

Machine Learning: il micro-corso Cattolica Assicurazioni

Introduzione al corso

Fabio Mardero

7 novembre 2019



Fabio Mardero



*Data scientist e addetto valutazioni riserve danni IAS presso
Cattolica Assicurazioni*

*Collaboratore e formatore all'interno del gruppo di studio
"Tarallucci, Vino e Machine Learning"*



*Formatore nell'ambito del Machine Learning, sia dal punto
di vista tecnico che divulgativo*

*Laureato in **fisica** e in **scienze statistiche e attuariali**,
certificazione GIMI sull'innovation management*



Io?



Non sono un programmatore!

*Vedo la programmazione come uno strumento
per mettere a terra idee e analisi sui dati*

Organizzazione del corso

5 incontri

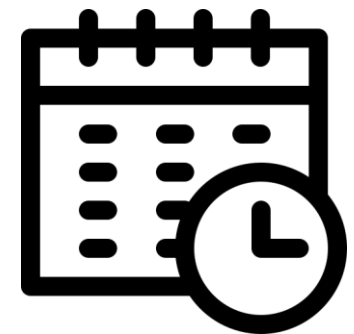
cadenza settimanale

ore 12:00-13:00

Giornate:

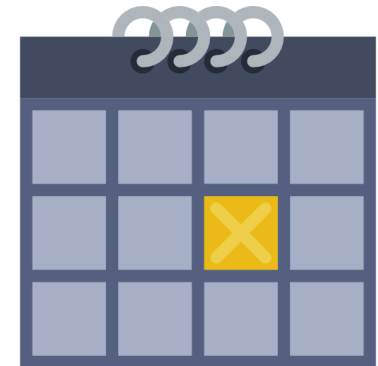
7-15-22-29 novembre

6 dicembre



Struttura del corso

1. *Introduzione all'Intelligenza Artificiale, Machine Learning e Deep Learning*
2. *Data Mining & Analytics*
3. *Un esempio di algoritmo di Machine Learning*
4. *Alcuni esempi di algoritmi di Deep Learning*
5. *Deployment e scalabilità nel Machine Learning*



Materiale necessario per seguire il corso



1. Un PC portatile
2. Un account google.com

Disclaimer

«Se la pizza è italiana allora il Machine Learning è anglosassone.»



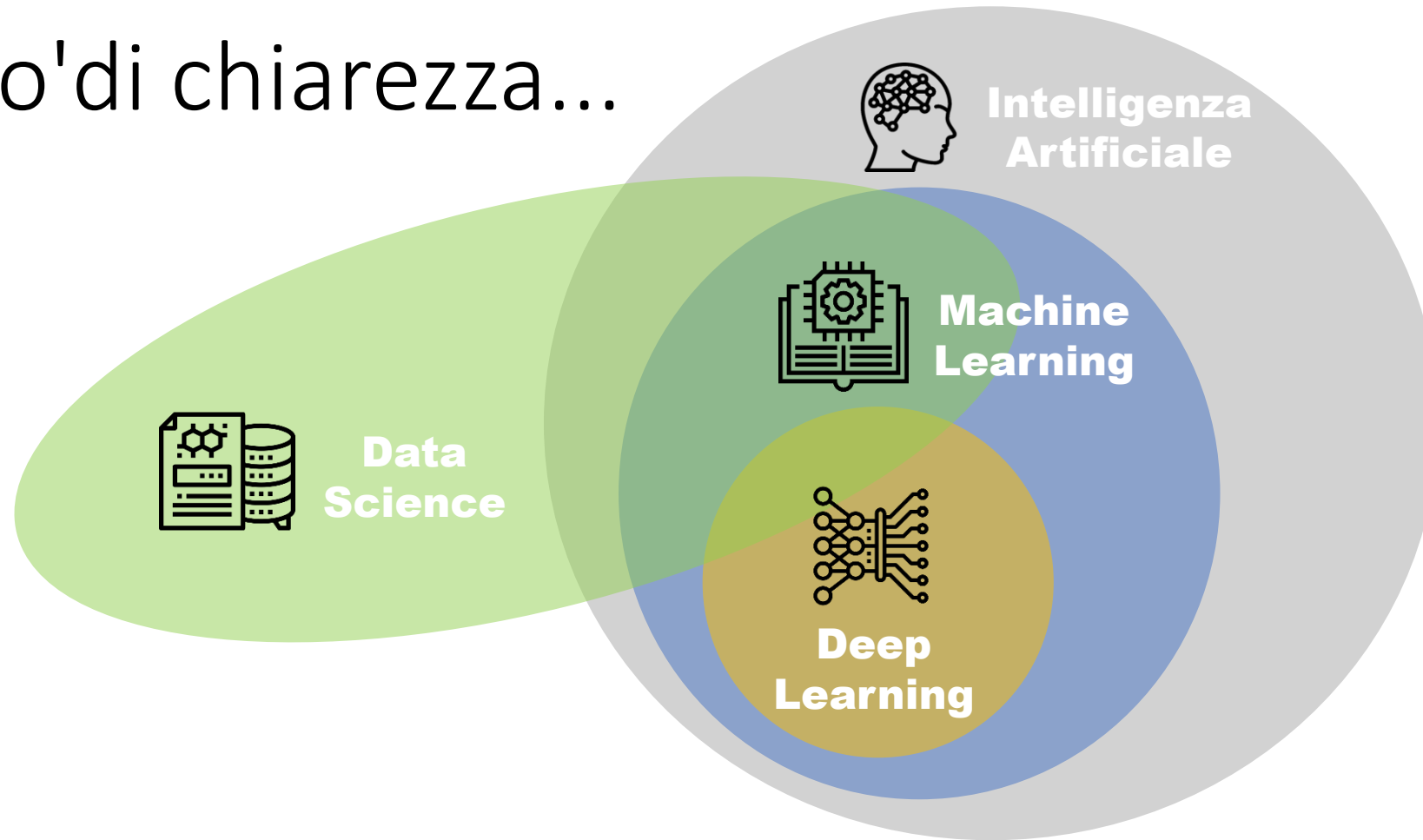
Le migliori fonti, la documentazione e le pubblicazioni
sull'argomento sono in inglese

Ciò non significa che nel panorama italiano non esistano documenti/libri validi

Machine Learning: un nuovo approccio al Data Mining

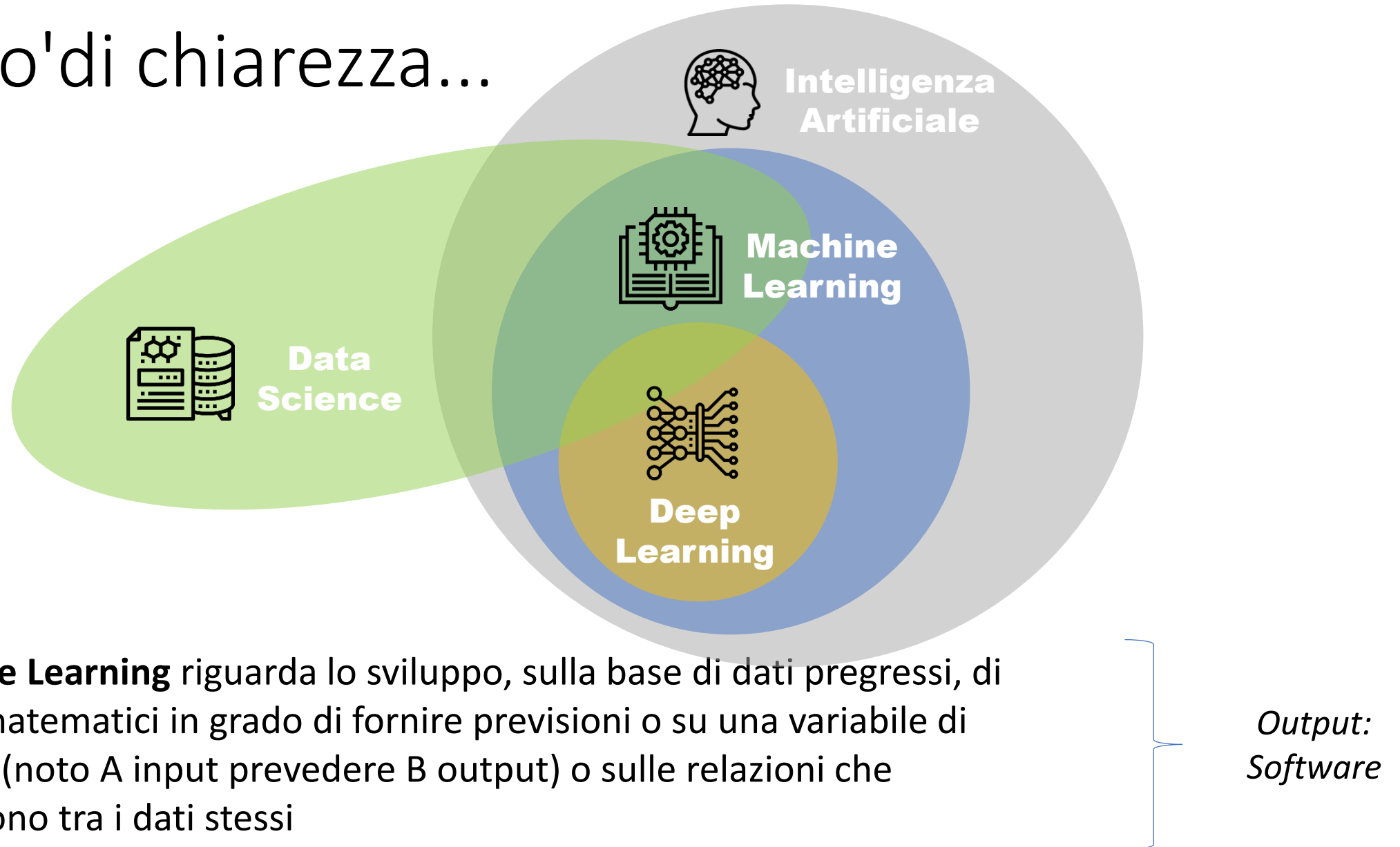
Introduzione e contestualizzazione

Un po' di chiarezza...



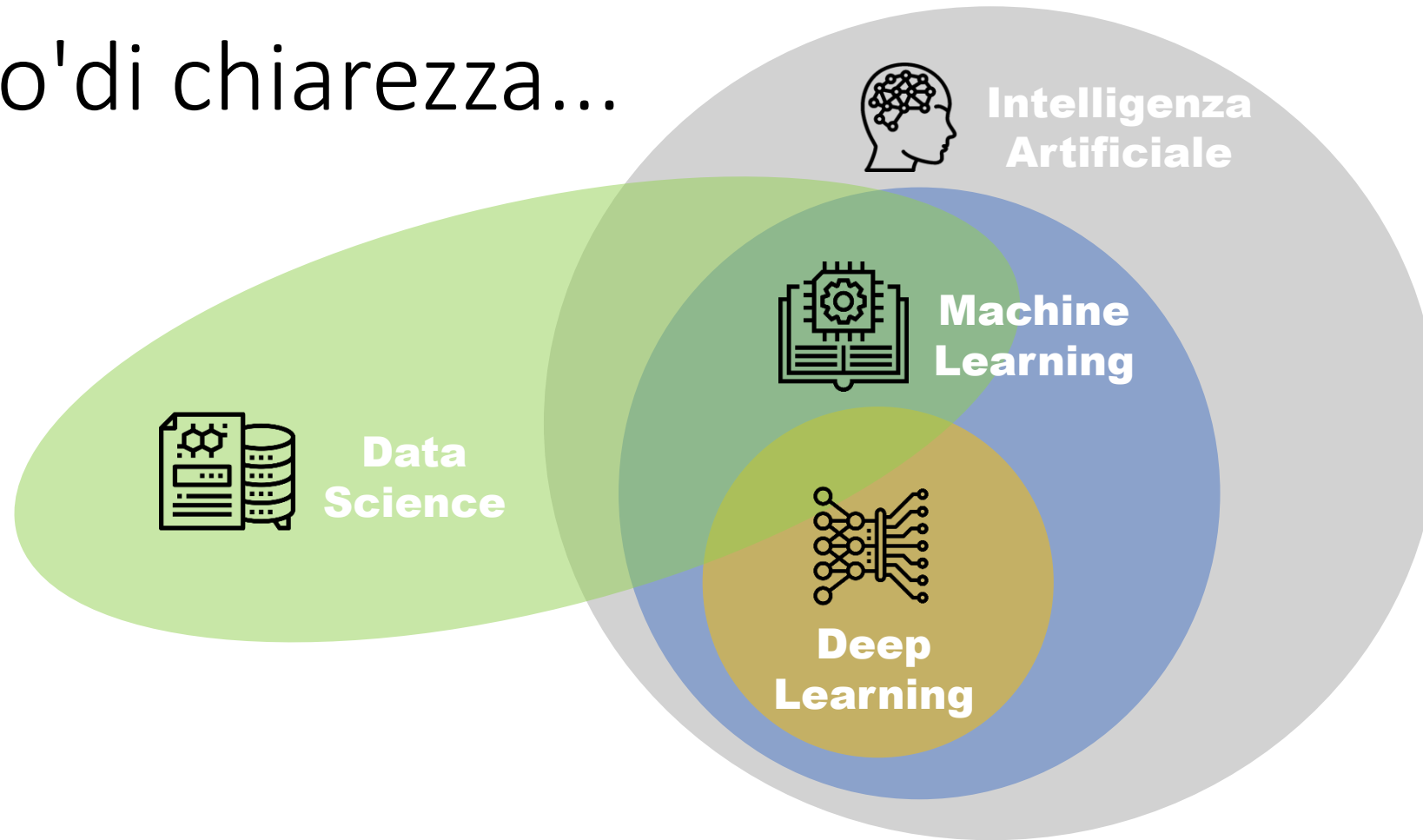
L'Intelligenza Artificiale consiste nella progettazione di sistemi hardware e software capaci di fornire all'elaboratore elettronico prestazioni che, a un osservatore comune, sembrerebbero essere di pertinenza esclusiva dell'intelligenza umana

Un po' di chiarezza...



