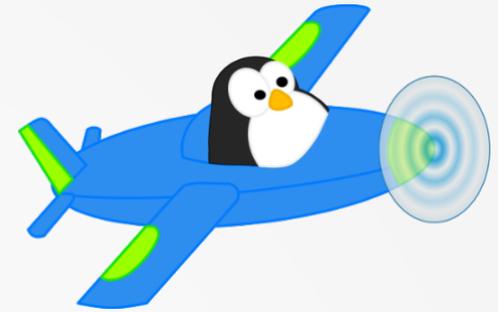


LINUX WAR

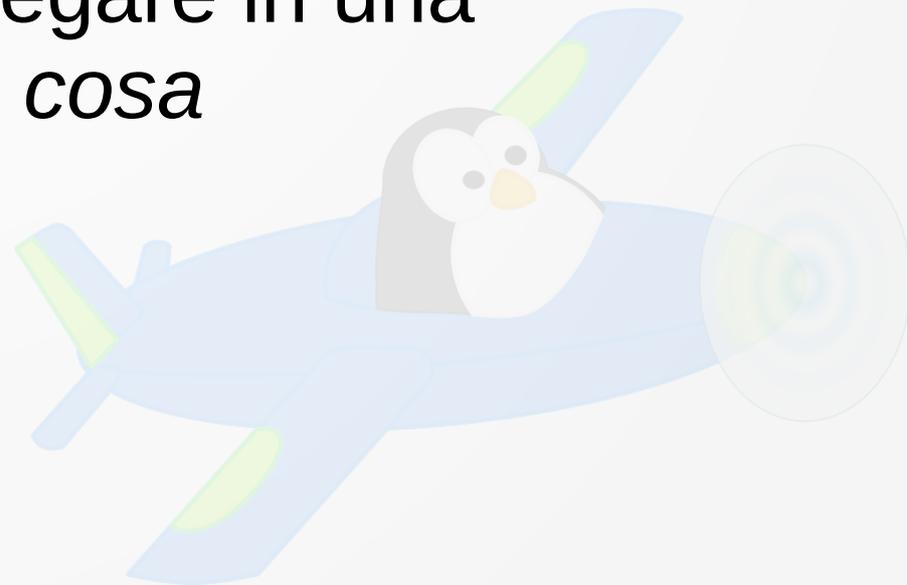


Introduzione al machine learning

di Riccardo Macoratti

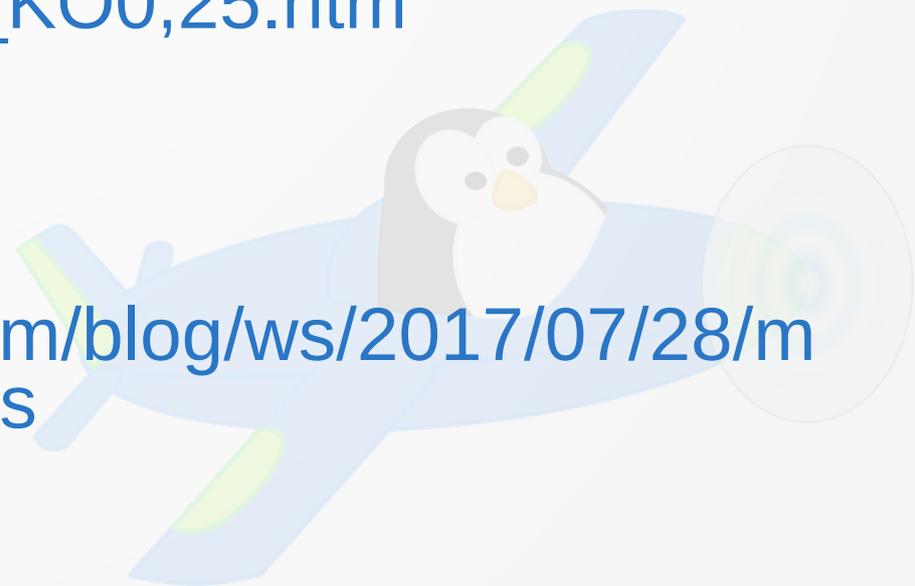
PremessaONA

- Sono uno studente, non un esperto di *machine learning*
- Non ho mai implementato programmi di apprendimento in ambito lavorativo
- Il *come* non è facile da spiegare in una presentazione, parlerò del *cosa*



Perché?

- Perché parlare di *machine learning*?
 - ↳ *Top buzzword* legate alla *data science*:
<https://www.gartner.com/newsroom/id/3812063>
 - ↳ Quanto guadagna un *machine learning engineer*?
https://www.glassdoor.com/Salaries/machine-learning-engineer-salary-SRCH_KO0,25.htm
- Non basta?
 - ↳ È oggettivamente *cool*:
<https://www.wordstream.com/blog/ws/2017/07/28/machine-learning-applications>



Quando

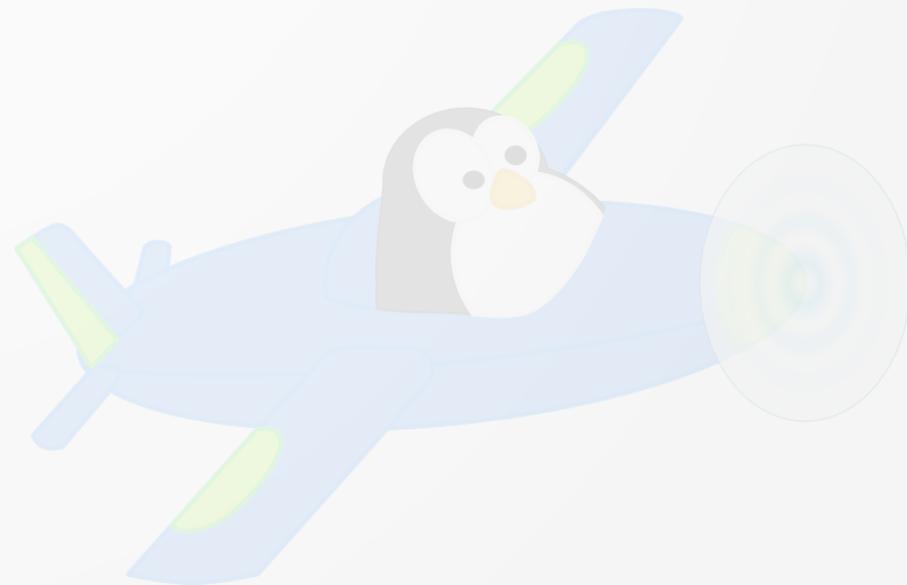
- È un *trend* recente?
 - ↘ “*Machine learning*” coniato nel 1959 da Arthur Lee Samuel
 - ↘ Per capirsi, negli anni ‘50 c’era questo https://en.wikipedia.org/wiki/IBM_701
- Sviluppato solo “ora” (anni ‘90)
 - ↘ I mezzi lo consentono (CPU, GPU, ...)
 - ↘ Sono abbastanza a buon mercato



