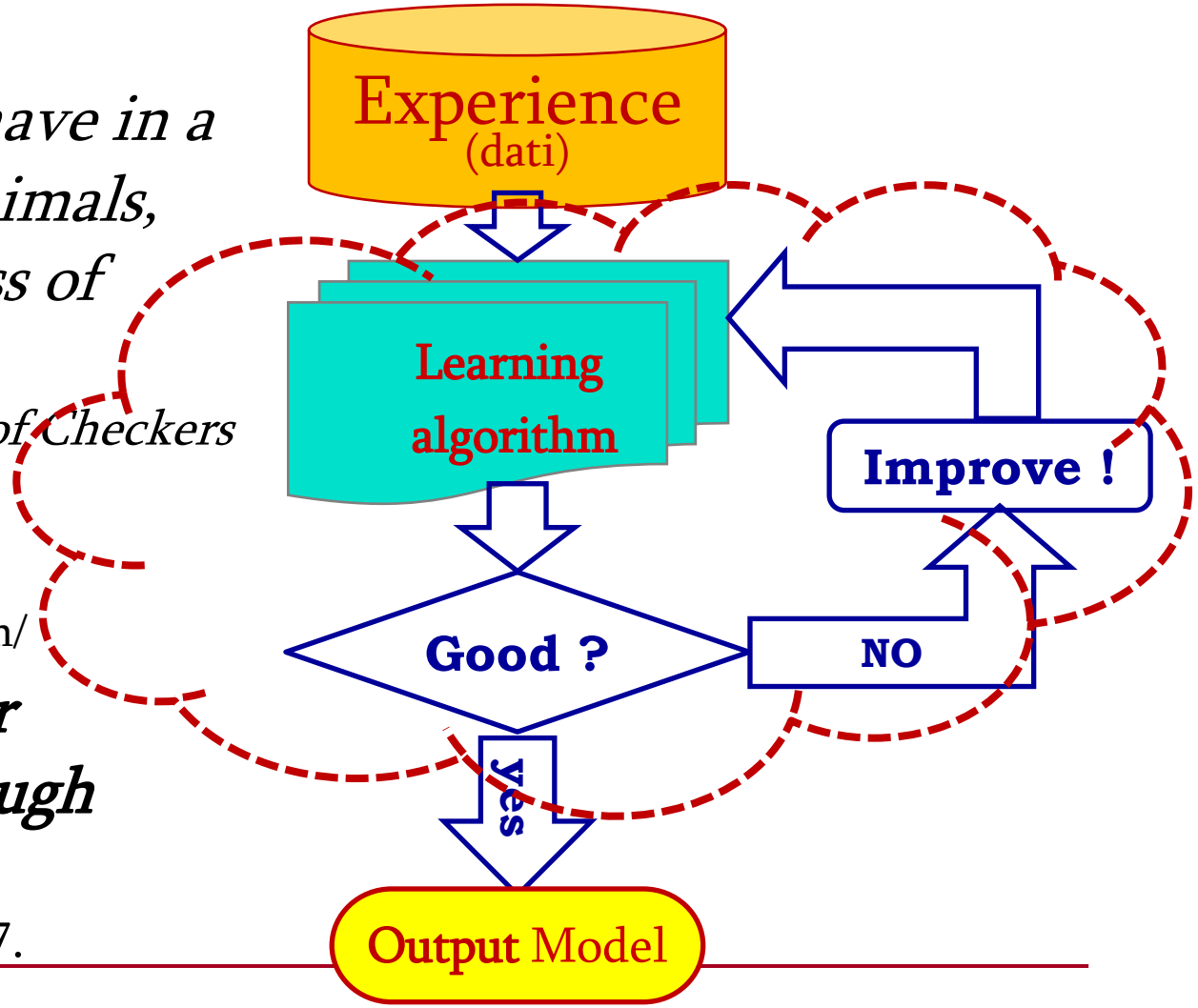
"programming of a digital computer to behave in a way which, if done by human beings or animals, would be described as involving the process of learning"

In Some Studies In Machine Learning Using the Game of Checkers ,1959

Tom Mitchell (1997) http://www.cs.cmu.edu/~tom/

"Machine Learning is the study of computer algorithms that improve automatically through experience"

in Machine Learning, Tom Mitchell, McGraw Hill, 1997.