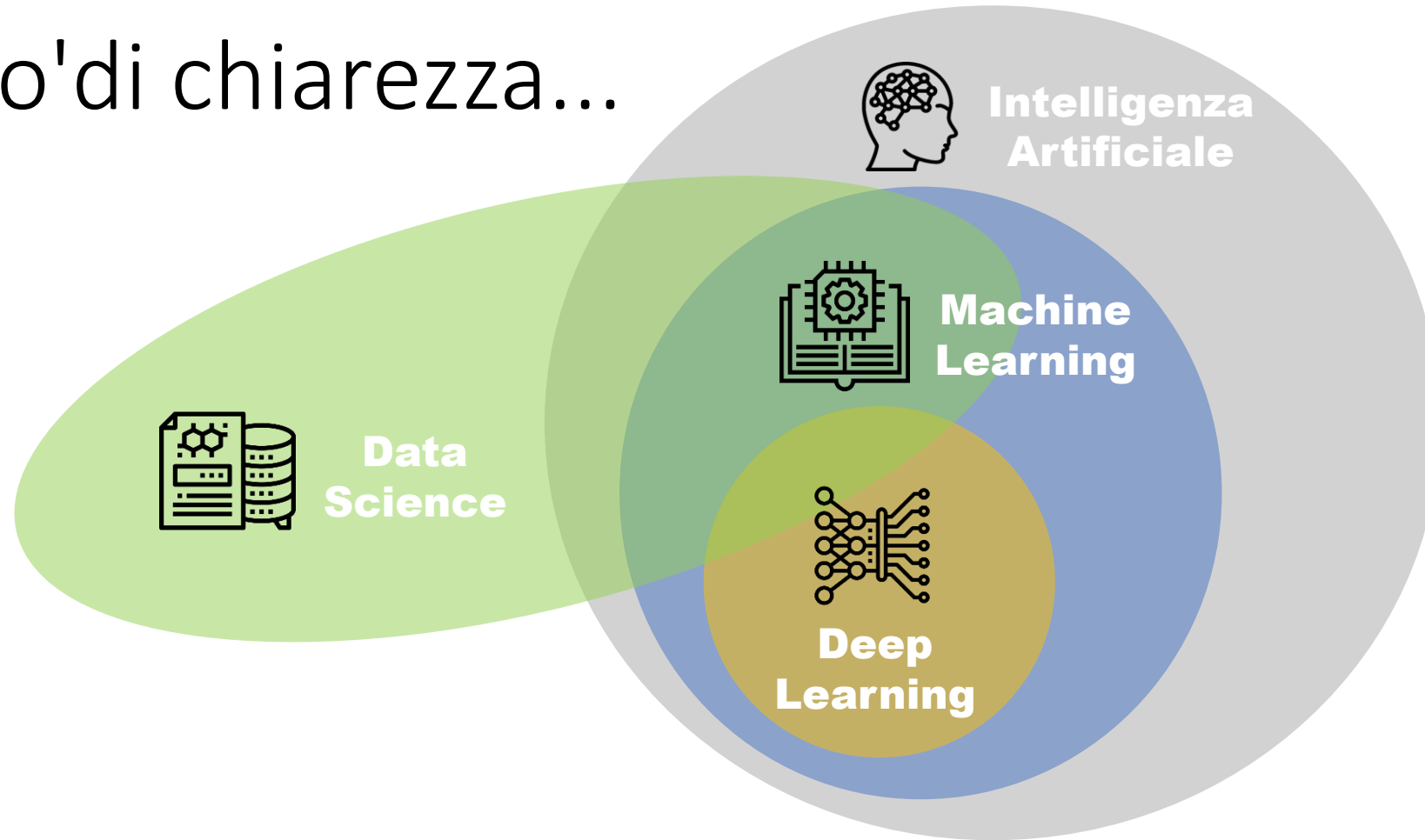
Il **Machine Learning** riguarda lo sviluppo, sulla base di dati pregressi, di modelli matematici in grado di fornire previsioni o su una variabile di interesse (noto A input prevedere B output) o sulle relazioni che intercorrono tra i dati stessi

Un po'di chiarezza...



Il **Deep Learning** include specifici modelli di machine learning, le reti neurali, che si sono guadagnati una sotto-classificazione dato il loro particolare funzionamento e i notevoli risultati che hanno raggiunto in tempi recenti

Un po' di chiarezza...



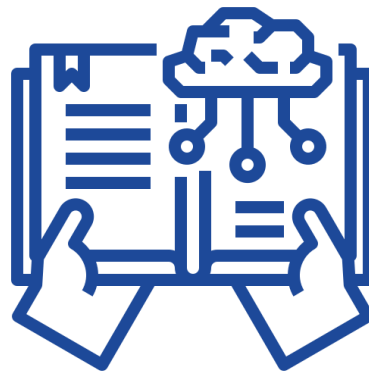
La **Data Science** consiste nell'estrazione dai dati di informazioni significative e/o utili al business

} *Output:
Decisioni
strategiche*

Machine learning

Il significato

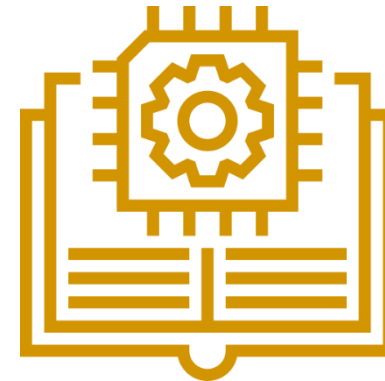
"Disciplina che utilizza metodi statistici per migliorare progressivamente, in seguito alla disponibilità di informazioni pregresse, le performance di un algoritmo in un dato compito."



Machine learning

Il significato

Un modello di machine learning è quindi in grado di “**apprendere**” dai **dati** allo scopo di eseguire, nel miglior modo possibile, un dato **compito**



Machine Learning

Elementi chiave

STEP 1

i dati

- 1. qualsiasi formato, anche immagini, audio o testo*
- 2. La quantità di dati necessaria per allenare un modello di Machine Learning dipende dall'algoritmo utilizzato e dalla difficoltà del compito*
- 3. Potrebbero essere necessarie alcune pre-elaborazioni*

Machine Learning

Elementi chiave

STEP 2

l'apprendimento

un modello apprende quando modifica la sua struttura, o i suoi parametri, per ridurre gli errori delle sue previsioni

- 1. Apprendimento per rinforzo*
- 2. Apprendimento supervisionato*
- 3. Apprendimento non supervisionato*

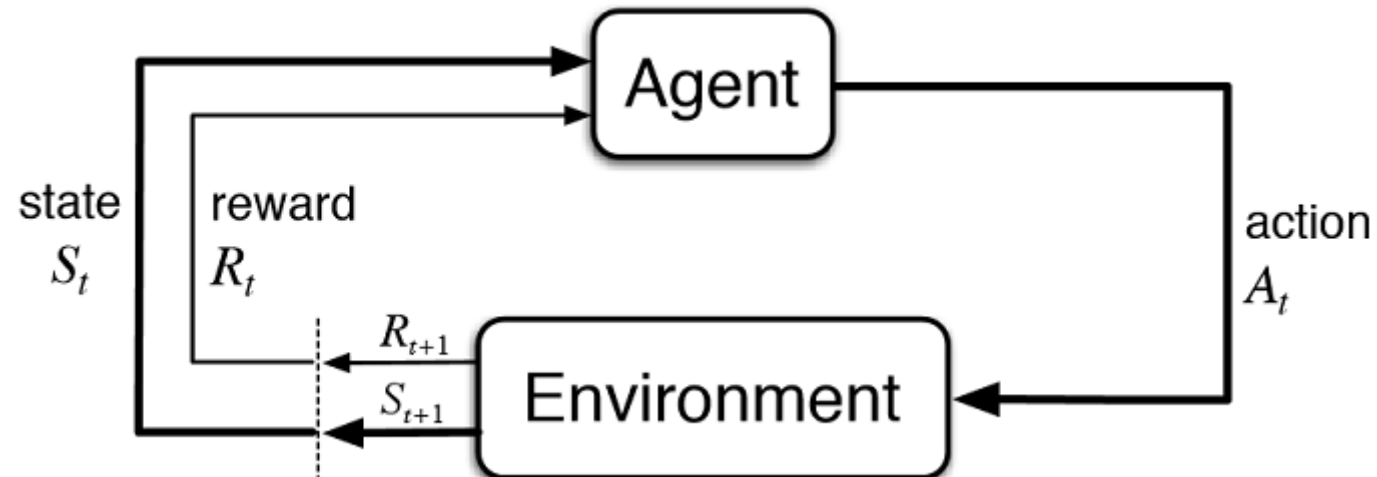
Machine Learning

Apprendimento per rinforzo

- *L'agente interagisce con l'environment e ogni sua azione modifica l'ambiente*
- *Il modello interagisce con il sistema e ogni sua previsione modifica il suo stato*

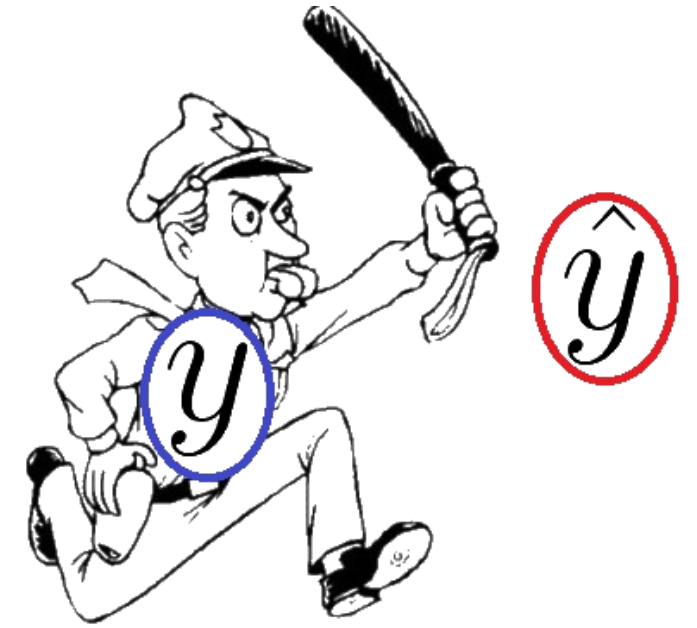
Nel tempo, non necessariamente ad ogni interazione con l'ambiente, l'agente riceve un feedback sul suo comportamento. Egli modifica quindi le sue future azioni, sulla base delle precedenti, tentando di massimizzare quelle che hanno portato a risultati positivi e minimizzando quelle risultate negative.

L'apprendimento dipende quindi da un sistema di rewards e punishments



Machine Learning

Apprendimento supervisionato



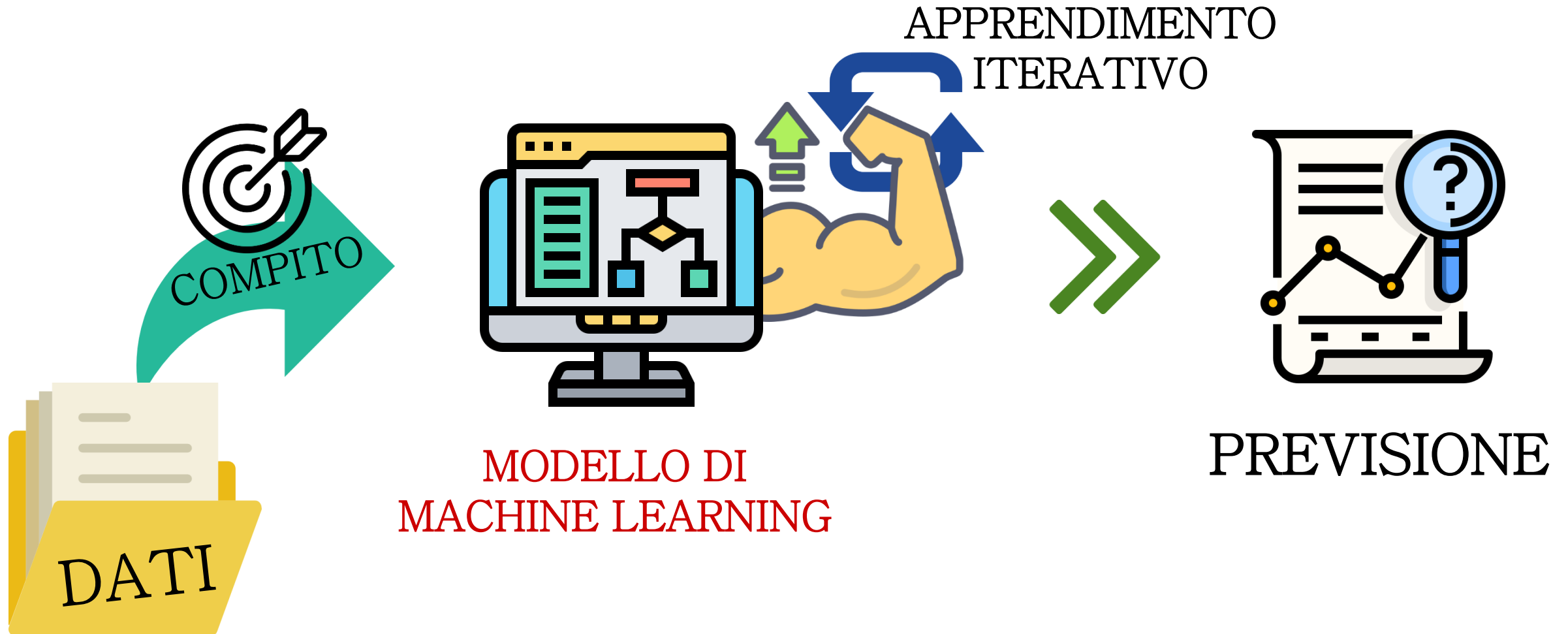
Il modello subisce l'ambiente

Nel caso dell'apprendimento supervisionato il modello mira a predire il comportamento di una o più variabili osservate rispetto alle altre

Indicata con \hat{y} la previsione e con y il valore osservato, il modello apprende a minimizzare l'errore tra \hat{y} e y . L'apprendimento è, informalmente, "supervisionato" dai valori di \hat{y}

Machine Learning

Apprendimento supervisionato in sintesi

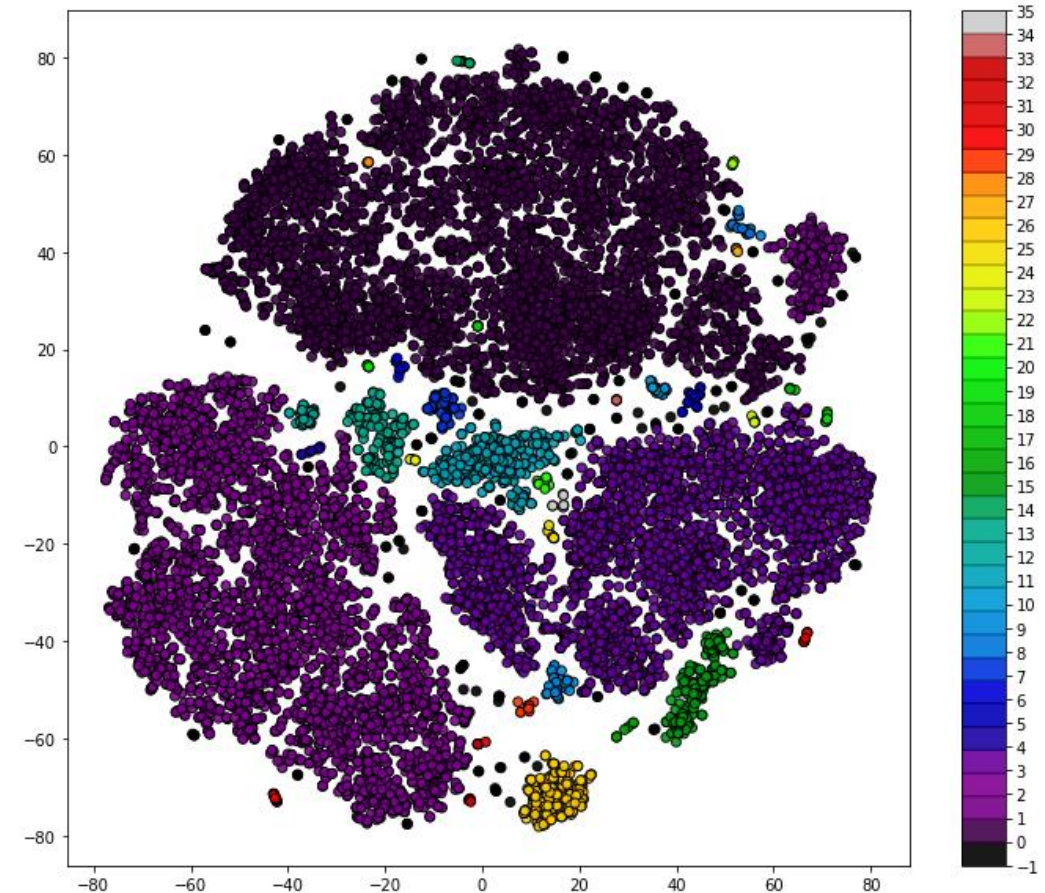


Machine Learning

Apprendimento non supervisionato

Il modello subisce l'ambiente ma non è allenato per fornire una previsione

L'apprendimento non supervisionato prevede che l'algoritmo ricerchi strutture informative (*pattern*) tra i dati



Machine Learning

Elementi chiave

STEP 3

il compito

definisce su cosa il modello è allenato e con quali intenzioni, ad esempio fornire previsioni o trovare pattern di aggregazione dei dati

1. *Regressione*
2. *Classificazione*
3. *Clustering*
4. ...