IBM 701 da [Wikipedia](#)

Ambiti

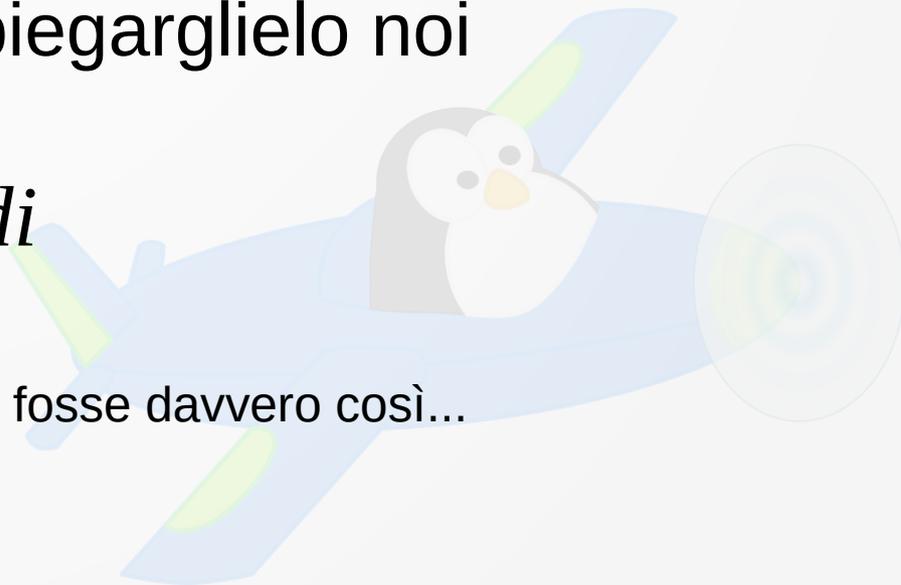
- Non è solo dominio dell'informatica
 - ↘ Anzi, più della *data science*, la “scienza dei dati” e quindi della statistica
- Ogni campo gli dà un nome originale e complicato
 - ↘ Statistica computazionale
 - ↘ Algoritmi adattivi
 - ↘ Sistemi dinamici
 - ↘ Analisi predittiva
 - ↘ ...



Machine learning?

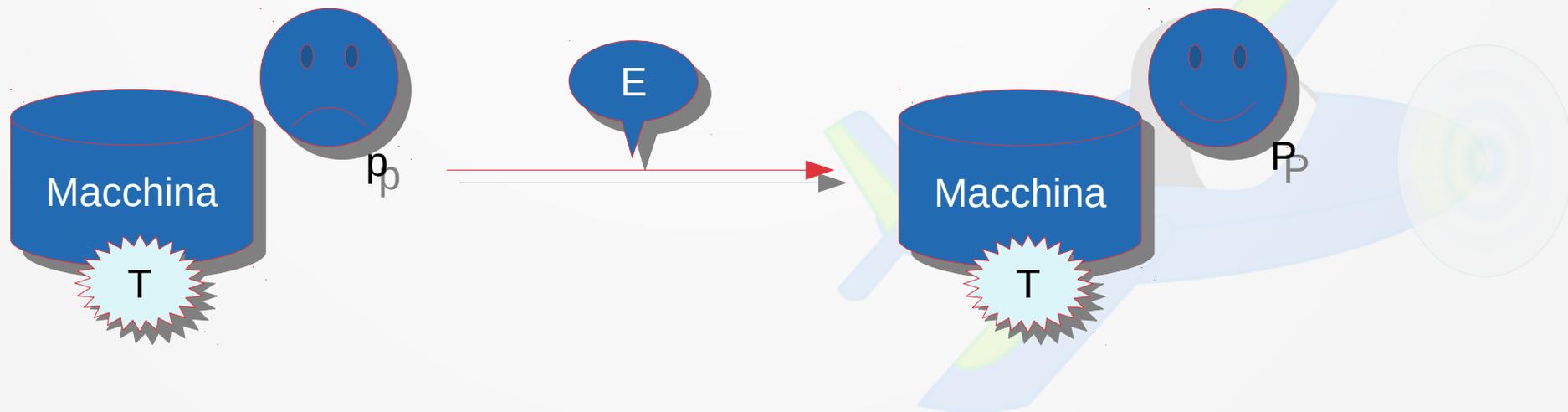
- Noi la chiameremo semplicemente *apprendimento automatico*
 - ↪ *Apprendimento*: insomma, c'è qualcosa da imparare
 - ↪ *Automatico*: ci pensa la macchina a capire come imparare, non dobbiamo spiegarglielo noi

“fornisce ai computer l’abilità di apprendere senza essere stati esplicitamente programati” Magari fosse davvero così...



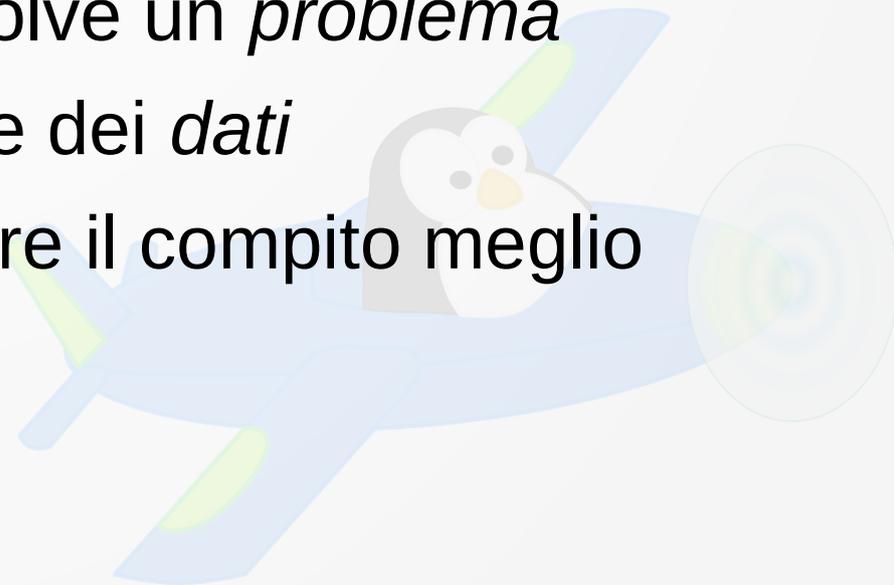
Finalmente, cos'è?