Apprendimento Supervisionato

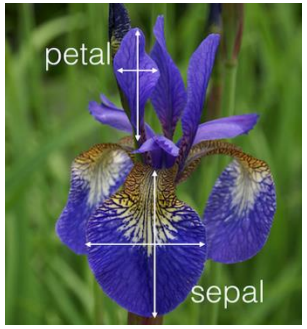
I dati sono descritti da attributi e tra gli attributi di ogni esempio c'è un valore di label/target che li caratterizza.

Gli attributi sono coppie (x, y) con $x \in R^n$ e y che può avere valori categorici (classificazione) o continui (regressione)

Iris Data Set

Download: Data Folder, Data Set Description

Abstract: Famous database; from Fisher, 1936



1. sepal length in cm
2. sepal width in cm
3. petal length in cm
4. petal width in cm

$$x \in R^4$$

Tre Classi: $y \in \{1,2,3\}$
-- Iris Setosa
-- Iris Versicolour
-- Iris Virginica

1. sepal length in cm
2. sepal width in cm
3. petal length in cm
4. petal width in cm

$x \rightarrow$ **Unknown relationship** $\rightarrow y$

-- Iris Setosa
-- Iris Versicolour
-- Iris Virginica

Apprendimento supervisionato

Past Label data (x^i, y^i)

1	5.1,3.5,1.4,0.2,Iris-setosa
2	5.0,3.3,1.4,0.2,Iris-setosa
3	7.0,3.2,4.7,1.4,Iris-versicolor
4	5.8,2.7,5.1,1.9,Iris-virginica
5	5.9,3.0,5.1,1.8,Iris-virginica

1. sepal length in cm
2. sepal width in cm
3. petal length in cm
4. petal width in cm

x^i

Unknown relationship

y^i

-- Iris Setosa
-- Iris Versicolour
-- Iris Virginica

Learning algorithm

x^i

New data

Predictive model

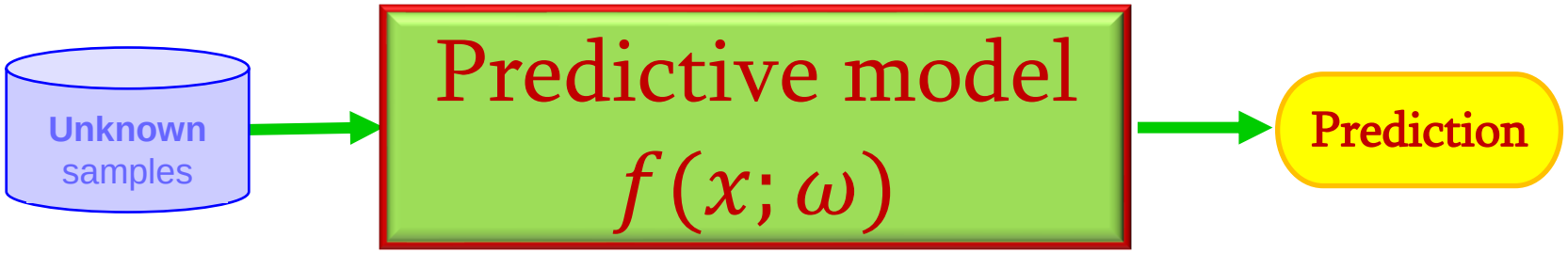
$\tilde{y}^i$

Prediction

It is a Setosa !

Obiettivo modello predittivo

Obiettivo: ai nuovi campioni sia assegnato un valore il più preciso possibile.