Machine Learning

Elementi chiave

STEP 3

il compito

Si riconoscono due casi:

- si individuano delle variabili più importanti, dette *variabili target/risposta*, rispetto alle altre, chiamate *variabili esplicative/covariate/features*
- tutte le variabili sono intese come significative (o potenzialmente tali)

Dato un insieme di dati, spetta all'osservatore decidere come intende interpretarli e se assegnare particolare importanza a qualcuna delle variabili disponibili

Machine Learning

Elementi chiave

STEP 3

il compito

Si riconoscono due casi:

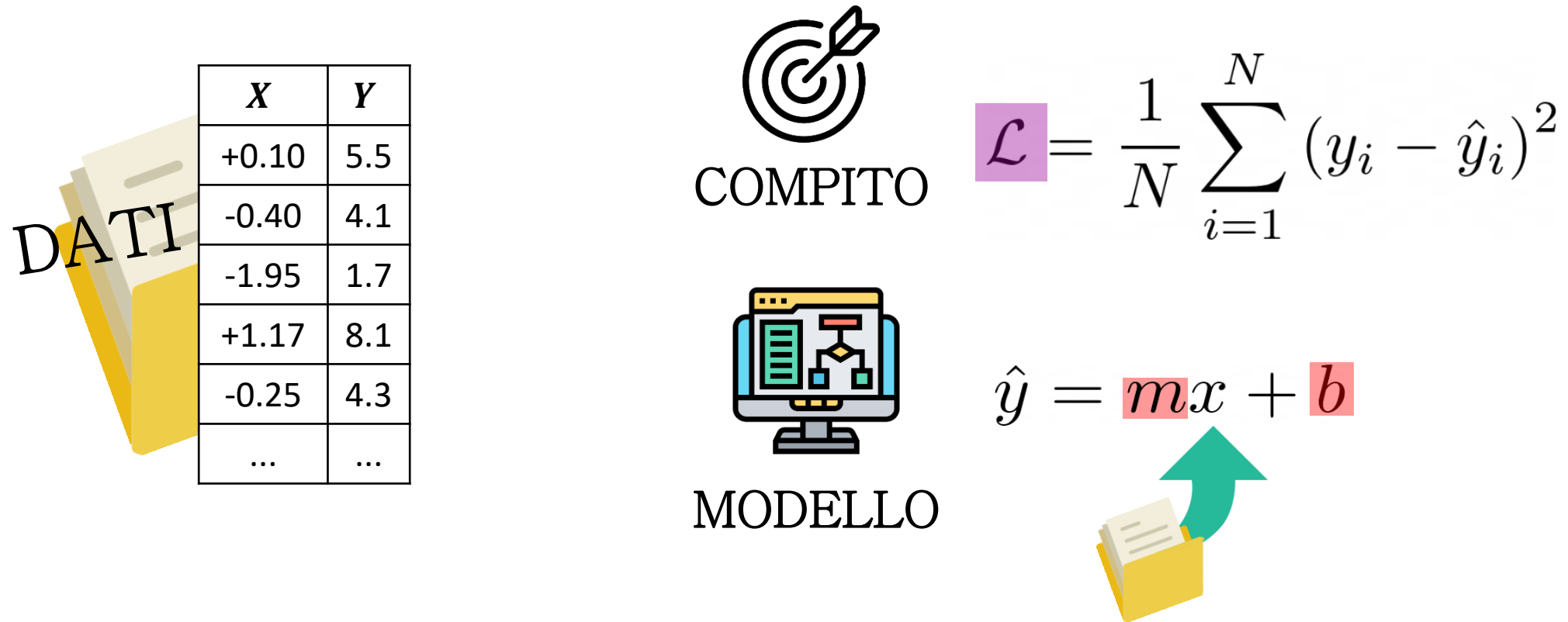
1. Compiti di regressione o classificazione

Mirando a fornire una previsione accurata delle variabili target, il modello spiega il fenomeno che genera \hat{y}

2. Compiti legati all'estrazione di informazione dai dati e ad una loro rappresentazione, ad esempio il clustering

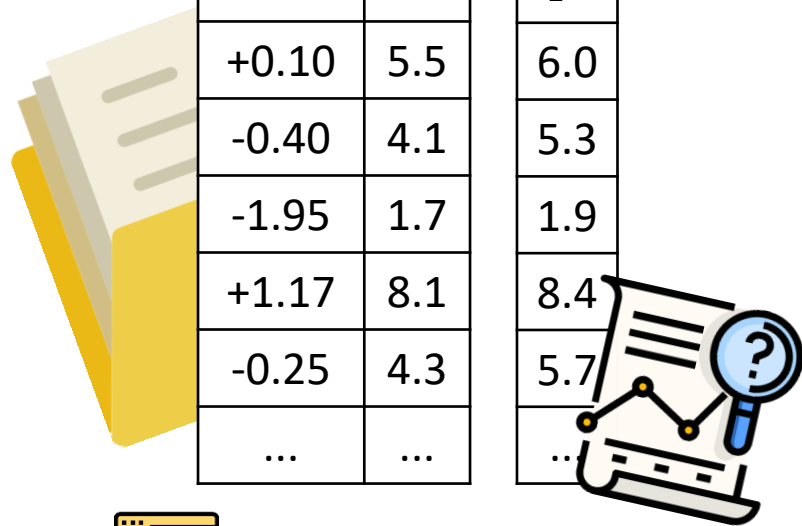
Ad esempio si individuano relazioni tra i dati contenuti in un dataset

Un esempio – Regressione Lineare



L'algoritmo ricerca i **parametri** che permettono di minimizzare la **funzione d'errore** calcolata sui dati noti

Un esempio – Regressione Lineare



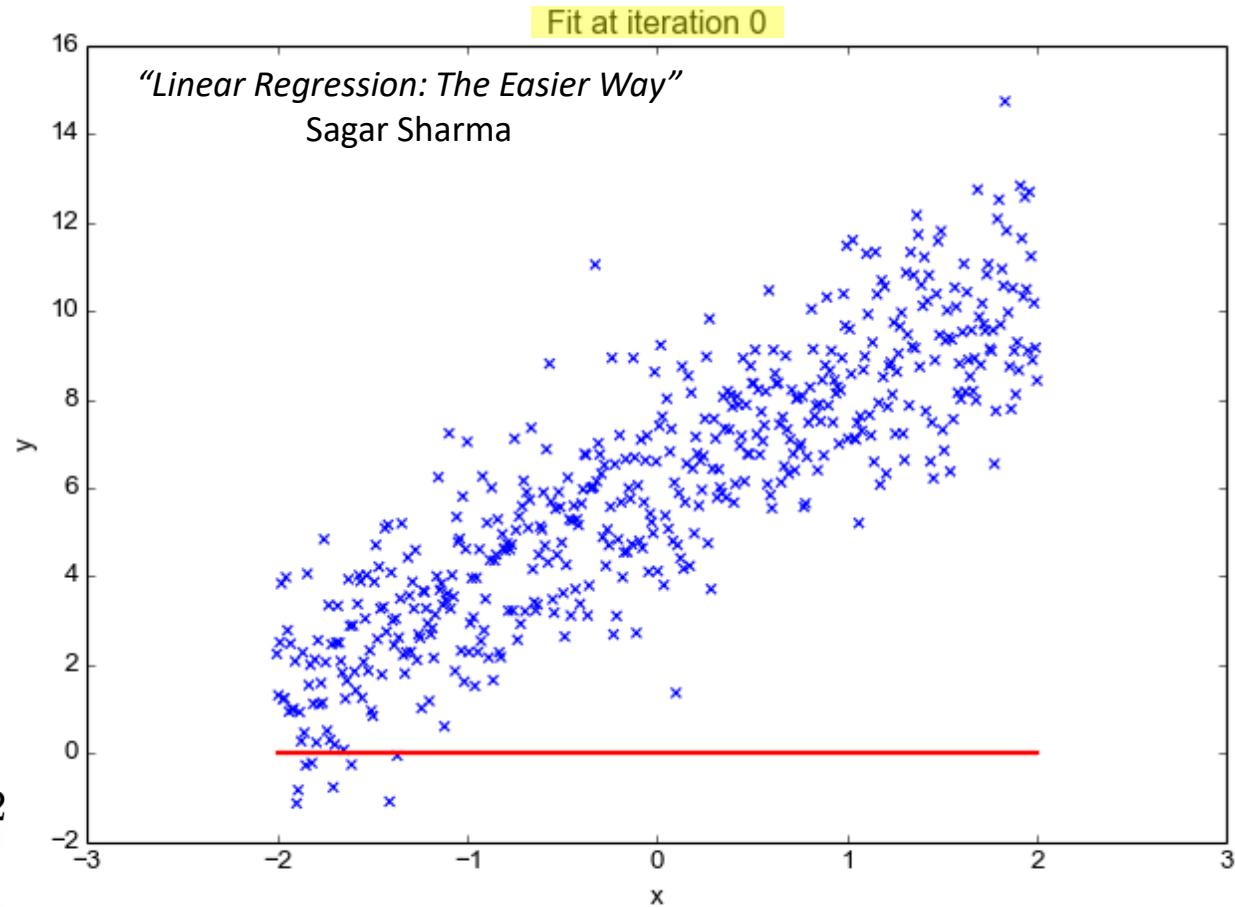
X	Y	\hat{Y}
+0.10	5.5	6.0
-0.40	4.1	5.3
-1.95	1.7	1.9
+1.17	8.1	8.4
-0.25	4.3	5.7
...



$$\hat{y} = mx + b$$



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



L'algoritmo ricerca i **parametri** che permettono di minimizzare la **funzione d'errore** calcolata sui dati noti

Capacità del Machine Learning

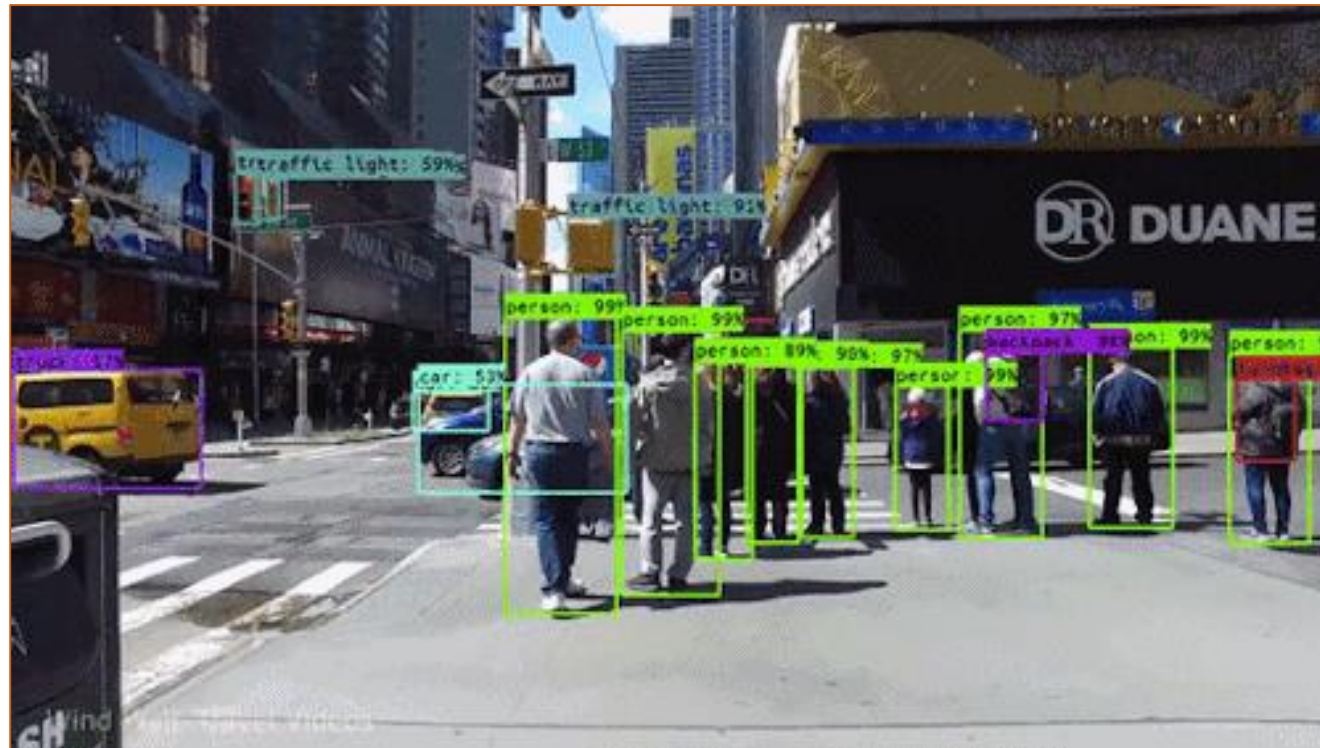


Un algoritmo di ML può:

- utilizzare sensori più precisi dei sensi umani
- comandare dispositivi in qualsiasi condizione con una precisione maggiore a quella umana
- agire più velocemente di un operatore
- lavorare in background anche 24h/24h (non si stanca mai)
- *raggiungere, per molti compiti, capacità predittive simili o superiori a quelle di un persona*