- Diamo una definizione operativa:
“si dice che una macchina apprende dall'esperienza E , con riferimento ad un compito T e con misurazione della performance p , se le sue performance nel compito T migliorano l'esperienza con E e diventano P ”



Disclaimer

- Ti sembra di saperne quanto prima, vero?
 - ↪ Non è vero, qualcosa abbiamo imparato
- Sappiamo che:
 - ↪ C'è una macchina, quindi un *programma*
 - ↪ Svolge un compito, cioè risolve un *problema*
 - ↪ Fa un'esperienza, cioè vede dei *dati*
 - ↪ Grazie a questi, può svolgere il compito meglio



Un esempio

- Meglio fare un esempio
 - Siamo Zuckerberg, dobbiamo aiutare gli utenti a *taggare* le foto
 - Il **programma** è Facebook
 - Il **problema** è dire se un'immagine contiene un volto umano (risposta SÌ o NO)
 - Le **esperienze** sono le foto già taggate manualmente dagli utenti → sappiamo già che c'è un volto
 - Guardandone tante, miglioriamo l'**output**, cioè diciamo SÌ quando c'è veramente un volto e NO quando non c'è

SÌ e NO

Output del programma



SÌ



Vero positivo

GNU+Linux



Verità assoluta

NO



Falso negativo



SÌ

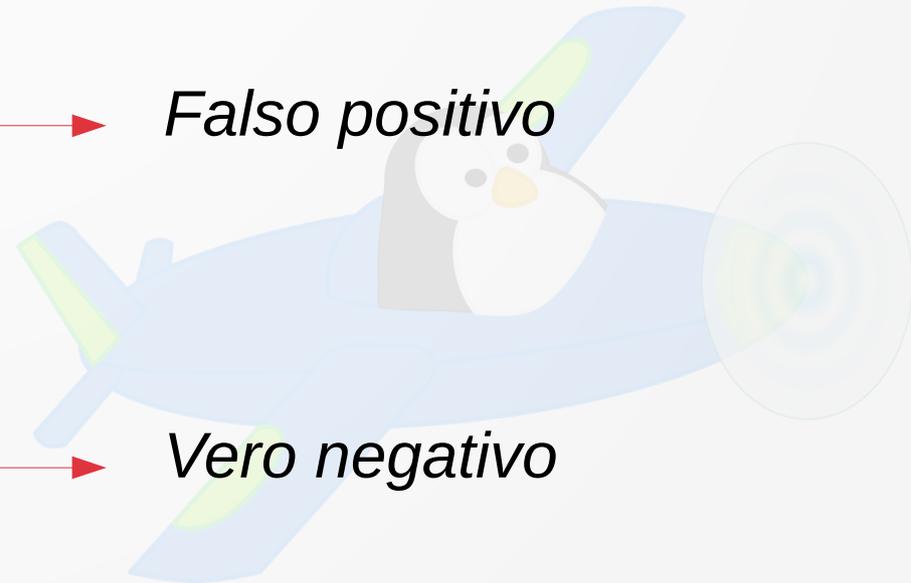


Falso positivo

NO



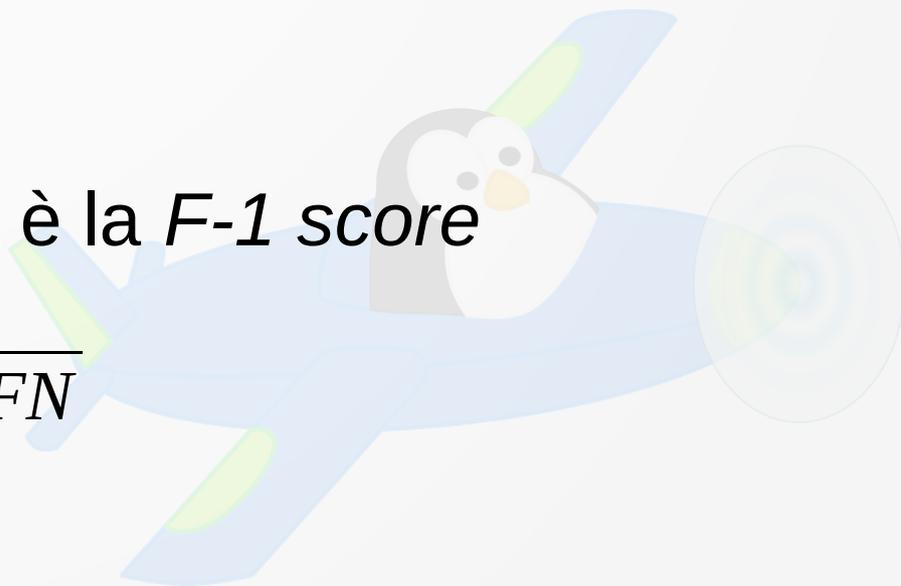
Vero negativo



Performance

- Quindi come si valutano le performance?
 - ↪ Valuto l'errore commesso
 - ↪ Commistione di veri / falsi e positivi / negativi
 - ↪ Cose complicatissime, ma sono da citare
 - $Sensitivity = \frac{VP}{VP+FN}$
 - $Precision = \frac{VP}{VP+FP}$
 - ↪ Un'altra formula usatissima è la *F-1 score*