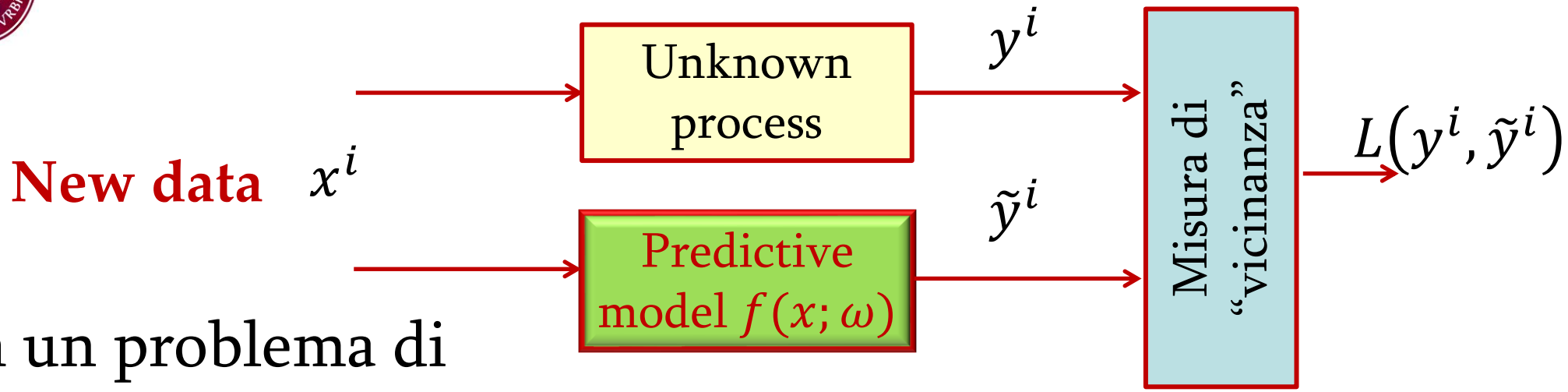


Proprietà di generalizzazione: si misura con la capacità del modello di dare la "giusta risposta" su nuove "istanze" (out-of-sample generalization).

Per **misurare la performance** di un modello predittivo è necessario definire una misura quantitativa di «somiglianza» (dipende dal Task).

Performance measure P - classificazione

Ammettendo di conoscere la risposta corretta y^i



In un problema di classificazione:

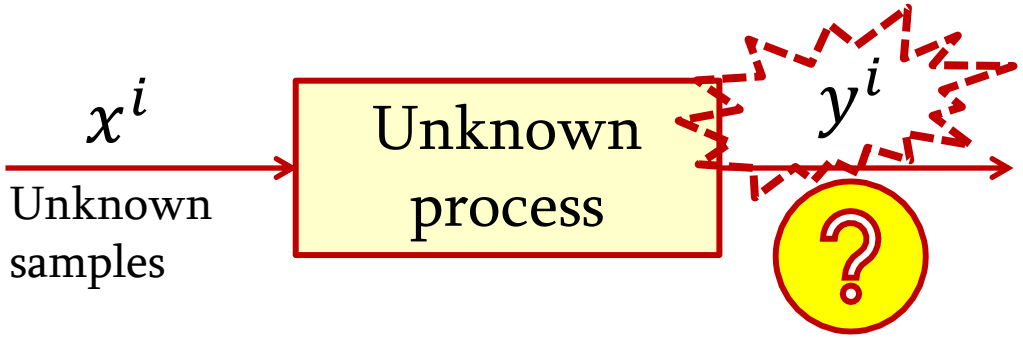
0-1 loss function

$$L(y, \tilde{y}) = \begin{cases} 0, & y = \tilde{y} \\ 1, & y \neq \tilde{y} \end{cases}$$

Obiettivo: **minimizzare *error rate*** cioè la proporzione di valori sbagliati in uscita rispetto a quanti se ne presentano