Capacità del Machine Learning

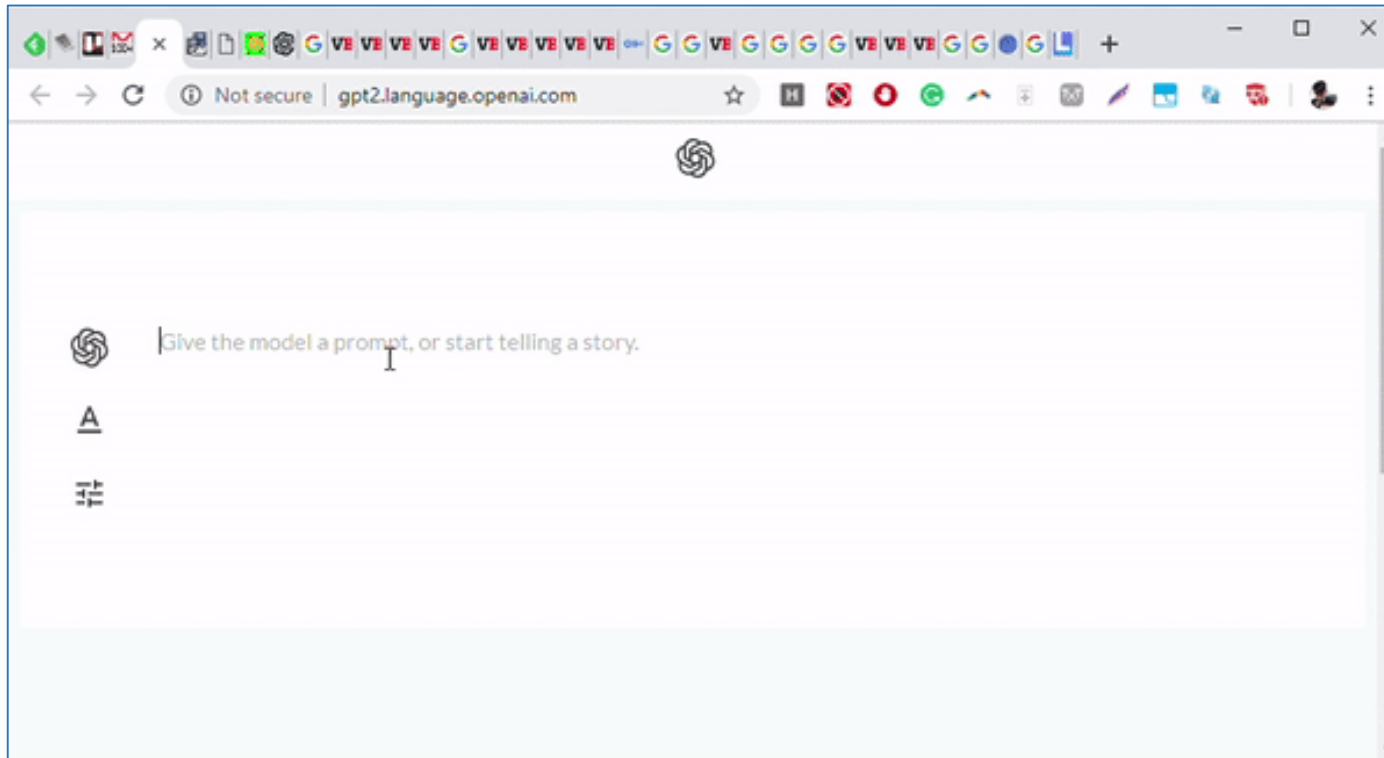
Rilevazione di oggetti



Capacità del Machine Learning



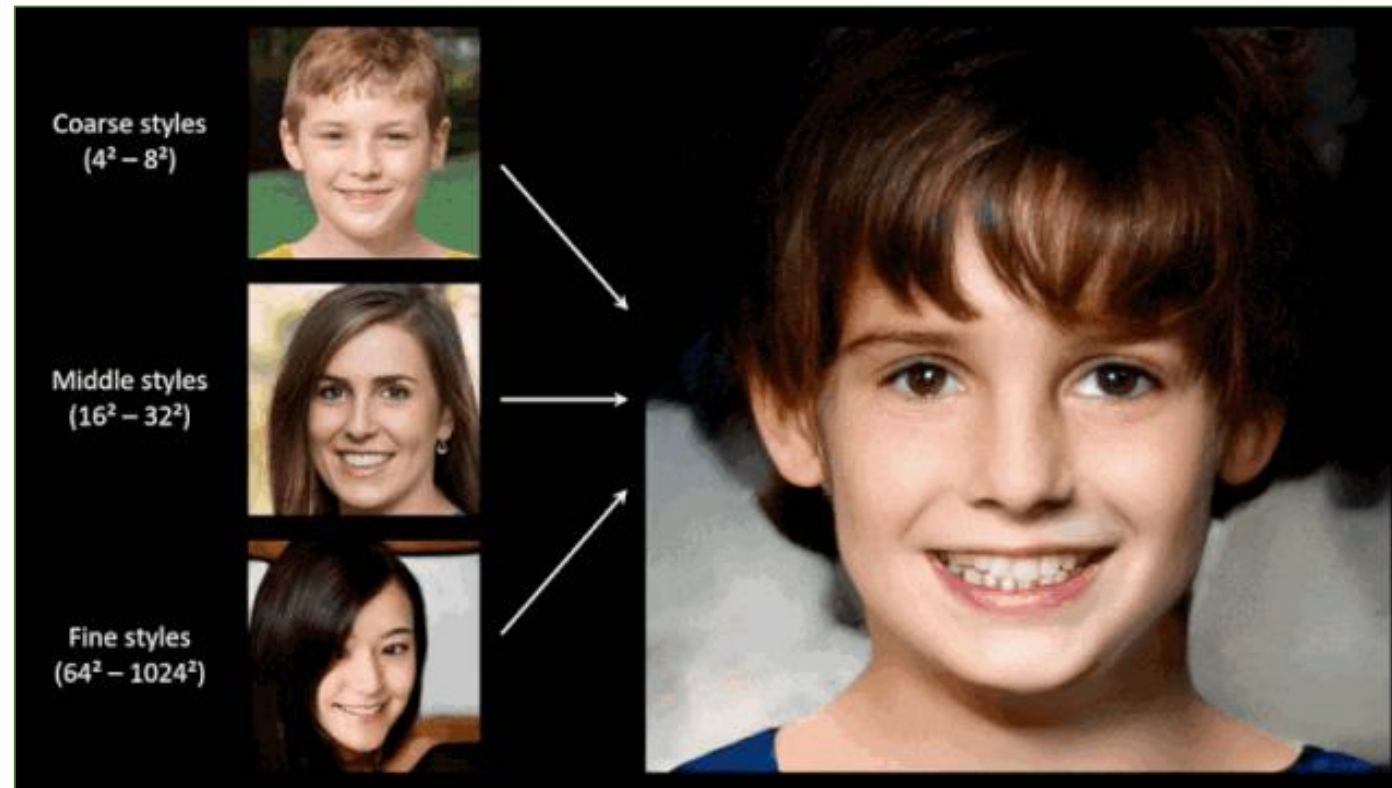
Generazione di testo



Capacità del Machine Learning



Generazione di volti



Capacità del Machine Learning

Alcune demo



Tramite webcam insegna all'algoritmo la
comprensione dei gesti

<https://teachablemachine.withgoogle.com/>

Capacità del Machine Learning

Alcune demo



Replica come l'algoritmo ha modellizzato la
similarità tra diversi vocaboli

<https://research.google.com/semantis/>

Capacità del Machine Learning

Alcune demo



<https://quickdraw.withgoogle.com/>

Capacità del Machine Learning

Alcune demo

E ancora...

- Pix2Pix

<https://affinelayer.com/pixsrv/>

- Replicatore dello stile di scrittura

<http://www.cs.toronto.edu/~graves/handwriting.html>

- Motore di ricerca


<https://books.google.com/talktobooks/>

Limiti del Machine Learning



Basta qualche secondo!

Limiti del Machine Learning

 *"Gli algoritmi di Machine Learning non sono in grado di svolgere compiti che richiederebbero per una comune persona più di qualche secondo per essere completati." Cit. Andrew Ag*

- Richiedono domain knowledge da parte degli sviluppatori
- Potenziale esposizione ad attacchi informatici
- Potenziali bias e inefficienze
- Difficoltà nell'interpretazione dei risultati prodotti dai modelli



Big Data e Machine Learning

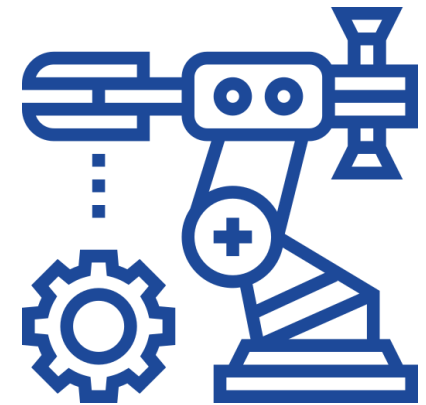
La massiccia raccolta di dati, anche tramite sistemi connessi e IoT, è estremamente legata all'implementazione di algoritmi di Machine Learning

È possibile sviluppare modelli di Machine Learning anche senza disporre di centinaia di GigaByte di dati (Big Data)

Si possono ottenere vantaggi economici per l'azienda già per la sola applicazione del Machine Learning, in particolare nelle aree dove prima non era previsto l'utilizzo di alcun algoritmo

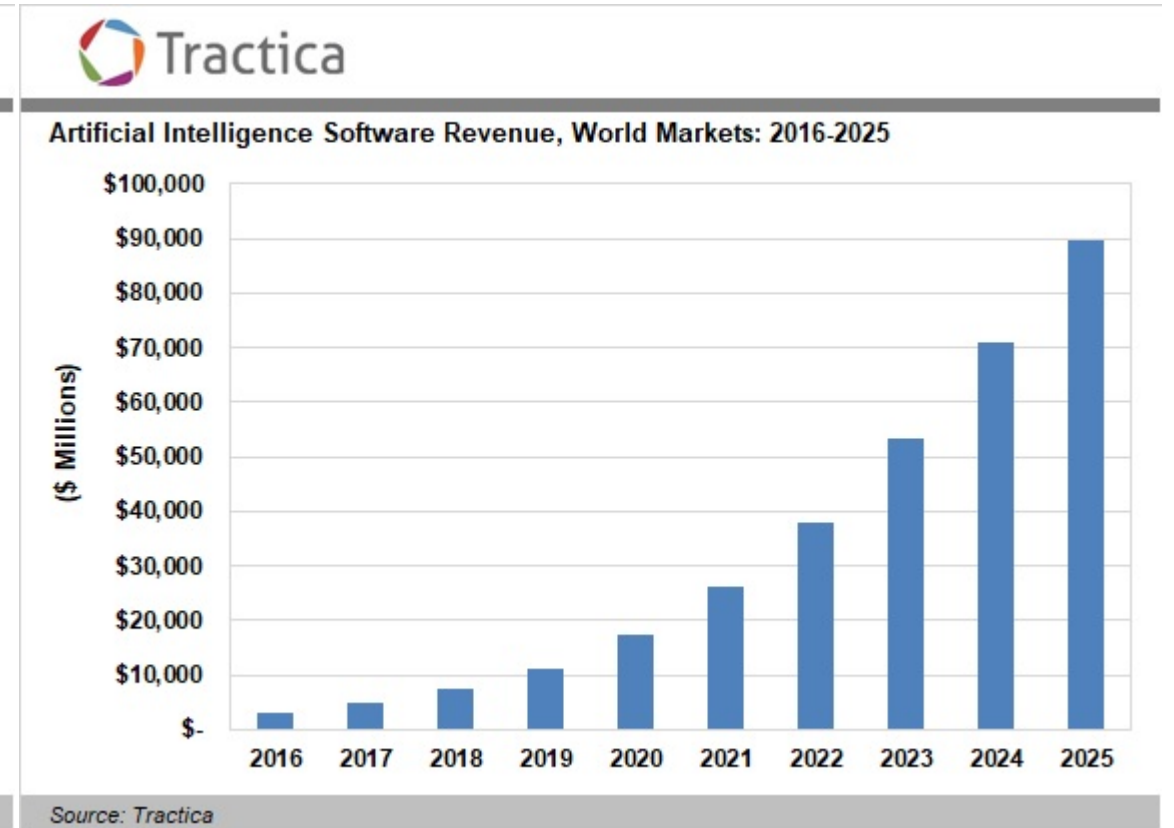
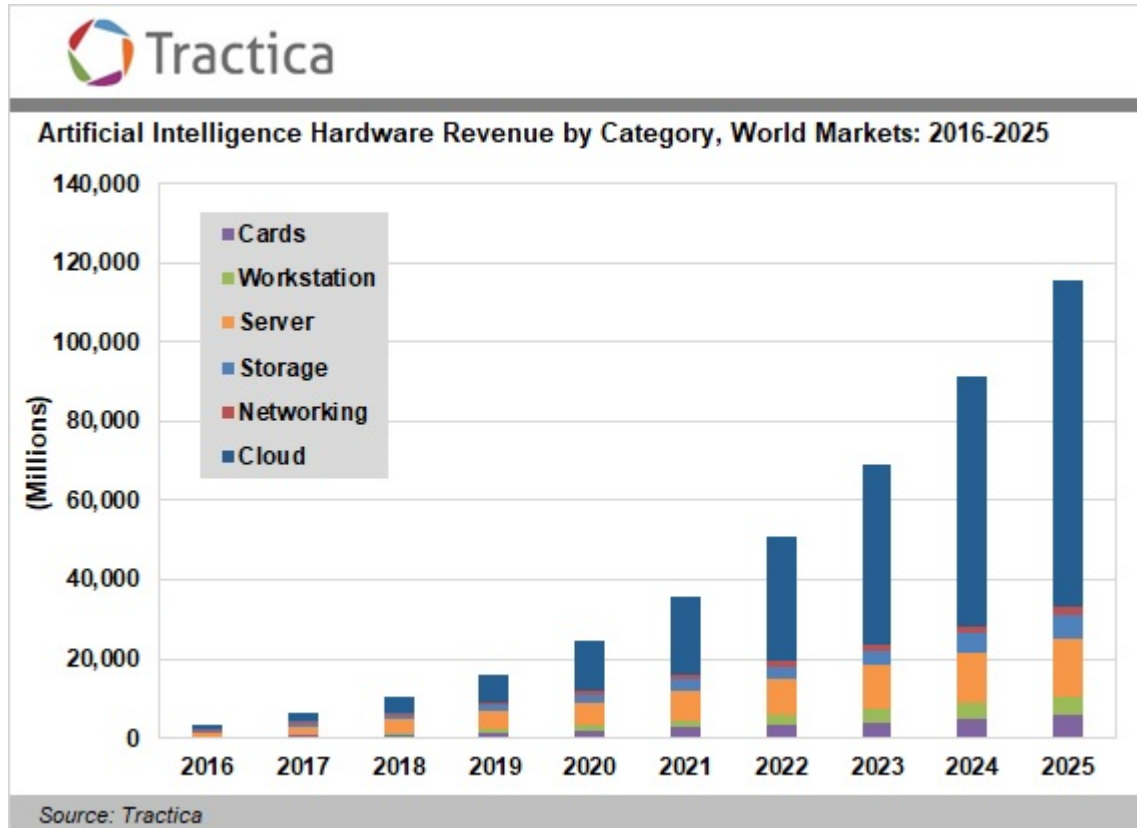
Industria 4.0 e Machine Learning

Il Machine Learning è strettamente legato all'Industria 4.0 in quanto consente all'impresa di implementare o evolvere i propri processi rendendoli più veloci ed efficienti, anche grazie all'automazione



Il mercato del Machine Learning

Nel presente e in futuro

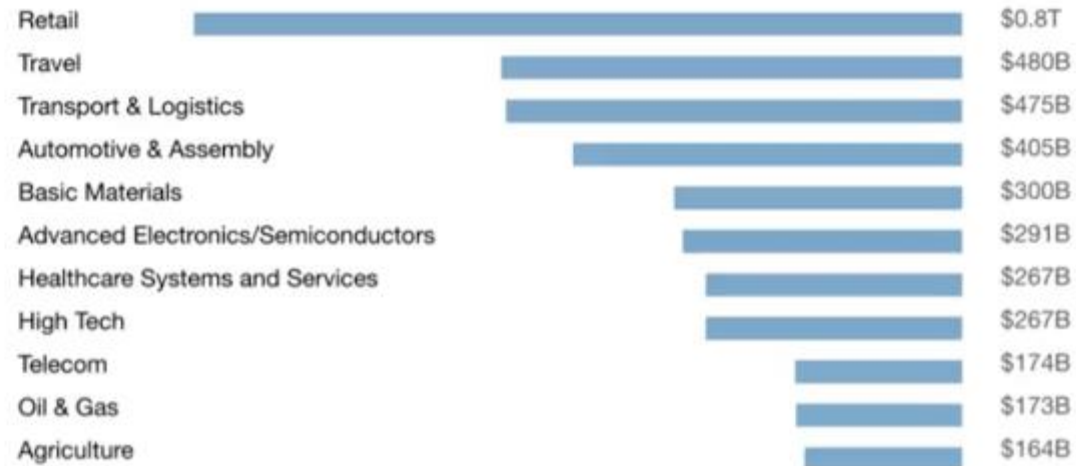


Il mercato del Machine Learning

In futuro

AI value creation
by 2030

**\$13
trillion**



“Notes from the AI frontier: Modeling the impact of AI on the world economy”

September 2018, mckinsey.com

Il mercato del Machine Learning

Nel presente e in futuro

*Il mercato mondiale della **data annotation** è stato valutato nel 2018 pari a più di 316 milioni di USD e si stima che nel 2025 possa arrivare a valere oltre 1.6 miliardi di USD (*)*

Secondo le previsioni, la crescita sarà trainata soprattutto dal mercato dell'automotive, quello retail, e i settori healthcare