$$\frac{2VP}{2VP+FP+FN}$$

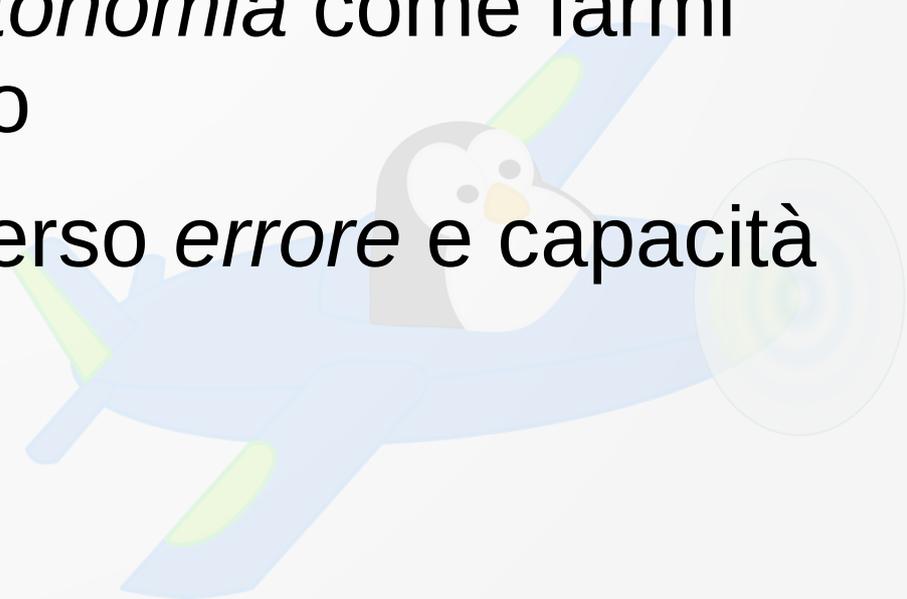


Riassunto

- Vediamo cosa è stato fatto:
 - ✦ Ho preso una serie di coppie $(Im_1 \rightarrow sì), (Im_2 \rightarrow no), (Im_3 \rightarrow sì), \dots$
 - ✦ La macchina le ha usate per imparare una *mappatura* del tipo immagine \rightarrow *vero / falso*
 - ✦ Ho testato quanto la macchina è capace di dirmi la *verità*
 - ✦ Adesso, se le faccio analizzare un'immagine nuova, mai vista, mi sa dire *vero* o *falso* con una maggiore precisione

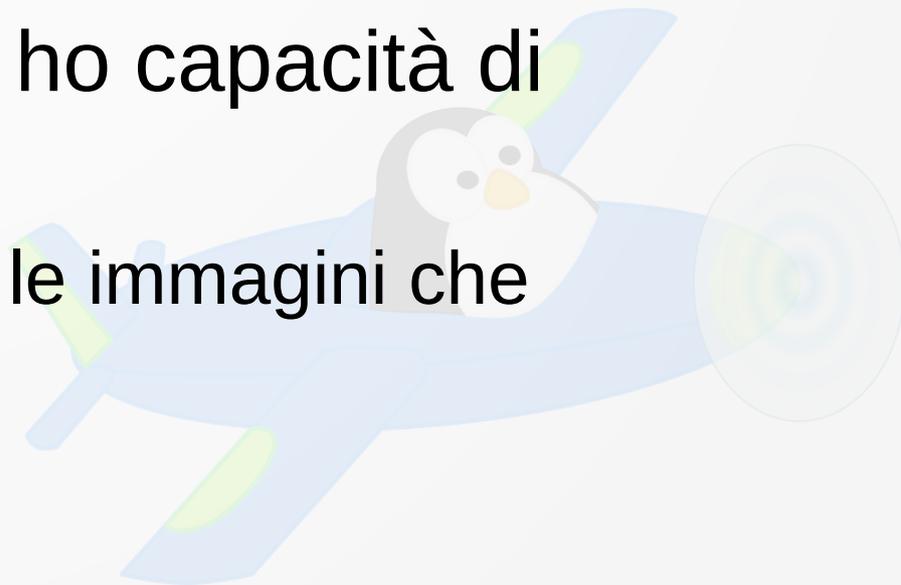
Il *machine learning*, quindi

- Ho una macchina, che prende in input *qualcosa* e mi restituisce *qualcos'altro*
- Le fornisco un input e, allo stesso tempo, l'*output che desidero avere*
- La macchina impara *in autonomia* come farmi avere l'*output che desidero*
- Valuto la macchina, attraverso *errore* e capacità di *generalizzazione*



Generalizzazione

- Ho fatto il passo più lungo della gamba?
 - ↪ Ho imparato davvero a dire se un'immagine è un volto?
 - ↪ Oppure ho barato e ho imparato benissimo come sono fatte le poche immagini iniziali?
- Valuto l'*overfitting*, cioè se ho capacità di *generalizzazione*
 - ↪ A me non serve catalogare le immagini che conosco, ma quelle nuove



Mapping + Generalizzazione

→ Perciò l'apprendimento automatico è:

- ↳ Imparare un *mapping*, una *funzione matematica*, che dato un input mi dà un certo output

1 → a

2 → b

3 → c

...

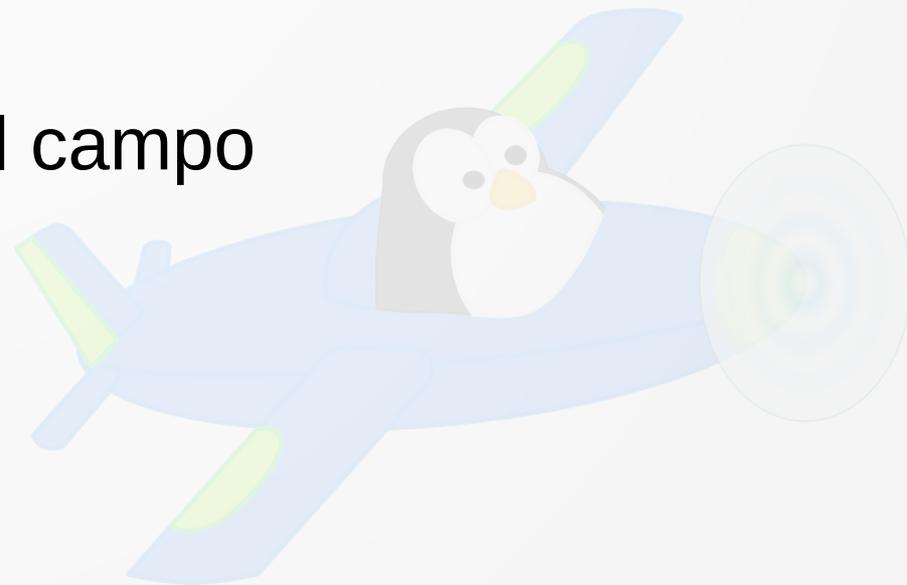
- ↳ Ma saper dire qualcosa di *probabilmente giusto* anche in caso l'input non sia conosciuto

-1 → ?

3.5 → ?