$$error\ rate = \frac{\sum_{i=1}^{P} L(y^i, \tilde{y}^i)}{P}$$

Obiettivo effettivo vs obiettivo surrogato



$$R(f,\omega) = \int L(y; f(x;\omega))dp(x,y)$$

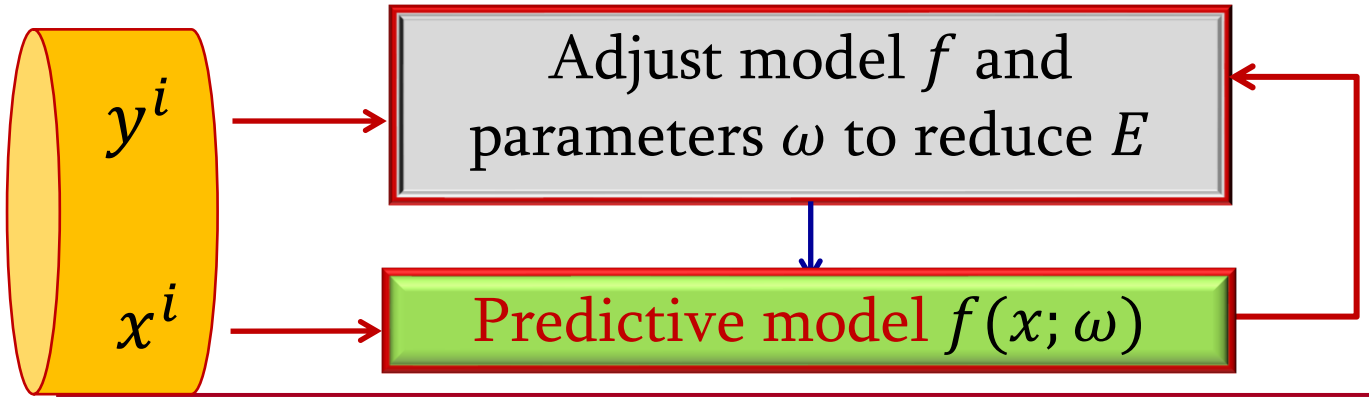
Misura di performance surrogata

Errore E sui dati disponibili

Empirical Risk Minimization

$$E(f,\omega) = \frac{1}{P}\sum_{i=1}^{P} L(y^i, f(x^i;\omega))$$

Available data



$f(x^i;\omega)$

"An evaluation function (also called objective function or scoring function) is needed to distinguish good classifiers from bad ones. **The evaluation function used internally by the algorithm may differ from the external one** that we want the classifier to optimize, for ease of optimization"

Pedro Domingos, A Few Useful Things to Know about Machine Learning, 2012

ideal

$$(f^*, \omega^*) = \arg\min_{f,\omega} R(f, \omega)$$

practical

$$(f_{emp}, \omega_{emp}) = \arg\min_{f,\omega} E(f, \omega)$$

Se $E = 0$, $y^i = f(x^i; \omega)$

Che succede su nuovi dati ?

E' necessario bilanciare la complessità del modello e l'accuratezza

Modelli troppo complessi sono in grado di minimizzare l'errore di training ma tendono a generalizzare

OVERFITTING
Modello impara a memoria

Modelli troppo semplici possono non essere in grado di descrivere il fenomeno

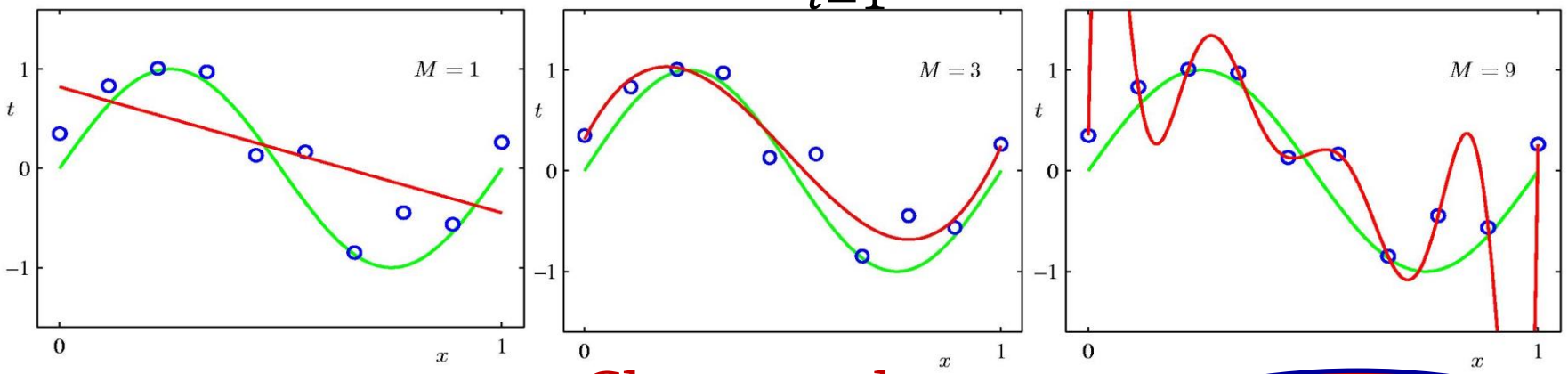
UNDERFITTING
Modello non ha studiato

Regressione polinomiale (fissata f varia ω)

$$f(x;w) = w_o + w_1 x + w_2 x^2 + \cdots + w_n x^n$$

Minimizza MSE

$$E(w) = \frac{1}{P}\sum_{i=1}^{P}\left(f\left(x^i;w\right) - y^i\right)^2$$



Underfitting

Che succede su nuovi dati ?