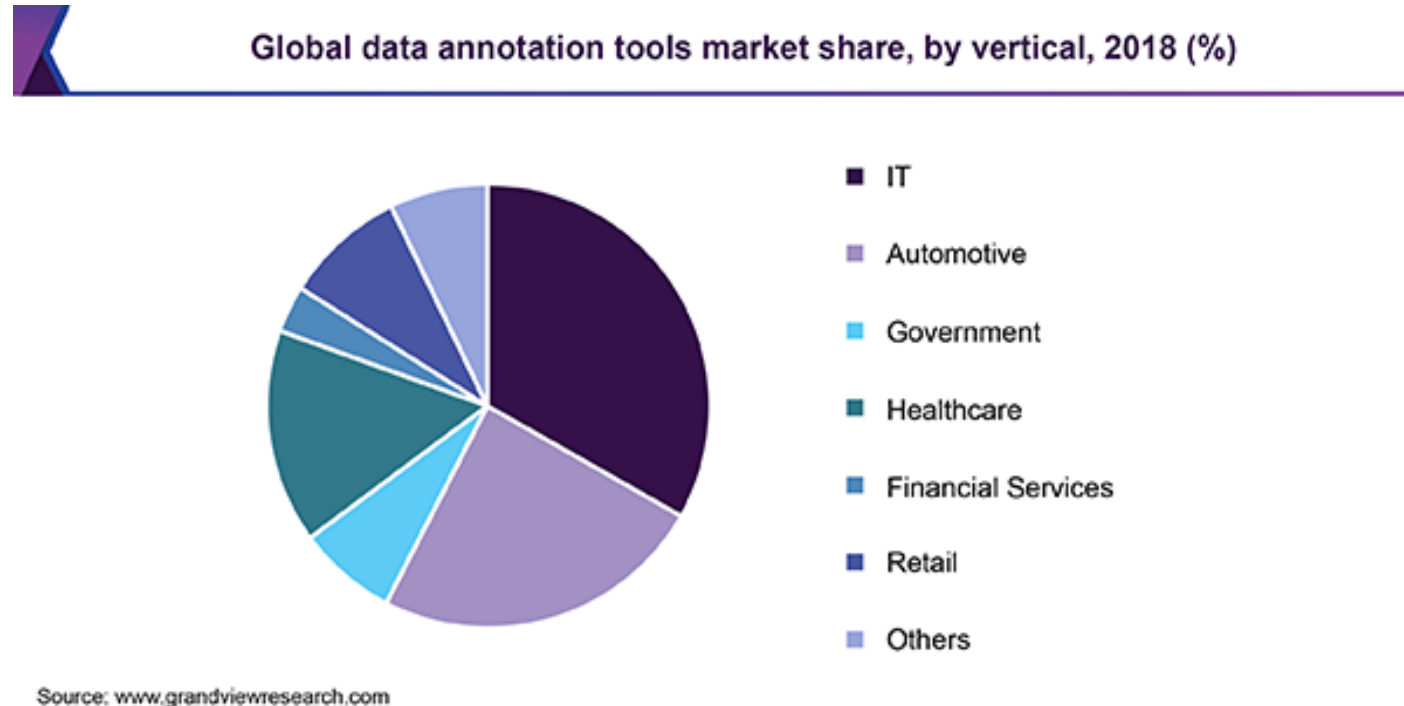
Esempi di aziende interamente dedicate alla data annotation:

- **Testin**
- **Scale AI** (valutata circa 100 milioni di dollari US)
- **Mighty AI** (acquisita nel 06/2019 da Uber)

(*) "Data Annotation Tools Market Size, Share & Trends Analysis Report, 2019 – 2025", [grandviewresearch.com](https://www.grandviewresearch.com)

Il mercato del Machine Learning

Nel presente



“Desperate Venezuelans are making money by training AI for self-driving cars”

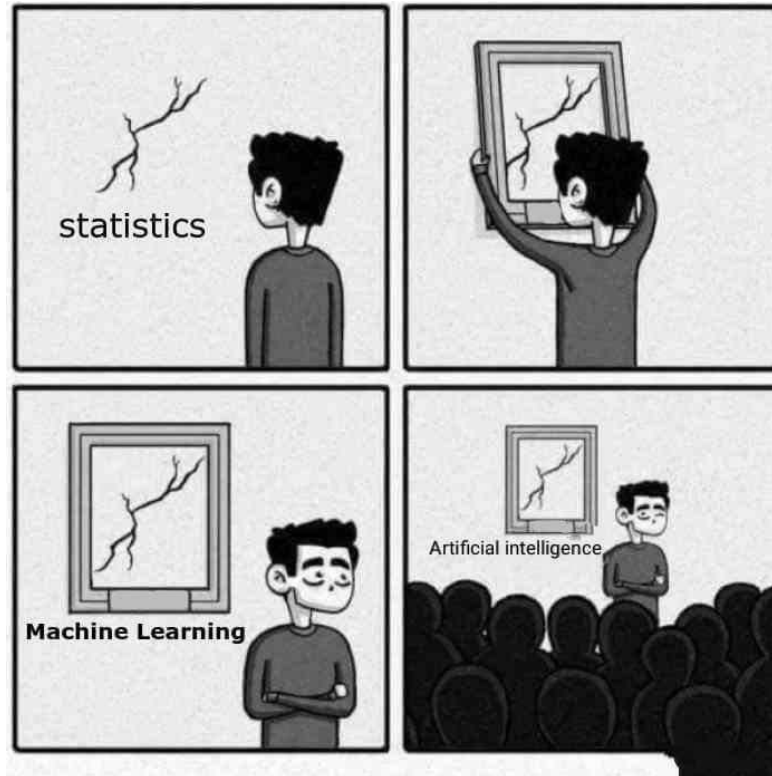
Agosto 2019, technologyreview.com

Machine Learning: un nuovo approccio al Data Mining

Introduzione teorica

Introduzione teorica al Machine Learning

La statistica

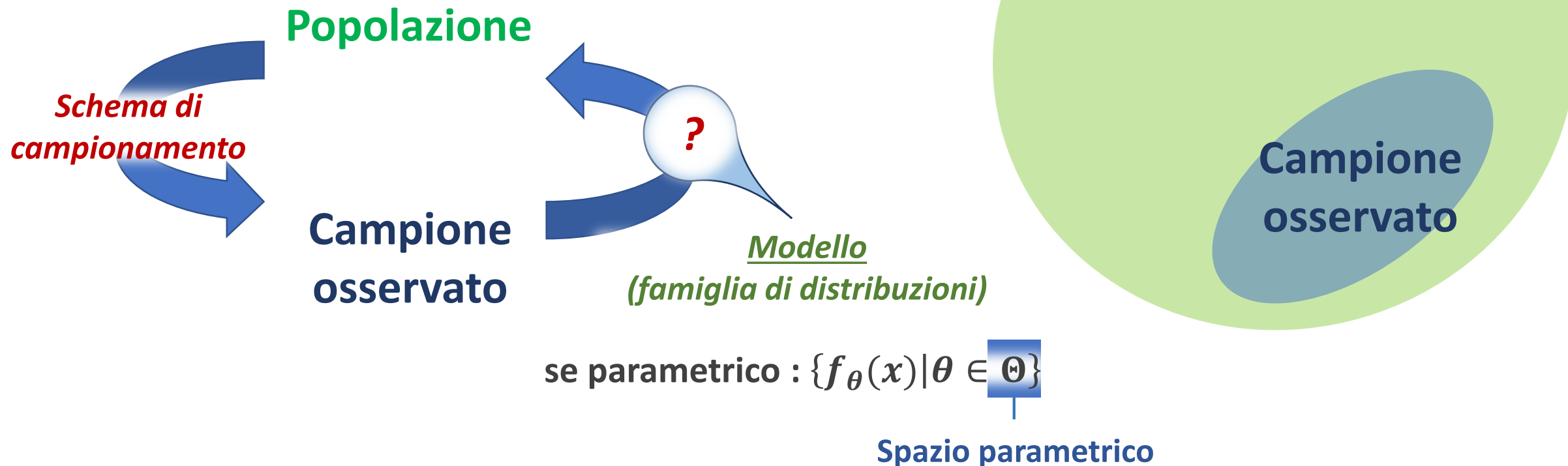


Definiamo incerto/aleatorio tutto ciò che non possiamo verificare in maniera deterministica, per mancanza di informazioni o per proprietà intrinseca del fenomeno

Introduzione teorica al Machine Learning

Inferenza statistica

Molto spesso...



Introduzione teorica al Machine Learning

Apprendimento supervisionato

Dataset: $\mathcal{D} = \{ (x_i, y_i) \}_{i=1}^N$ con $x_i \in \mathcal{X}, y_i \in \mathcal{Y}$.

Variabile target descritta tramite un modello predittivo:

Componente deterministica

$$Y = \boxed{f(x)} + \boxed{\varepsilon}, \quad x \in \mathcal{X}$$

Componente stocastica

Ipotesi induttiva

$$f: \mathcal{X} \longrightarrow \mathcal{Y}$$

$$x \longmapsto f(x) = \hat{y}$$

Modello parametrico

$$\mathcal{F} = \{ f(x; \theta); \theta \in \Theta \}, \quad \text{per ogni } x \in \mathcal{X}$$

Introduzione teorica al Machine Learning

Apprendimento supervisionato

Loss function:

$$L: \mathcal{Y} \times \mathcal{Y} \longrightarrow \mathbb{R}$$
$$(y_1, y_2) \longmapsto L(y_1, y_2)$$

$$f: \mathcal{X} \longrightarrow \mathcal{Y}$$
$$\mathbf{x} \longmapsto f(\mathbf{x}) = \hat{y}$$

$$L(\hat{y}, y) = L(f(\mathbf{x}; \boldsymbol{\theta}), y), \quad \text{per ogni } (\mathbf{x}, y) \in \mathcal{D}$$

Introduzione teorica al Machine Learning

Apprendimento supervisionato

Risk function:

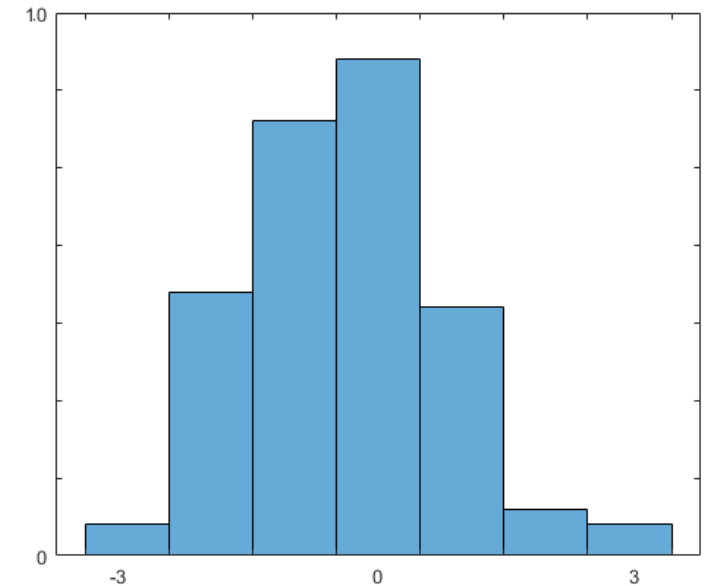
$$R(\boldsymbol{\theta}) = \mathbb{E} [L (f(\mathbf{X}; \boldsymbol{\theta}), Y)]$$
$$= \int L (f(\mathbf{x}; \boldsymbol{\theta}), y) dF(\mathbf{x}, y)$$

approssimando

$$F_{emp}(\mathbf{x}, y) = \frac{1}{N} \sum_{i=1}^N I(\mathbf{x} \leq \mathbf{x}_i, y \leq y_i)$$

➡ $R(\boldsymbol{\theta}) \approx R_{emp}(\mathcal{D}, \boldsymbol{\theta}) = \frac{1}{N} \sum_{i=1}^N L (f(\mathbf{x}_i; \boldsymbol{\theta}), y_i)$

Distribuzione di probabilità empirica



Introduzione teorica al Machine Learning

Apprendimento supervisionato

$$R(\boldsymbol{\theta}) \approx R_{emp}(\mathcal{D}, \boldsymbol{\theta}) = \frac{1}{N} \sum_{i=1}^N L(f(\mathbf{x}_i; \boldsymbol{\theta}), y_i)$$

Alcuni esempi:

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

$$MAE = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i| \quad \text{Mean Absolute Error}$$

Introduzione teorica al Machine Learning

Apprendimento supervisionato

Empirical Risk Minimization *(ERM)*

$$\hat{\theta} = \arg \min_{\theta \in \Theta} R_{emp}(\mathcal{D}, \theta)$$

Modello parametrico

$$\mathcal{F} = \{f(x; \theta); \theta \in \Theta\}, \quad \text{per ogni } x \in \mathcal{X}$$

$$R(\theta) \approx R_{emp}(\mathcal{D}, \theta) = \frac{1}{N} \sum_{i=1}^N L(f(x_i; \theta), y_i)$$

Introduzione teorica al Machine Learning

Apprendimento supervisionato

Empirical Risk Minimization (ERM)

$$\hat{\theta} = \arg \min_{\theta \in \Theta} R_{emp}(\mathcal{D}, \theta)$$



**Risoluzione del
problema di
ottimizzazione**

$$\nabla_{\theta} = \left(\frac{\partial}{\partial \theta_1}, \dots, \frac{\partial}{\partial \theta_p} \right)^T$$

gradiente

$$0 = \nabla_{\theta} R_{emp}(\mathcal{D}, \theta)$$

$$0 = \frac{1}{N} \sum_{i=1}^N \nabla_{\theta} L(f(x_i; \theta), y_i)$$

$$0 = \frac{1}{N} \sum_{i=1}^N \frac{\partial L(t, y_i)}{\partial t} \nabla_{\theta} f(x_i; \theta)$$

Introduzione teorica al Machine Learning

Regole di derivazione

$$(k \cdot f)'(x) = k \cdot f'(x) \quad k \in \mathbb{R}$$

$$(f \pm g)'(x) = f'(x) \pm g'(x)$$

$$(f \cdot g)'(x) = f'(x) \cdot g(x) + f(x) \cdot g'(x)$$

$$\left(\frac{f}{g}\right)'(x) = \frac{f'(x)g(x) - f(x)g'(x)}{(g(x))^2}$$

$$\text{Chain rule: } (g \circ f)'(x) = g'(f(x)) \cdot f'(x)$$

$$\text{Constant Rule: } \frac{d}{dx}[c] = 0$$

$$\text{Power Rule: } \frac{d}{dx}[x^n] = n \cdot x^{n-1}$$

Introduzione teorica al Machine Learning

Teorema di Fermat (1 / 3)

Enunciato e dimostrazione del teorema di Fermat

youmath.it

Sia $y = f(x)$ una funzione con dominio $Dom(f) \subseteq \mathbb{R}$. Se $x_0 \in Dom(f)$ è un punto estremante per f , e la funzione è derivabile in quel punto, allora si ha che

$$f'(x_0) = 0$$

Dimostrazione

Prima di tutto osserviamo che per ipotesi $f(x)$ è derivabile nel punto x_0 , dunque vale la condizione

$$\lim_{x \rightarrow x_0^-} f'(x) = \lim_{x \rightarrow x_0^+} f'(x)$$

Dimostriamo il teorema nel caso in cui x_0 sia un punto di massimo relativo; il caso in cui è un punto di minimo si dimostra in maniera del tutto analoga.

Poiché x_0 è un punto di massimo relativo, dato un incremento h vale

$$f(x_0 + h) - f(x_0) \leq 0$$

Introduzione teorica al Machine Learning

Teorema di Fermat (2 / 3)

Infatti se x_0 è un punto di massimo spostandoci sull'asse delle ascisse troveremo, localmente, valori della funzione più piccoli di $f(x_0)$.

Dividiamo la disuguaglianza per h . Otteniamo:

youmath.it

- se h è positivo

$$\frac{f(x_0 + h) - f(x_0)}{h} \leq 0$$

- se h è negativo

$$\frac{f(x_0 + h) - f(x_0)}{h} \geq 0$$

Ora: se passiamo al limite per $h \rightarrow 0$ in entrambe le disuguaglianze, otteniamo

$$\lim_{h \rightarrow 0} \frac{f(x_0 + h) - f(x_0)}{h} \leq 0 \quad (h > 0)$$

$$\lim_{h \rightarrow 0} \frac{f(x_0 + h) - f(x_0)}{h} \geq 0 \quad (h < 0)$$

Introduzione teorica al Machine Learning

Teorema di Fermat (3 / 3)

I due limiti sono rispettivamente limite destro e limite sinistro della derivata prima,

$$f'_+(x_0) = f'_-(x_0)$$

Per l'ipotesi di derivabilità di f in x_0 in due limiti devono coincidere, quindi essendo

$$f'_+(x_0) \leq 0 \quad \text{e} \quad f'_-(x_0) \geq 0$$

l'unico caso possibile è

$$f'_+(x_0) = 0 = f'_-(x_0)$$

Ossia

youmath.it

$$f'(x_0) = 0$$

Introduzione teorica al Machine Learning

Problema di ottimizzazione

Un esempio...

$$\mathcal{R}(\theta_1, \theta_2) = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

$$(\hat{\theta}_1, \hat{\theta}_2) = \arg \min_{(\theta_1, \theta_2)} \mathcal{R}(\theta_1, \theta_2)$$

**problema di
ottimizzazione**

$$\left\{ \begin{array}{l} \frac{\partial \mathcal{R}}{\partial \theta_1}(\hat{\theta}_1, \hat{\theta}_2) = 0 \\ \frac{\partial \mathcal{R}}{\partial \theta_2}(\hat{\theta}_1, \hat{\theta}_2) = 0 \end{array} \right.$$

Introduzione teorica al Machine Learning

Problema di ottimizzazione

Un esempio...