Classificazione

- L'esempio appena visto è un esempio di *classificazione binaria*
 - ↪ Per ogni immagine viene detto sì o no su qualche *predicato*
 - ↪ Non è l'unica “modalità” di apprendimento automatico
 - ↪ C'è ancora molto studio nel campo



Machine learning

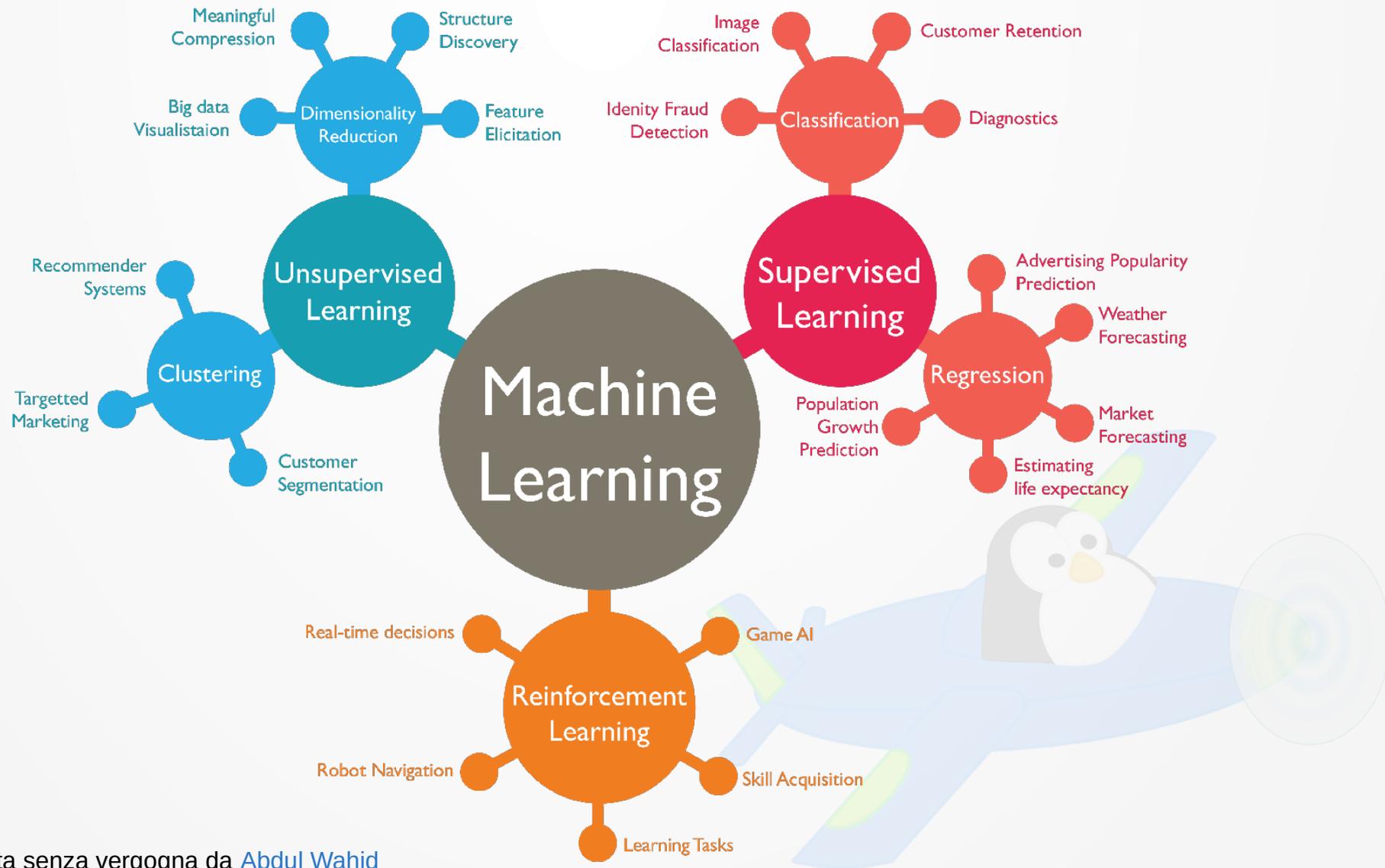


Immagine rubata senza vergogna da [Abdul Wahid](#)

Tipi di algoritmi

→ Supervisionati

- ✦ Fornisco input e output desiderato
- ✦ L'obiettivo è estrarre una regola generale

→ Non supervisionati

- ✦ Fornisco solo l'input
- ✦ La macchina dovrebbe trovare una struttura nei dati

→ Con rinforzo

- ✦ Si cerca di raggiungere un obiettivo con un *insegnante* che dice se l'obiettivo è stato raggiunto

Tipi di output

→ Classificazione

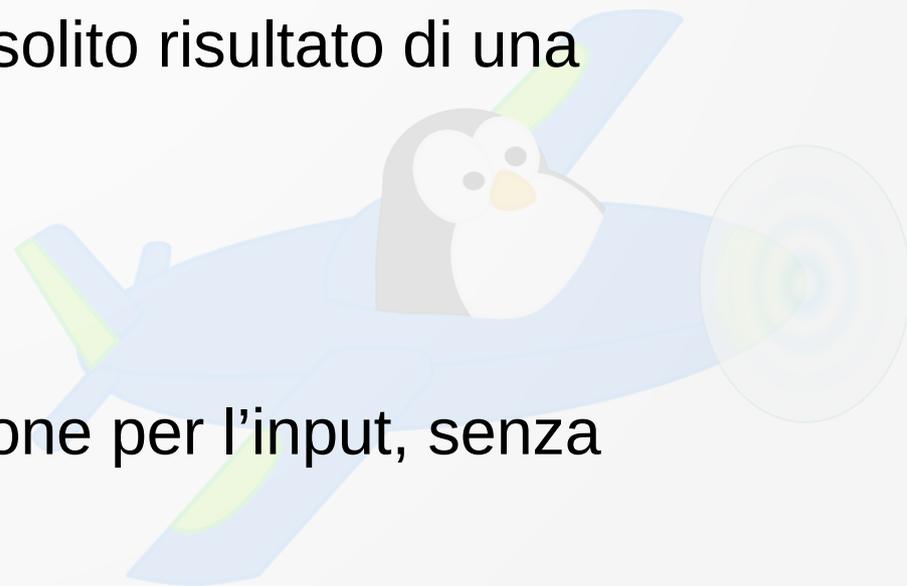
- ✦ Output *discreto*
- ✦ La macchina cataloga l'input con *etichette* già note

→ Regressione

- ✦ Output *continuo*
- ✦ L'output fornito è un numero, di solito risultato di una predizione