Overfitting

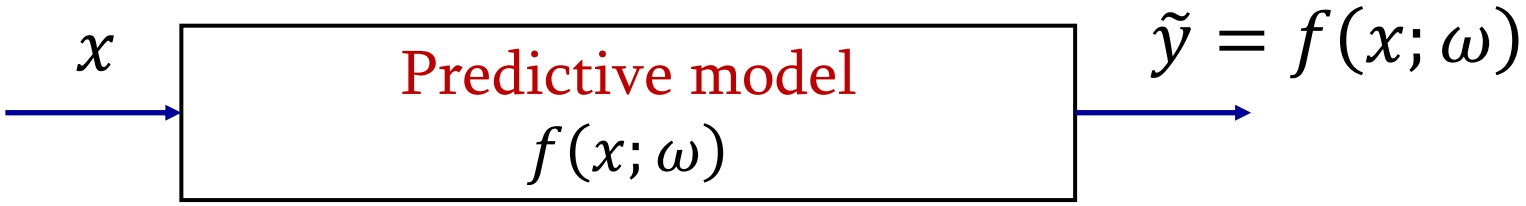
Individua la tipologia di funzioni

il grado ***n***

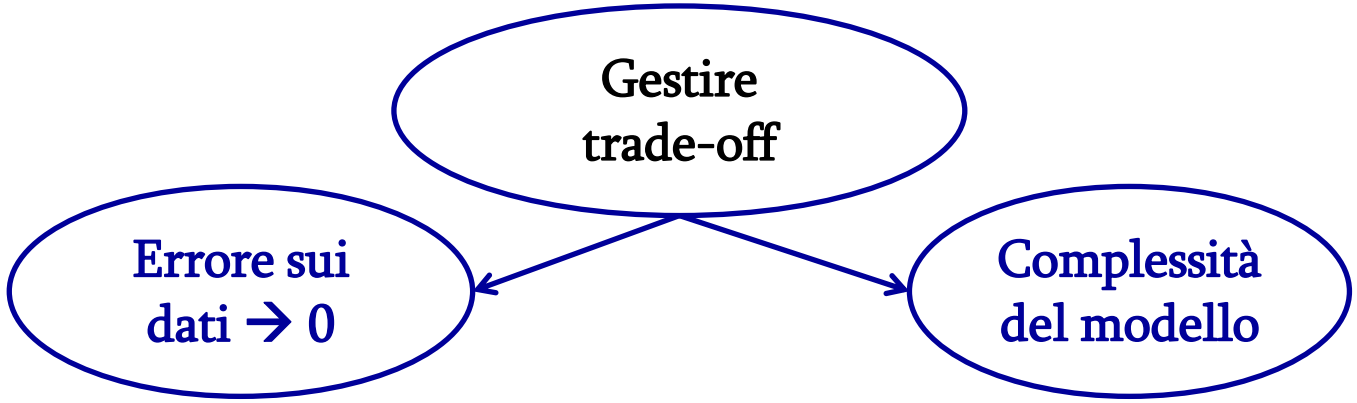
il coefficiente ***w***

Individua una specifica *f*

Riassumendo



Addestrare significa scegliere $f(\cdot;\omega)$ Come ?



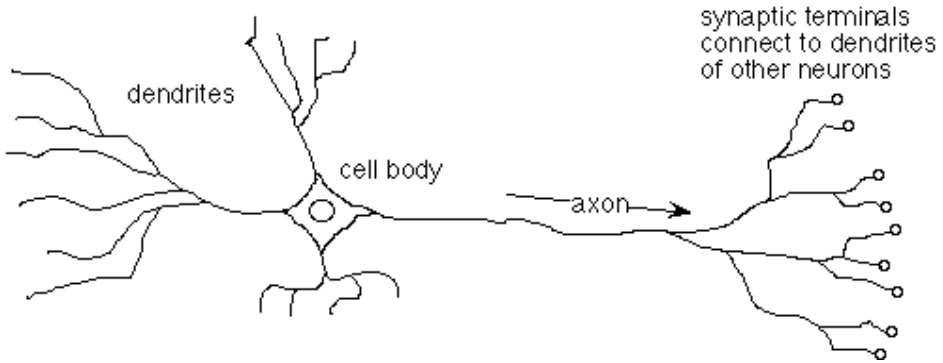
Ottimizzazione di un modello PRESCRITTIVO per la scelta di f, ω