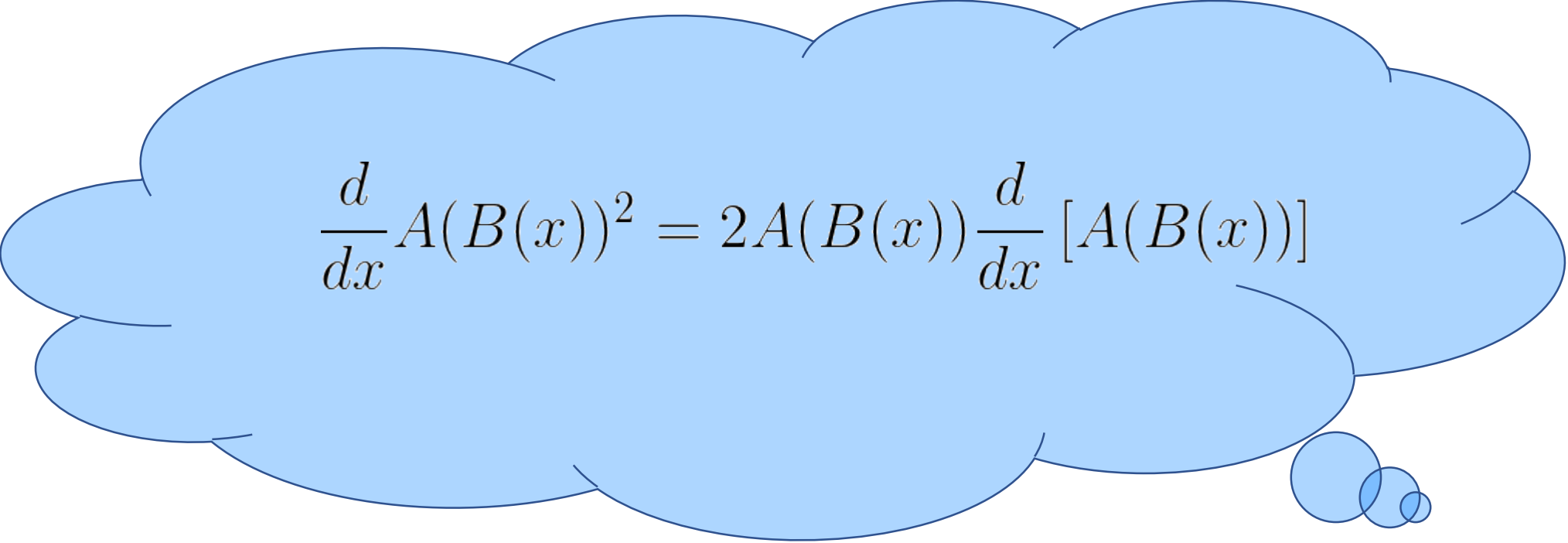
$$\frac{\partial \mathcal{R}}{\partial \theta_1}(\hat{\theta}_1, \hat{\theta}_2) = 0$$

$$\mathcal{R}(\theta_1, \theta_2) = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

$$\frac{\partial}{\partial \theta_1} \left[\frac{1}{N} \sum_{i=1}^N \left(y_i - f(\mathbf{x}_i, \hat{\theta}_1, \hat{\theta}_2) \right)^2 \right] = 0$$

Introduzione teorica al Machine Learning

Problema di ottimizzazione


$$\frac{d}{dx} A(B(x))^2 = 2A(B(x)) \frac{d}{dx} [A(B(x))]$$

Introduzione teorica al Machine Learning

Problema di ottimizzazione

Un esempio...

$$0 = \frac{\partial}{\partial \theta_1} \left[\frac{1}{N} \sum_{i=1}^N \left(y_i - f(\mathbf{x}_i, \hat{\theta}_1, \hat{\theta}_2) \right)^2 \right] = \frac{1}{N} \sum_{i=1}^N \frac{\partial}{\partial \theta_1} \left[\left(y_i - f(\mathbf{x}_i, \hat{\theta}_1, \hat{\theta}_2) \right)^2 \right]$$

$$0 = \frac{2}{N} \sum_{i=1}^N \left(y_i - f(\mathbf{x}_i, \hat{\theta}_1, \hat{\theta}_2) \right) \frac{\partial}{\partial \theta_1} \left(y_i - f(\mathbf{x}_i, \hat{\theta}_1, \hat{\theta}_2) \right)$$

Introduzione teorica al Machine Learning

Problema di ottimizzazione

Un esempio...

$$0 = \frac{2}{N} \sum_{i=1}^N \left(y_i - f(\mathbf{x}_i, \hat{\theta}_1, \hat{\theta}_2) \right) \frac{\partial}{\partial \theta_1} \left(y_i - f(\mathbf{x}_i, \hat{\theta}_1, \hat{\theta}_2) \right)$$

Regressione lineare: $f(\mathbf{x}_i, \hat{\theta}_1, \hat{\theta}_2) = \hat{\theta}_1 \mathbf{x}_i + \hat{\theta}_2$



$$0 = -\frac{2}{N} \sum_{i=1}^N \mathbf{x}_i \left(y_i - \left(\hat{\theta}_1 \mathbf{x}_i + \hat{\theta}_2 \right) \right)$$

Introduzione teorica al Machine Learning

Problema di ottimizzazione

Un esempio con la regressione lineare...

$$0 = -\frac{2}{N} \sum_{i=1}^N \mathbf{x}_i \left(y_i - \left(\hat{\theta}_1 \mathbf{x}_i + \hat{\theta}_2 \right) \right)$$

*Svolgendo calcoli analoghi per l'altro parametro,
si ottengono due equazioni in due incognite*

...

*Per la regressione lin. si possono ottenere **soluzioni esatte**,
generalmente è necessario implementare algoritmi
per la ricerca di **soluzioni approssimate***

Introduzione teorica al Machine Learning

Problema di ottimizzazione

Non sempre i problemi di ottimizzazione sono «liberi»

*In alcuni casi è necessario imporre dei vincoli ai valori che i
parametri possono assumere*

Si parla di problemi di ottimizzazione «vincolati»

Introduzione teorica al Machine Learning

Problema di ottimizzazione vincolato

Un esempio...

$$(\hat{\theta}_1, \hat{\theta}_2) = \arg \min_{(\theta_1, \theta_2)} \mathcal{R}(\theta_1, \theta_2)$$

$$\mathcal{R}(\theta_1, \theta_2) = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

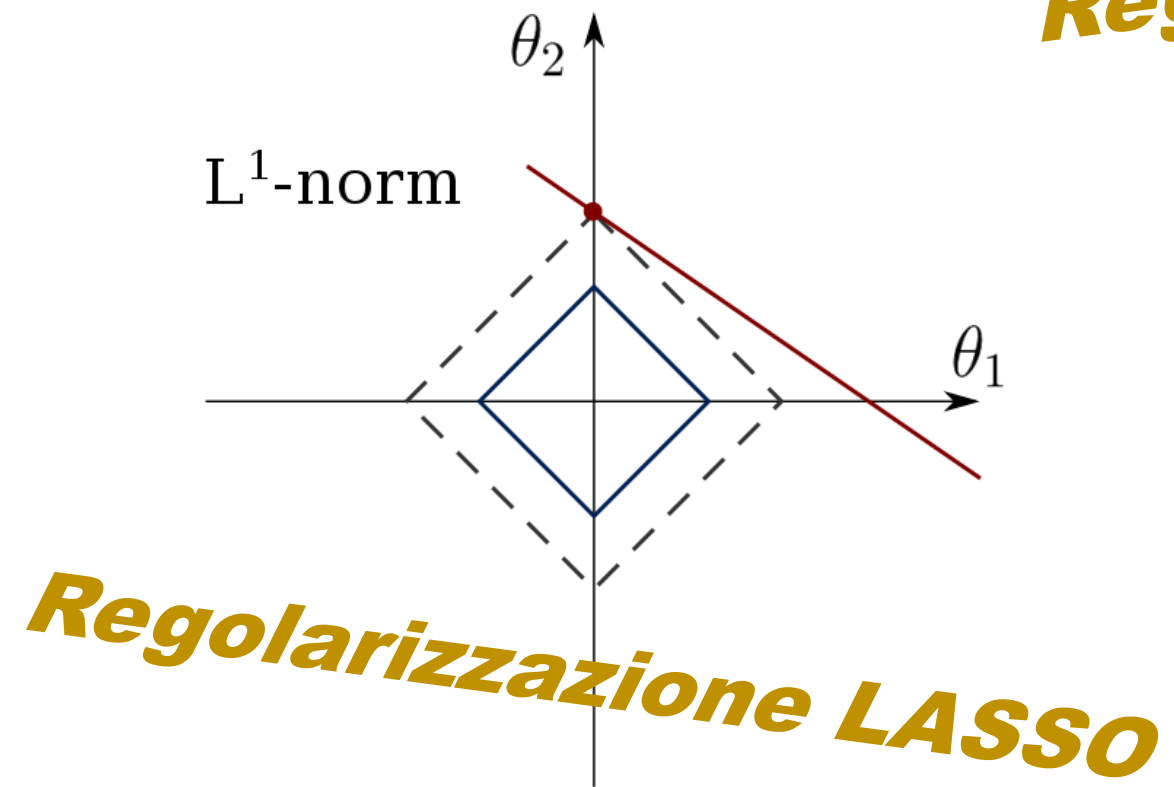
$$\text{con } (\theta_1)^2 + (\theta_2)^2 \leq 1$$

Regolarizzazione RIDGE

Introduzione teorica al Machine Learning

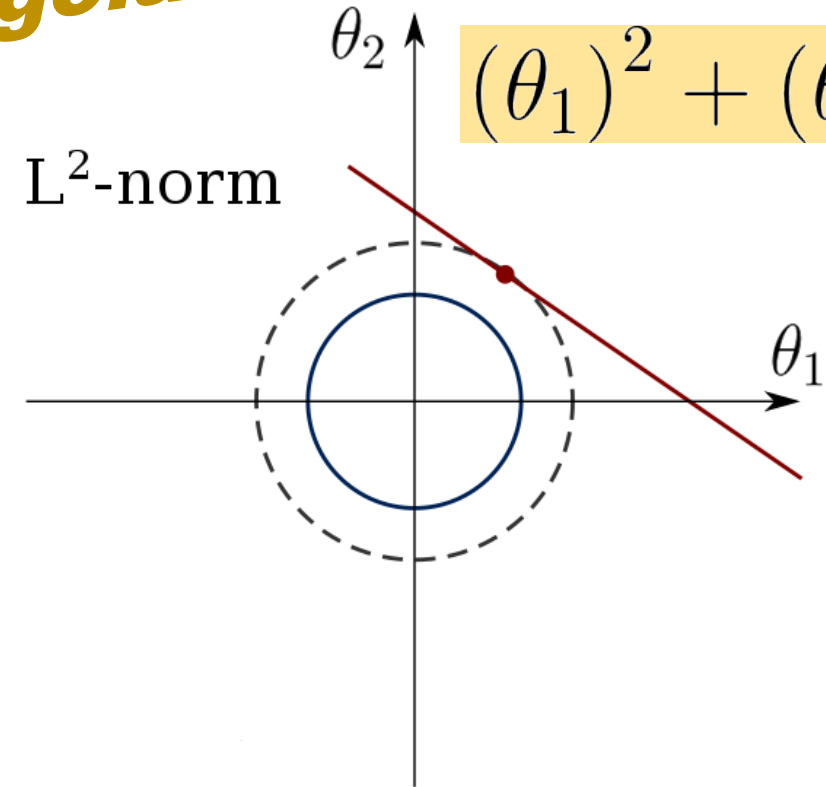
Problema di ottimizzazione vincolato

Un esempio...



Regularizzazione RIDGE

$(\theta_1)^2 + (\theta_2)^2 \leq 1$



Introduzione teorica al Machine Learning

Moltiplicatori di Lagrange

Il metodo dei **moltiplicatori di Lagrange** è una tecnica per studiare i massimi e minimi vincolati di una funzione a più variabili in riferimento ad un vincolo espresso mediante una o più equazioni, che individuano il vincolo come luogo geometrico di zeri.

youmath.it



problema vincolato

problema libero

Introduzione teorica al Machine Learning

Moltiplicatori di Lagrange

youmath.it

Moltiplicatori di Lagrange in due variabili con un vincolo

Sia $f : A \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}$ una funzione definita su un aperto $A \subseteq \mathbb{R}^2$, e sia $g(x, y) = 0$ un vincolo espresso sotto forma di **luogo geometrico**. Supponiamo che $f, g \in C^1(A)$, ossia che siano funzioni che ammettono **derivate parziali** continue su A .

Condizione necessaria ma non sufficiente affinché $(x_0, y_0) \in A$ sia un punto di estremo relativo per f rispetto al vincolo $g(x, y) = 0$ è che sussistano le seguenti condizioni:

1) $g(x_0, y_0) = 0$ e che inoltre il **gradiente** di g in (x_0, y_0) non sia nullo: $\nabla g(x_0, y_0) \neq 0$

2) Definita la *funzione lagrangiana*

$$L(x, y, \lambda) = f(x, y) - \lambda g(x, y)$$

esista un valore reale λ_0 tale per cui sia nullo il gradiente di L in (x_0, y_0, λ_0)

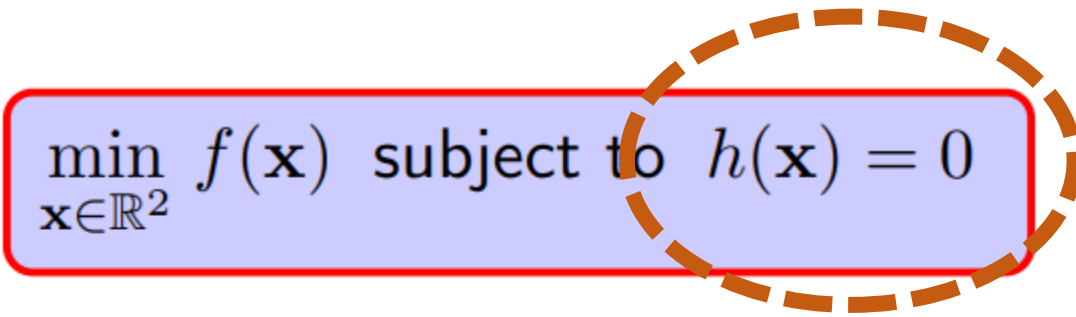
$$\nabla L(x_0, y_0, \lambda_0)$$

In particolare la variabile λ è detta *moltiplicatore di Lagrange*.