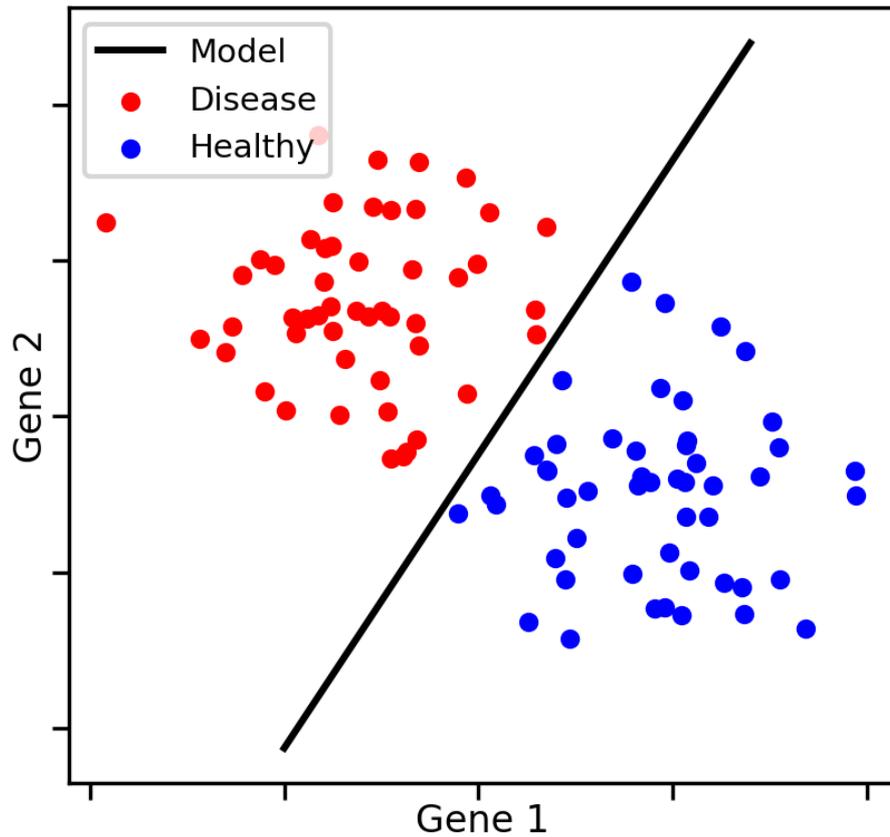
→ Clustering

- ✦ Output *discreto*
- ✦ La macchina trova la catalogazione per l'input, senza etichette già note



Classificazione vs Regressione

Classification



Regression

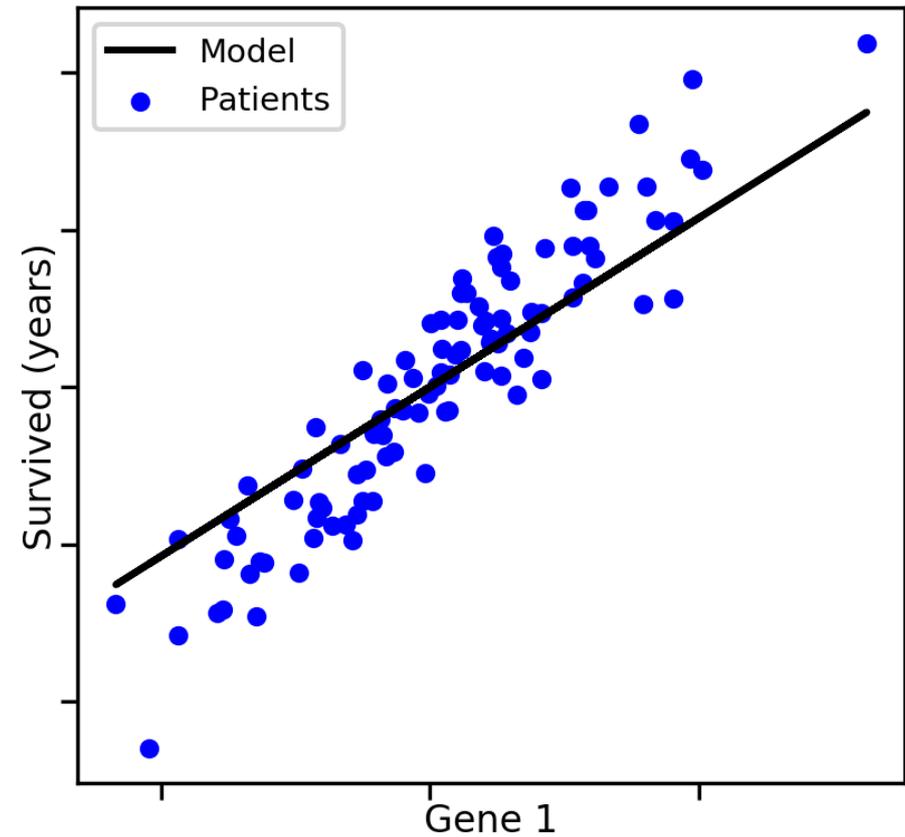
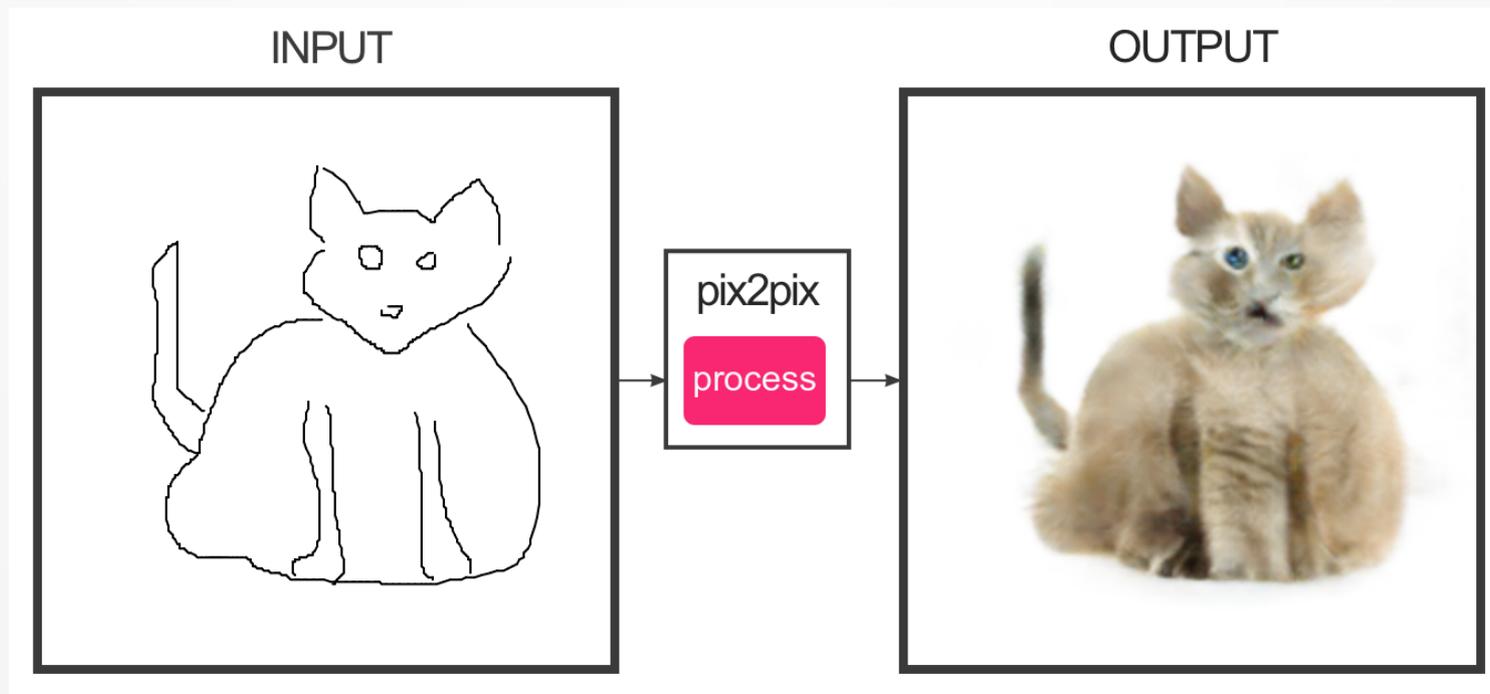