Minimizzazione del
- rischio empirico regolarizzato
- rischio strutturale

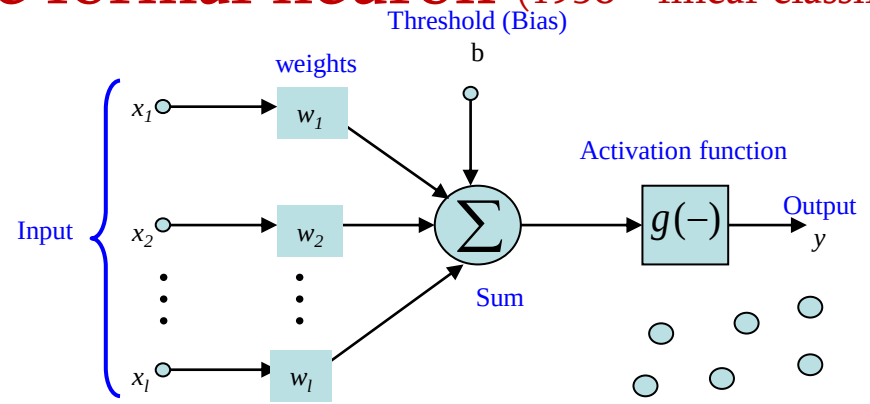
$$\min_{\omega} E(f,\omega) + \lambda \, \|\omega\|_p$$

Un po' di storia

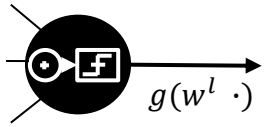
From the biological neuron ….

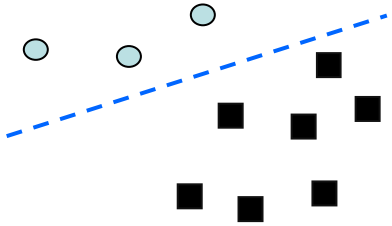


to the formal neuron (1958 – linear classifier)

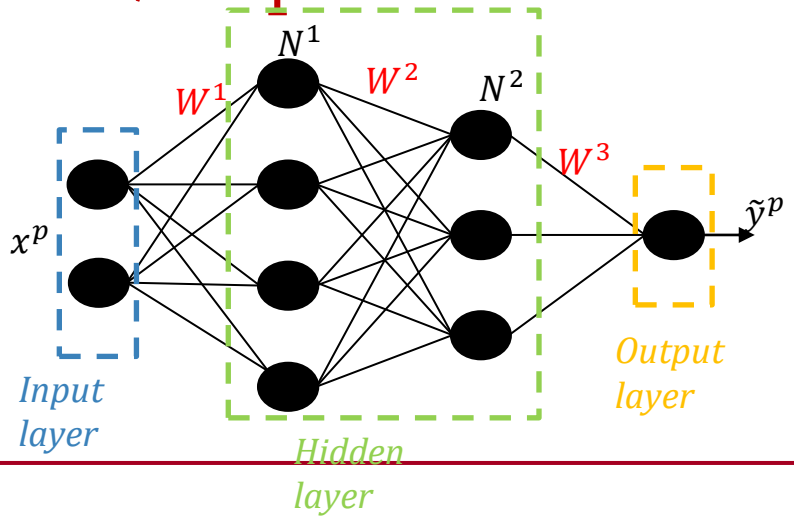


to NON linear classifier

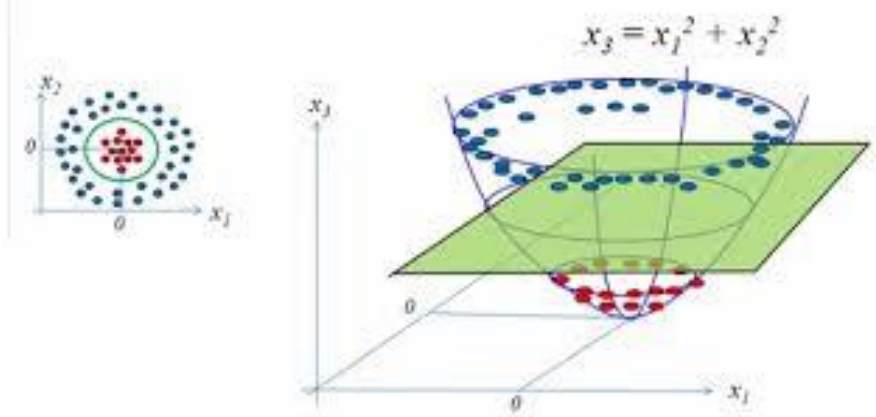




shallow/deep networks 1958-90/2006



Kernel methods 1990 - 2010



https://www.youtube.com/watch?v=ffF8UnbheLk

Optimization

$$\min_{\omega} E(f,\omega)+\lambda \|\omega\|_p$$

"Finally, we need a method to search among the classifiers in the language for the highest-scoring one.