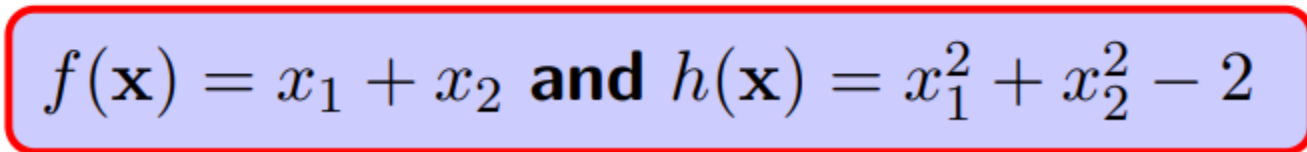
Introduzione teorica al Machine Learning

Problema di ottimizzazione vincolato

Un esempio...


$$\min_{\mathbf{x} \in \mathbb{R}^2} f(\mathbf{x}) \text{ subject to } h(\mathbf{x}) = 0$$

con

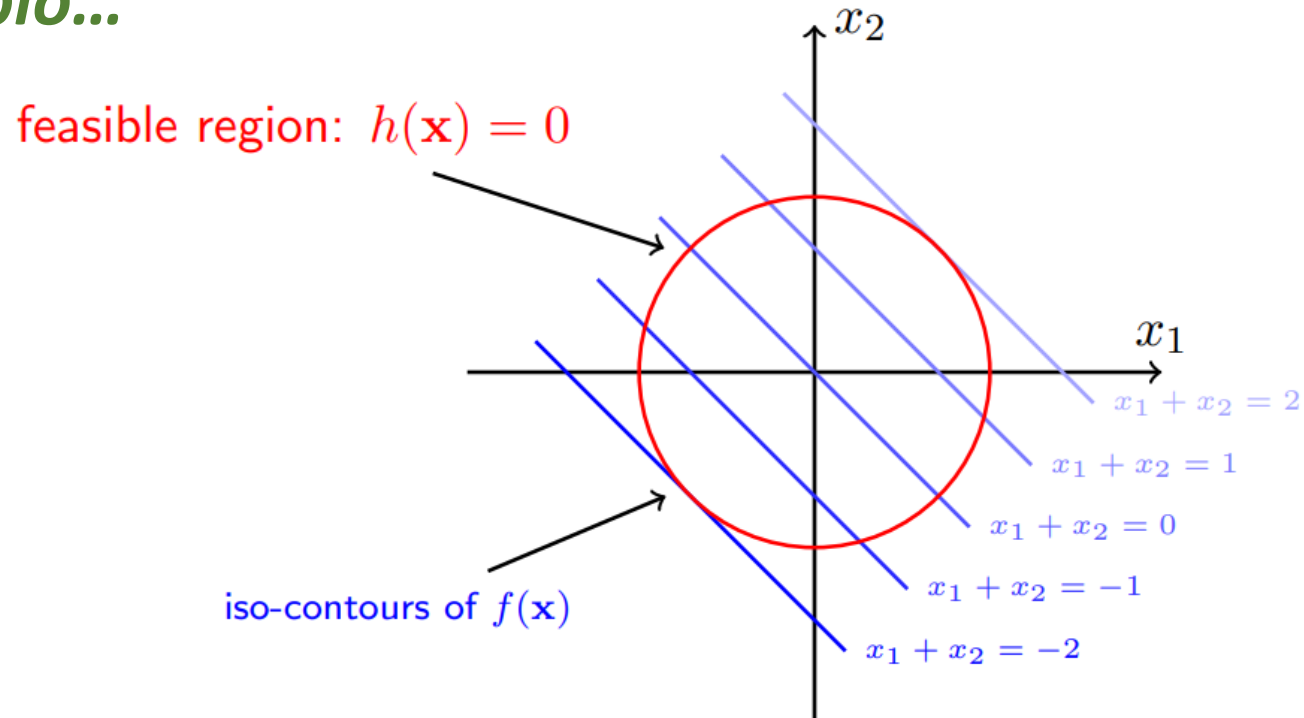

$$f(\mathbf{x}) = x_1 + x_2 \text{ and } h(\mathbf{x}) = x_1^2 + x_2^2 - 2$$

“Lagrange Multipliers and the Karush-Kuhn-Tucker (KKT) conditions”, deleeuwpx.net/pubfolders/dual/KKT.pdf

Introduzione teorica al Machine Learning

Problema di ottimizzazione vincolato

Un esempio...



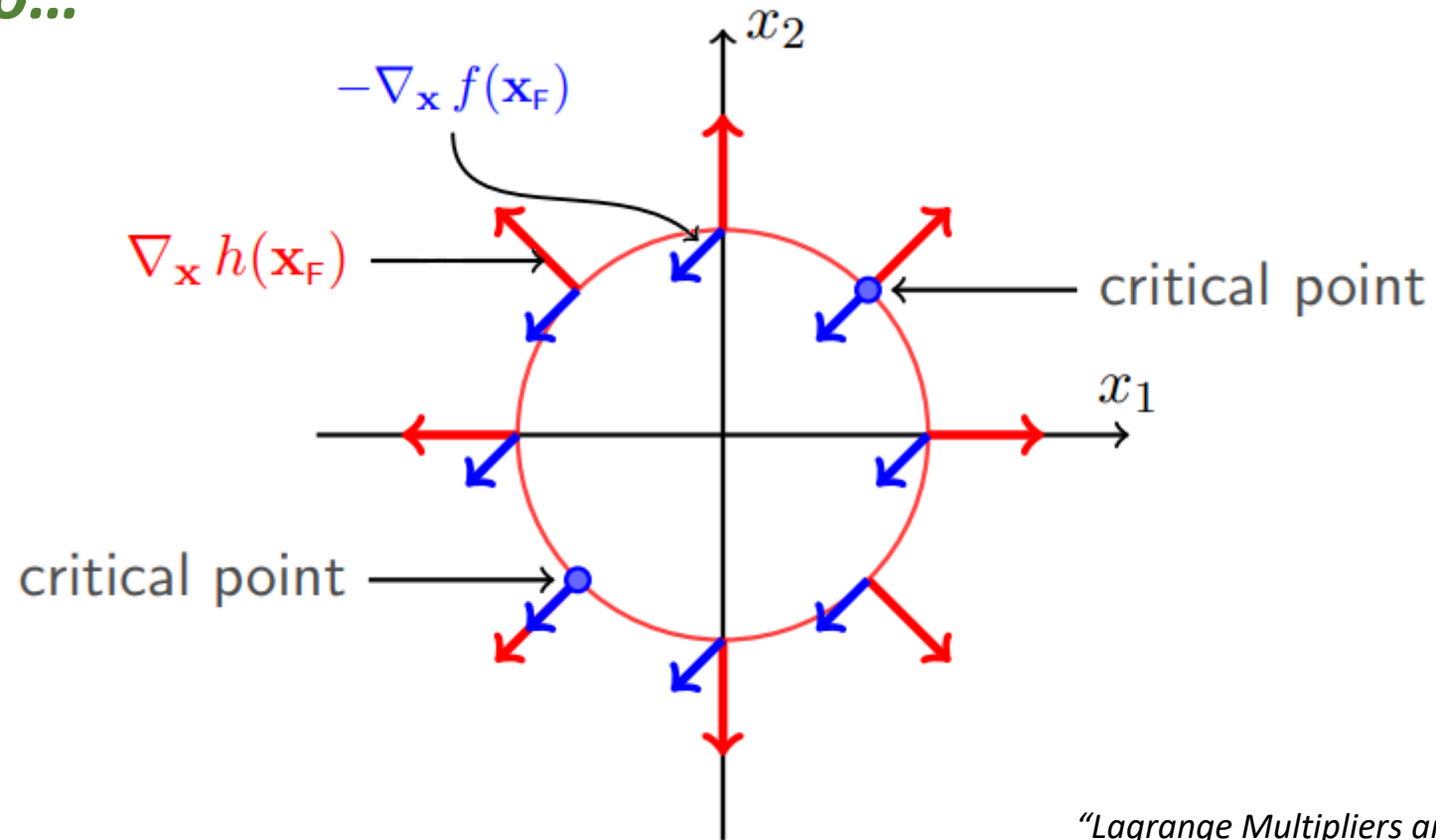
$$h(\mathbf{x}) = x_1^2 + x_2^2 - 2$$

“Lagrange Multipliers and the Karush-Kuhn-Tucker (KKT) conditions”, deleeuwpx.net/pubfolders/dual/KKT.pdf

Introduzione teorica al Machine Learning

Problema di ottimizzazione vincolato

Un esempio...



"Lagrange Multipliers and the Karush-Kuhn-Tucker (KKT) conditions", deleeuwpxd.net/pubfolders/dual/KKT.pdf

Introduzione teorica al Machine Learning

Problema di ottimizzazione vincolato

Un esempio...

$$\min_{\mathbf{x} \in \mathbb{R}^2} f(\mathbf{x}) \text{ subject to } h(\mathbf{x}) = 0$$

Define the **Lagrangian** as

$$\mathcal{L}(\mathbf{x}, \mu) = f(\mathbf{x}) + \mu h(\mathbf{x})$$

Then \mathbf{x}^* a local minimum \iff there exists a unique μ^* s.t.

$$\textcircled{1} \quad \nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}^*, \mu^*) = 0$$

$$\textcircled{2} \quad \nabla_{\mu} \mathcal{L}(\mathbf{x}^*, \mu^*) = 0$$

$$\textcircled{3} \quad \mathbf{y}^t (\nabla_{\mathbf{x}\mathbf{x}}^2 \mathcal{L}(\mathbf{x}^*, \mu^*)) \mathbf{y} \geq 0 \quad \forall \mathbf{y} \text{ s.t. } \nabla_{\mathbf{x}} h(\mathbf{x}^*)^t \mathbf{y} = 0$$

“Lagrange Multipliers and the Karush-Kuhn-Tucker (KKT) conditions”, deleeuwpx.net/pubfolders/dual/KKT.pdf

Introduzione teorica al Machine Learning

Problema di ottimizzazione vincolato

Una generalizzazione per vincoli con disuguaglianze:

Condizioni KKT

$$\min_{x \in \mathbb{R}} f(x) \quad \text{con} \quad g(x) \leq 0 \quad \Leftrightarrow \quad \min_{x \in \mathbb{R}} [f(x) + \lambda g(x)], \quad \lambda \geq 0$$

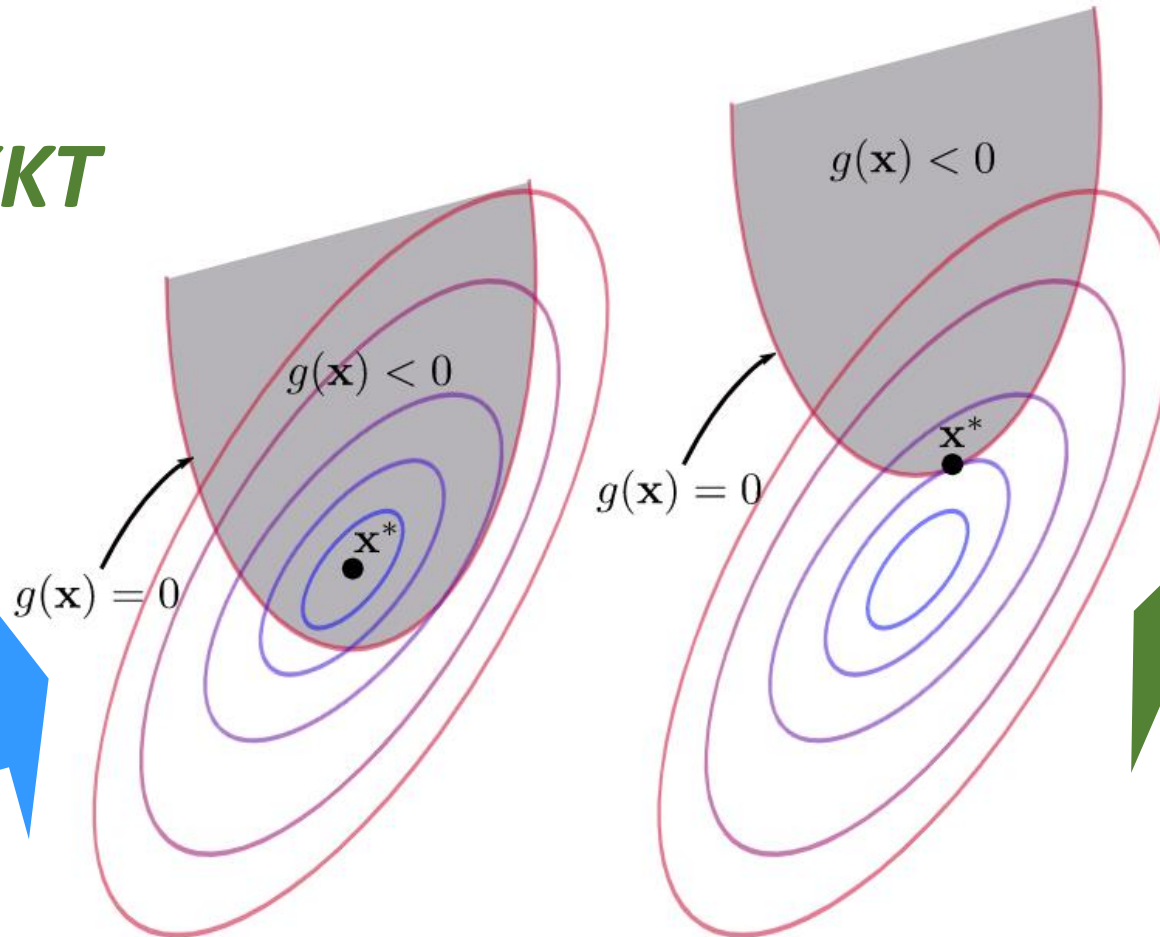
“Karush-Kuhn-Tucker (KKT) conditions”, onmyphd.com

Introduzione teorica al Machine Learning

Problema di ottimizzazione vincolato

Condizioni KKT

Se il punto di minimo libero è dentro la regione di vincolo allora il problema di ottimizzazione può essere trattato come libero



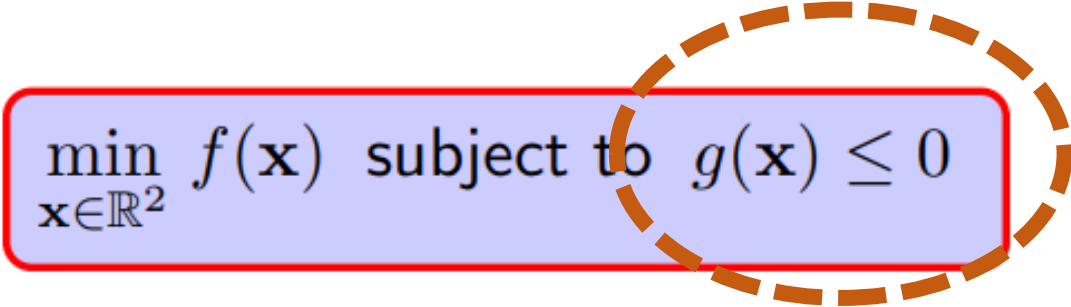
Se il punto di minimo globale libero è fuori la regione di vincolo, allora il punto che risolve il problema di ottimizzazione vincolato si trova sul bordo della regione

“Karush-Kuhn-Tucker (KKT) conditions”, onmyphd.com

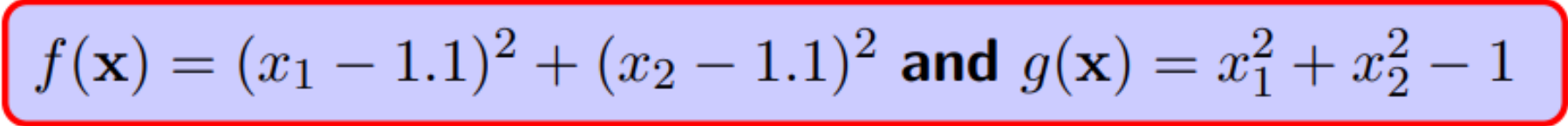
Introduzione teorica al Machine Learning

Problema di ottimizzazione vincolato

Un esempio...