Immagine trafugata senza rimpianti da [Microbiome Summer School 2017](#)

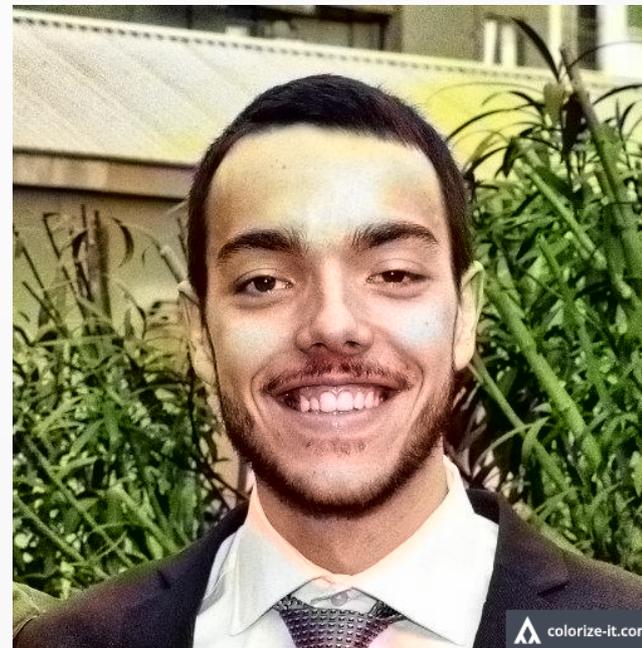
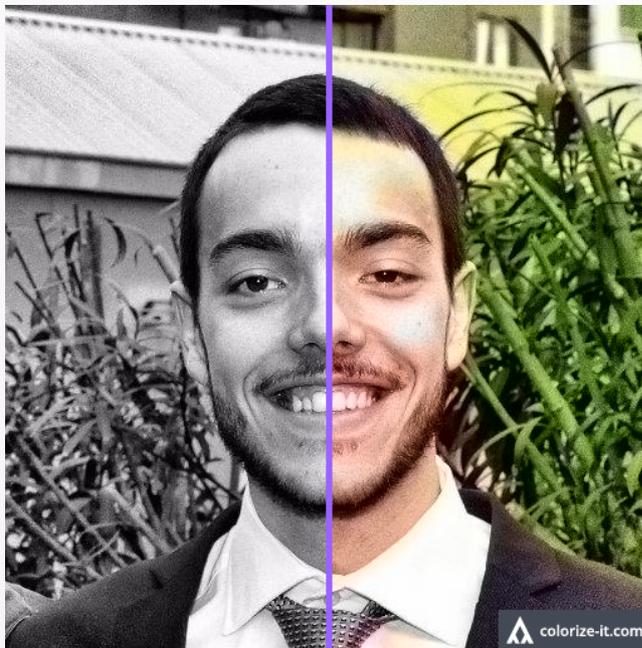
Relax slide (1)

- Rilassiamoci un po' prima di proseguire
 - ↘ Gatti e *machine learning*?
 - ↘ [edges2cats](#)



Relax slide (2)

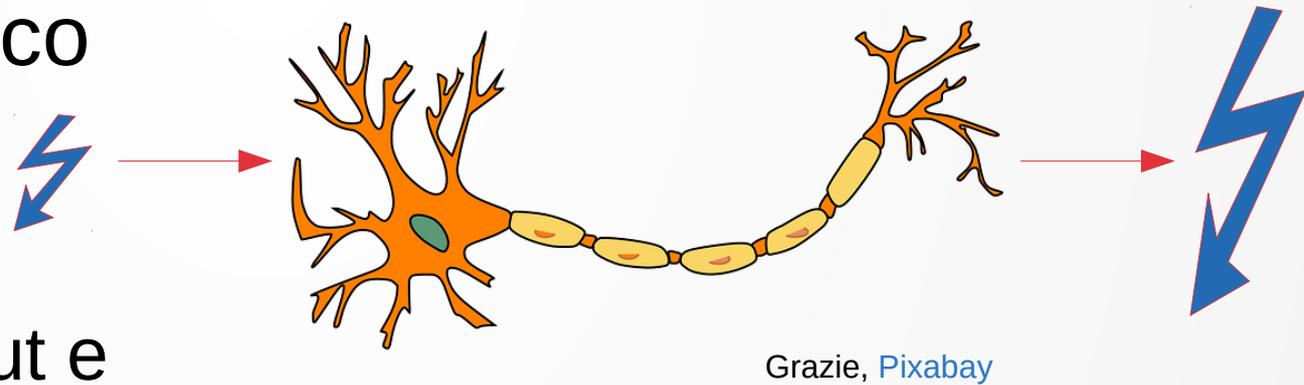
- Ancora un po' di relax
 - ↪ Una cosa utile questa volta, la colorazione automatica di immagini in bianco e nero
 - ↪ **Algoritmia** ([Richard Zhang](#))