The choice of optimization technique is key to the efficiency of the learner, and also helps determine the classifier produced if the evaluation function has more than one optimum.

It is common for new learners to start out using off-the-shelf optimizers, which are later replaced by custom-designed ones."

Table 1. The three components of learning algorithms.

Representation	Evaluation	Optimization
Instances	Accuracy/Error rate	Combinatorial optimization
K-nearest neighbor	Precision and recall	Greedy search
Support vector machines	Squared error	Beam search
Hyperplanes	Likelihood	Branch-and-bound
Naive Bayes	Posterior probability	Continuous optimization
Logistic regression	Information gain	Unconstrained
Decision trees	K-L divergence	Gradient descent
Sets of rules	Cost/Utility	Conjugate gradient
Propositional rules	Margin	Quasi-Newton methods
Logic programs		Constrained
Neural networks		Linear programming
Graphical models		Quadratic programming
Bayesian networks		
Conditional random fields		

Quoting: Pedro Domingos, *A Few Useful Things to Know about Machine Learning*

Algoritmi di ottimizzazione e AI

Cosa significa «addestrare» un modello di ML ?

- Definire una funzione obiettivo, eventuali vincoli
- Definire un algoritmo per risolvere il problema di ottimizzazione

Cosa rende il problema di ottimizzazione peculiare ?

- Molti sample P (big data)
- Molte variabili

Necessità di ripetere il processo di ottimizzazione molte volte per gestire in fase di addestramento la generalizzazione (k-fold cross validation)

Algoritmi di ottimizzazione standard possono fallire (sia per il tempo computazionale che di memoria necessari)

Algoritmi di ottimizzazione e AI

Sono molte le domande

- Come scegliere il modello ?
- Come scegliere un algoritmo ?
- Come verificare il corretto funzionamento di un algoritmo ?
- La soluzione è unica ? come scegliere tra le soluzioni?
- Se il risultato non è conforme alle aspettative, qual è la causa: il modello, l'algoritmo, la soluzione scelta ?

ML algoritms requires the solution of challenging optimization problems

- Sfruttare la struttura particolari
- Disponibilità di dati distribuita

SAPIENZA
UNIVERSITÀ DI ROMA

DEEP NETWORK
Highly non convex unconstrained minimization

Difficulties

- Highly non linear
- Plateau /local minimizer
- Heavy computation of gradient
- Heavy computation of objective function
- Difficult to manage overfitting

Structure

- Decompose
  - w.r.t samples

  Online methods
  - wr.t. layer

  Coordinatewise optimization

SUPPORT VECTOR MACHINEs
Convex constrained minimization

Difficulties

- Heavy computation of objective function and gradient
- Difficult to manage overfitting

Structure

- Dual formulation
- Sample decomposition

Learning Machine Learnig comics from google https://cloud.google.com/products/ai/ml-comic-1

Algoritmi & tecnologia

Dimitris Bertsimas

Machine learning and statistics via a modern optimization lens

https://www.youtube.com/watch?v=7w9aRrYgGEs

Progress of MIO

- Speed up between CPLEX 1.2 (1991) and CPLEX 11 (2007): 29 times
- Gurobi 1.0 (2009) comparable to CPLEX 11
- Speed up between Gurobi 1.0 and Gurobi 6.5 (2015): 48.7 times
- Total speedup 1991-2015: 1,400,000 times
- A MIO that would have taken 16 days to solve 25 years ago can now be solved on the same 25-year-old computer in less than one second.
- Hardware speedup: $10^{5.5}$= 570,000 times
- Total Speedup: **805 Billion times!**
- A MIO that would have taken 26,000 years to solve 25 years ago can now be solved in a modern computer in less than one second.

Ha spostato in avanti la frontiera dei problemi risolvibili e dunque anche degli ambiti di utilizzo

Algoritmi

Tecnologia

SAPIENZA
UNIVERSITÀ DI ROMA

Sono gli algoritmi il nuovo petrolio ?