Neuroni e Percettroni

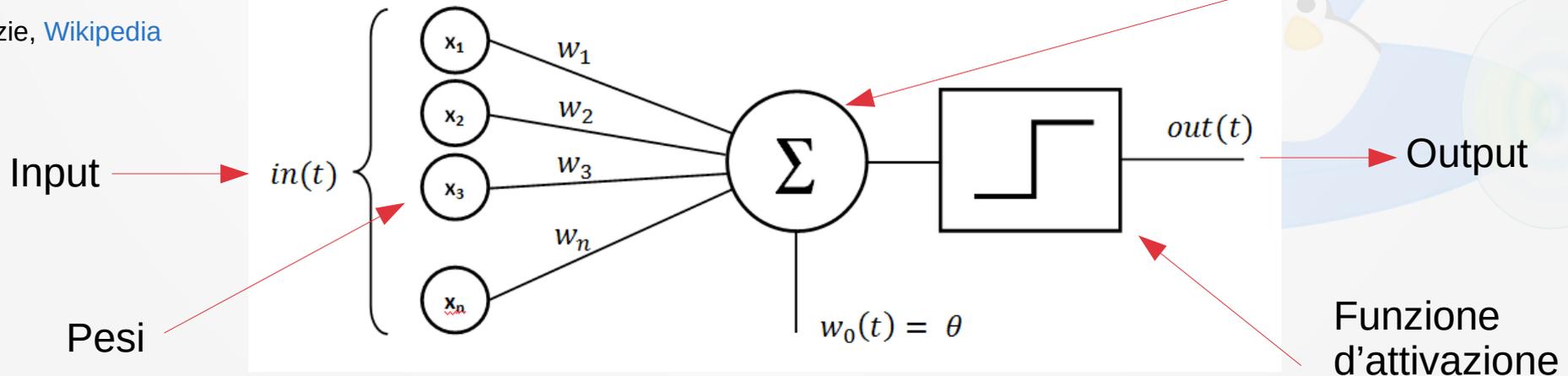
→ Neurone biologico

- ↳ Potenziale d'attivazione
- ↳ Corrente in input e output



→ Percettrone (neurone artificiale)

Grazie, Wikipedia



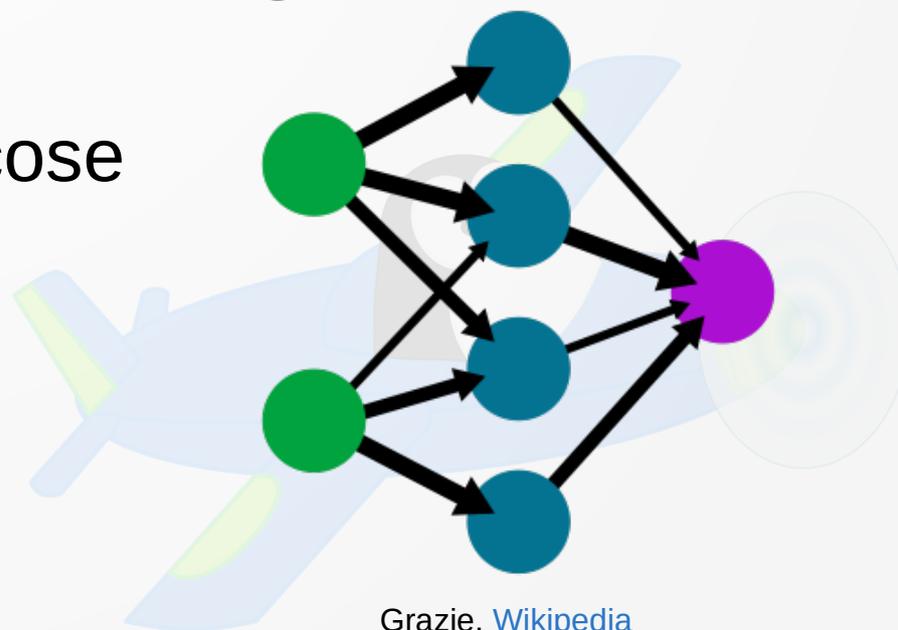
Reti neurali e *Deep learning*

→ Schemi non lineari di percettroni

- ↘ Aggregati con strategie anche complesse
- ↘ Funzioni di attivazione strane
- ↘ *Backpropagation*
- ↘ *Gradient descent*
- ↘ Si usano per tonnellate di cose
 - Visione artificiale
 - Riconoscimento vocale
 - Social networking
 - ...

Semplice rete neurale

strato di strato strato di
ingresso nascosto uscita

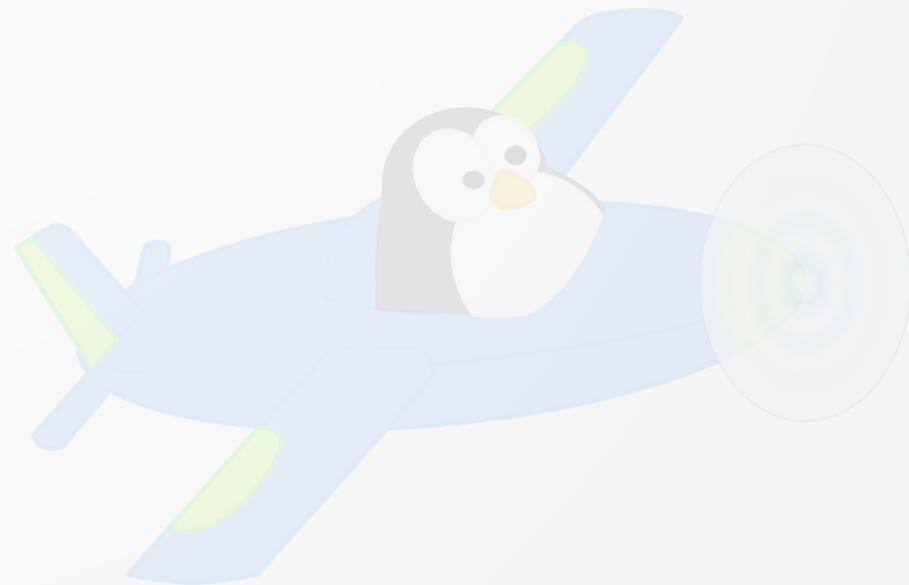


Altre domande?

→ Intanto

grazie!

→ Avete altre domande?



Riferimenti

- **Talk 17T50 all'End Summer Camp 2017**
<https://www.youtube.com/watch?v=IYOzg8BSy5c>
- **Pagina Wikipedia EN sul *machine learning***
https://en.wikipedia.org/wiki/Machine_learning
- **Slide sul *machine learning* e *big data***
<https://www.slideshare.net/awahid/big-data-and-machine-learning-for-businesses>
- **Appunti sulla regressione**
<http://malchiodi.di.unimi.it/teaching/big-scale-analytics/material/regression.pdf>
- **Capitolo 12 di “*Mining of Massive Datasets*”**
<http://www.mmds.org/>