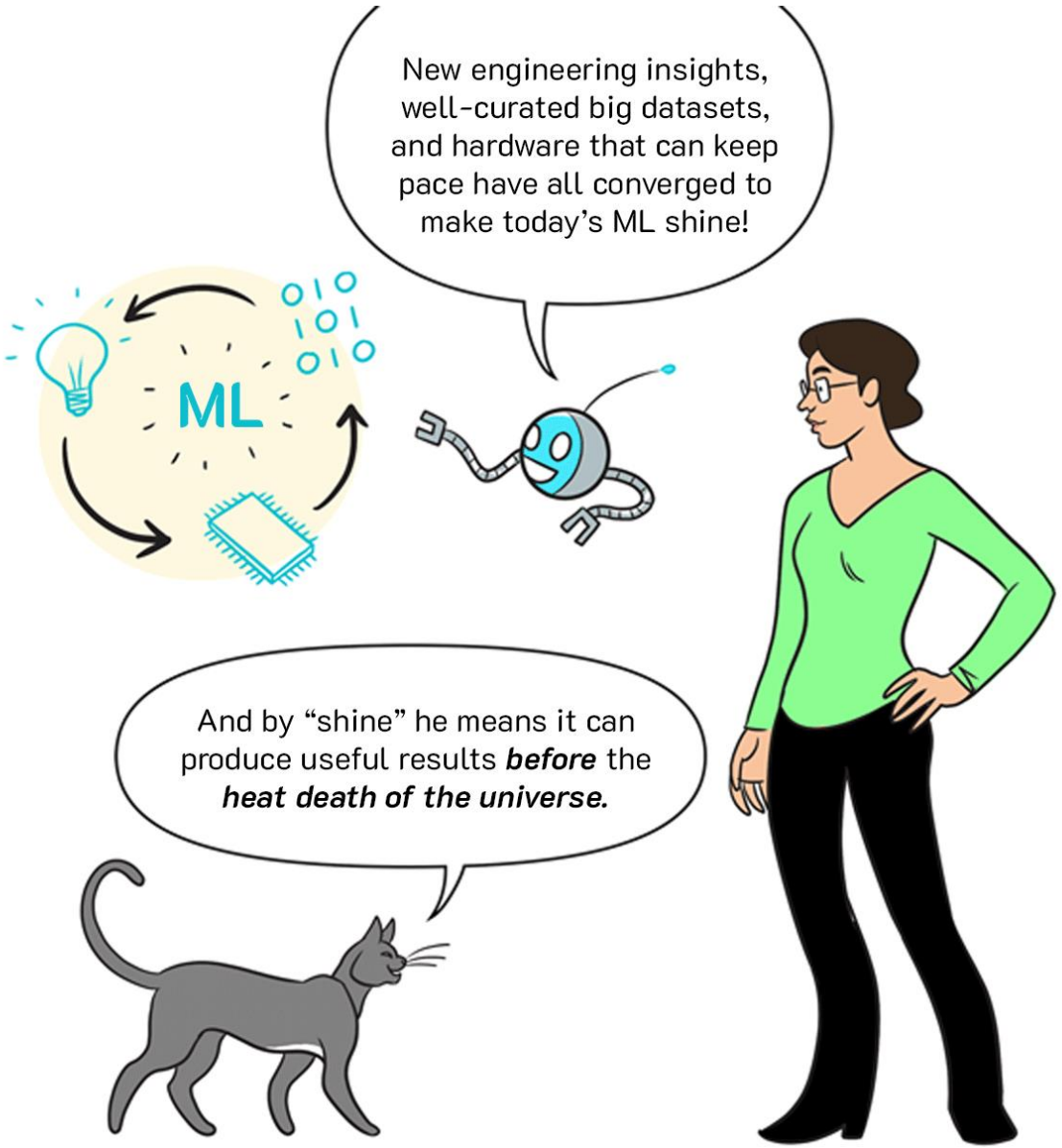
«Qualcuno dice: ***i dati sono il petrolio del futuro***. Mi sembra più corretto dire: così come il petrolio grezzo ha meno valore prima della sua trasformazione, così i dati, *anche i Big-Data*, sono valorizzati solo quando un algoritmo li utilizza e li trasforma in informazioni di qualità per trovare soluzioni a problemi rilevanti.

Saranno gli algoritmi, *la capacità di concepirli, farli funzionare e certificarne il comportamento corretto* ***a creare la ricchezza futura di nazioni e aziende***. Confondere la **qualità** dei metodi di ottimizzazione e degli algoritmi con la **quantità** dei dati a disposizione è un errore che potrebbe avere conseguenze molto negative.»

Prof Antonio Sassano – Presidente FUB

in occasione Audizione IX COMMISSIONE (TRASPORTI, POSTE E TELECOMUNICAZIONI)

luglio 2019



Grazie per l'attenzione

https://cloud.google.com/products/ai/ml-comic-1