$$\min_{\mathbf{x} \in \mathbb{R}^2} f(\mathbf{x}) \text{ subject to } g(\mathbf{x}) \leq 0$$

con

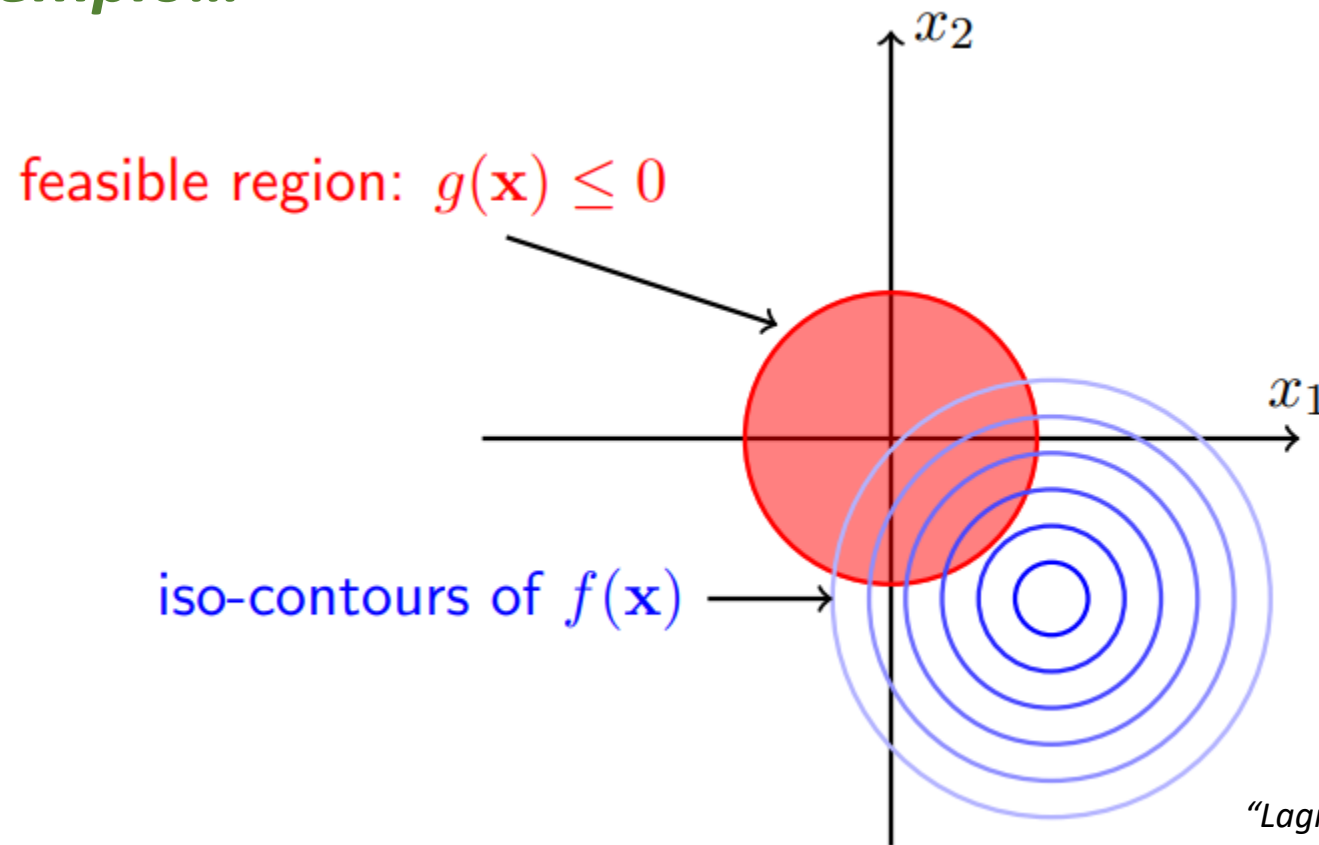

$$f(\mathbf{x}) = (x_1 - 1.1)^2 + (x_2 - 1.1)^2 \text{ and } g(\mathbf{x}) = x_1^2 + x_2^2 - 1$$

"Lagrange Multipliers and the Karush-Kuhn-Tucker (KKT) conditions", deleeuwpx.net/pubfolders/dual/KKT.pdf

Introduzione teorica al Machine Learning

Problema di ottimizzazione vincolato

Un esempio...

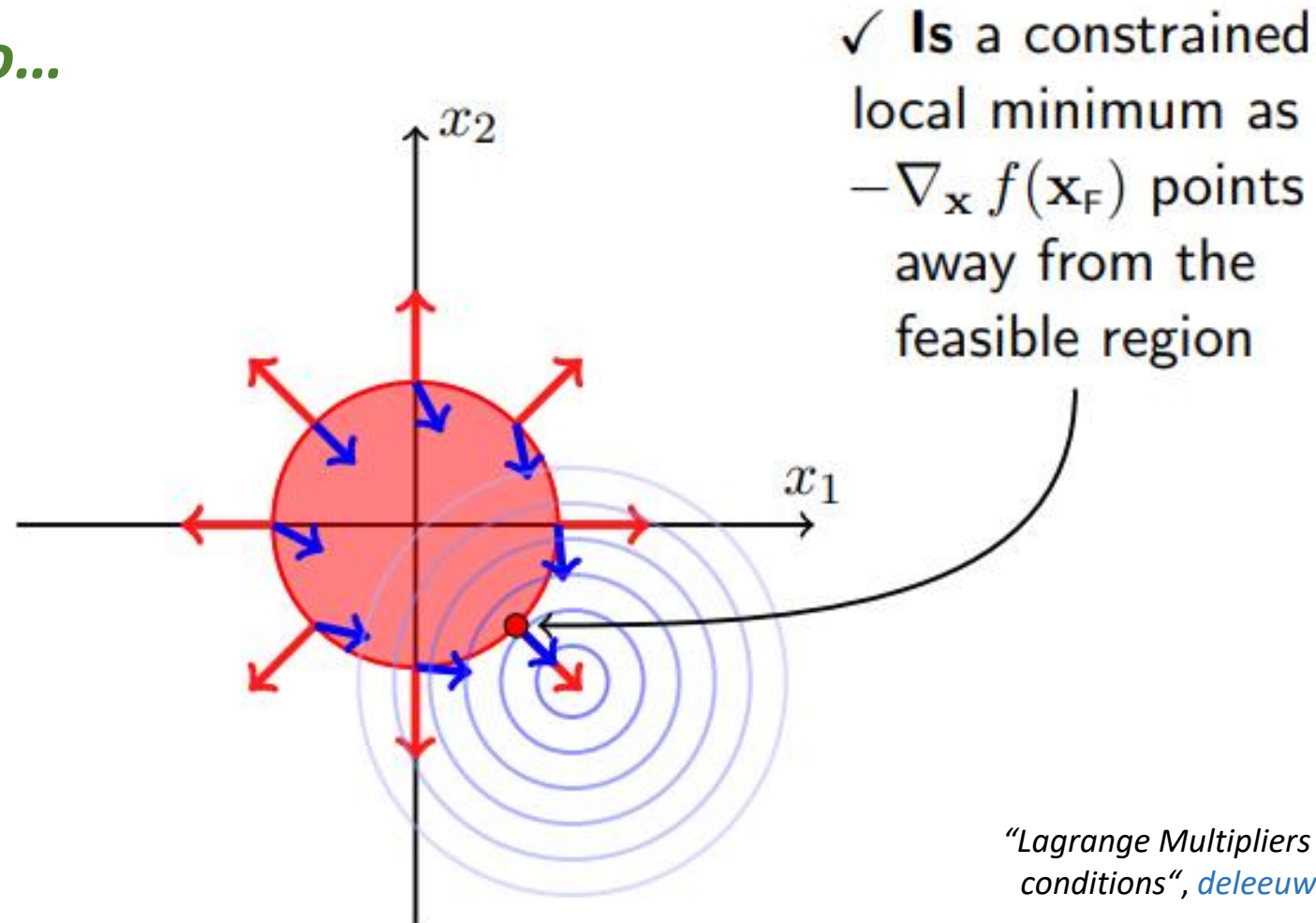


"Lagrange Multipliers and the Karush-Kuhn-Tucker (KKT) conditions", deleeuwpx.net/pubfolders/dual/KKT.pdf

Introduzione teorica al Machine Learning

Problema di ottimizzazione vincolato

Un esempio...



“Lagrange Multipliers and the Karush-Kuhn-Tucker (KKT) conditions”, deleeuwpx.net/pubfolders/dual/KKT.pdf

Introduzione teorica al Machine Learning

Problema di ottimizzazione vincolato

Un esempio...

$$\min_{\mathbf{x} \in \mathbb{R}^2} f(\mathbf{x}) \text{ subject to } g(\mathbf{x}) \leq 0$$

Define the **Lagrangian** as

$$\mathcal{L}(\mathbf{x}, \lambda) = f(\mathbf{x}) + \lambda g(\mathbf{x})$$

Then \mathbf{x}^* a local minimum \iff there exists a unique λ^* s.t.

$$\textcircled{1} \quad \nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}^*, \lambda^*) = 0$$

$$\textcircled{2} \quad \lambda^* \geq 0$$

$$\textcircled{3} \quad \lambda^* g(\mathbf{x}^*) = 0$$

$$\textcircled{4} \quad g(\mathbf{x}^*) \leq 0$$

$$\textcircled{5} \quad \text{Plus positive definite constraints on } \nabla_{\mathbf{xx}} \mathcal{L}(\mathbf{x}^*, \lambda^*).$$

“Lagrange Multipliers and the Karush-Kuhn-Tucker (KKT) conditions”, deleeuwpx.net/pubfolders/dual/KKT.pdf

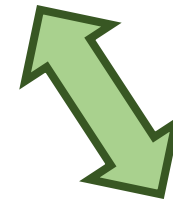
Introduzione teorica al Machine Learning

Problema di ottimizzazione vincolato

Un esempio... **Regolarizzazione RIDGE**

$$\min_{(\theta_1, \theta_2)} \mathcal{R}(\theta_1, \theta_2) \quad \text{con} \quad (\theta_1)^2 + (\theta_2)^2 \leq t$$

$$\mathcal{R}(\theta_1, \theta_2) = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

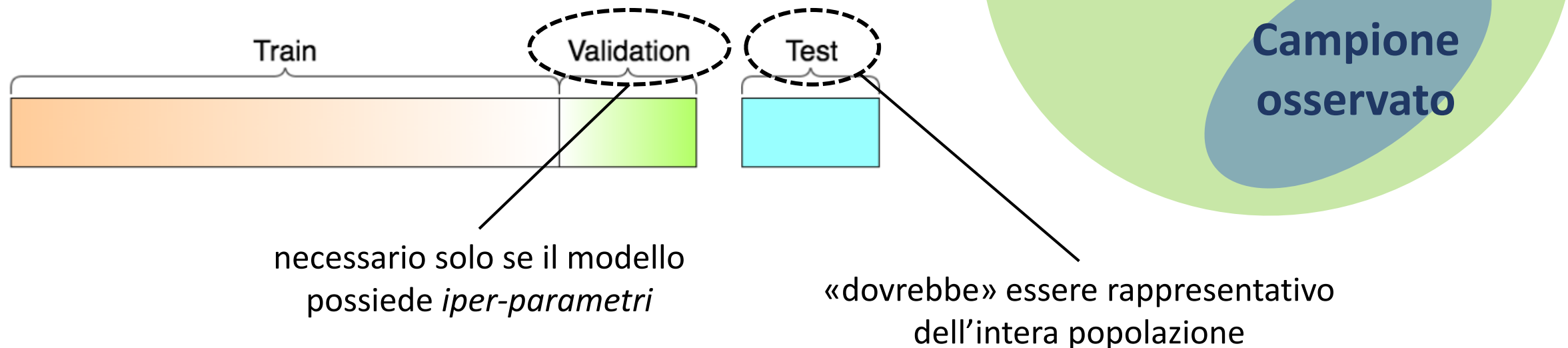


$$\min_{(\theta_1, \theta_2)} \left[\mathcal{R}(\theta_1, \theta_2) + \lambda [(\theta_1)^2 + (\theta_2)^2] \right], \quad \lambda \geq 0$$

Introduzione teorica al Machine Learning

Apprendimento supervisionato

Validazione delle capacità predittive del modello

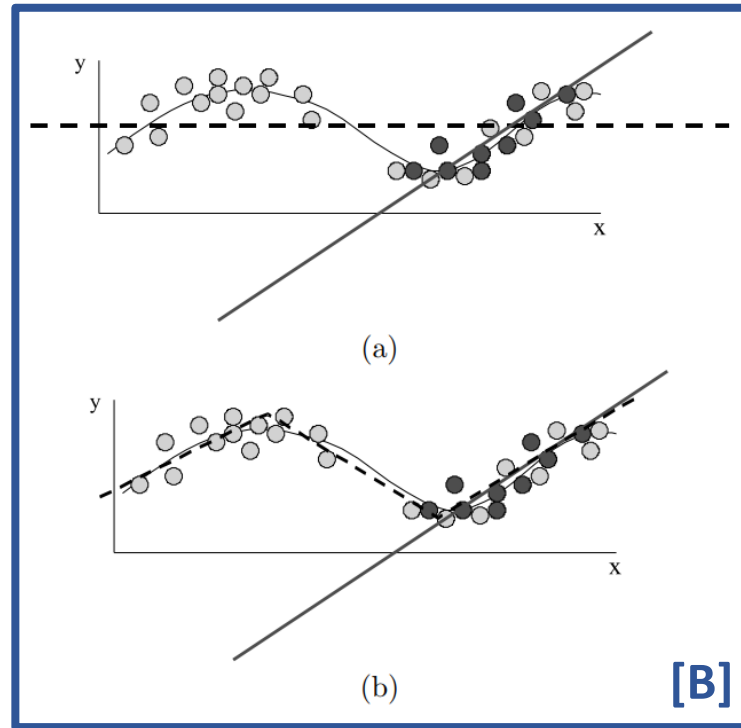


Introduzione teorica al Machine Learning

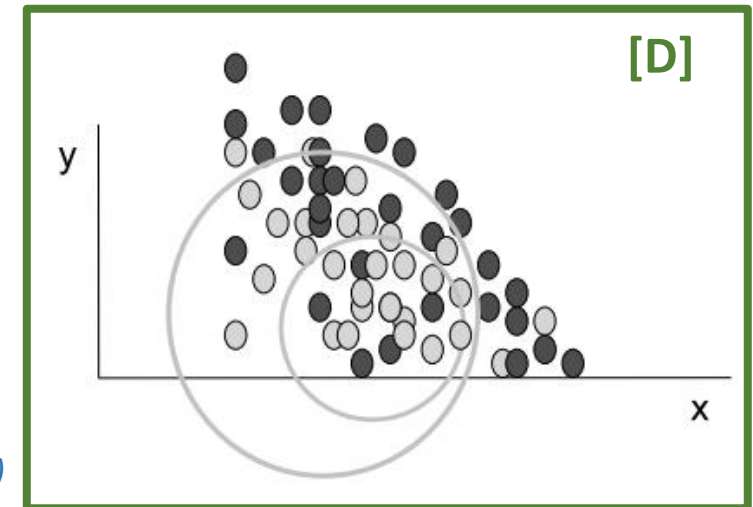
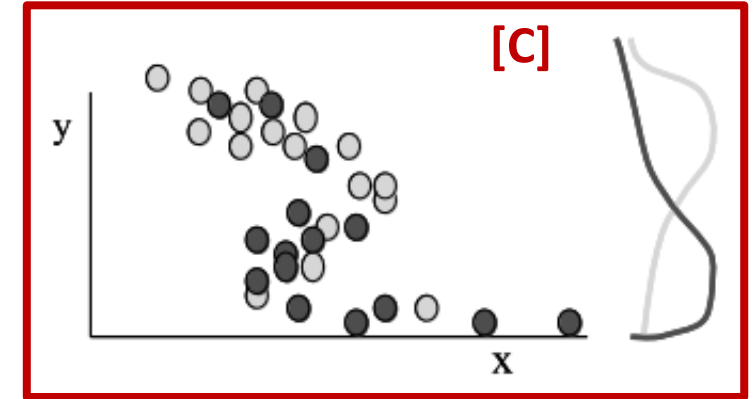
Problematiche nel campionamento

Problemi

- A. Errori nel partizionamento del dataset
- B. Covariate shift
- C. Probability shift
- D. Selection bias



“When training and test sets are different: characterizing learning transfer”, *Storkey, Amos (2013)*



Introduzione teorica al Machine Learning

Problematiche nel campionamento

Un esempio di ... ***Errore nel partizionamento del dataset***

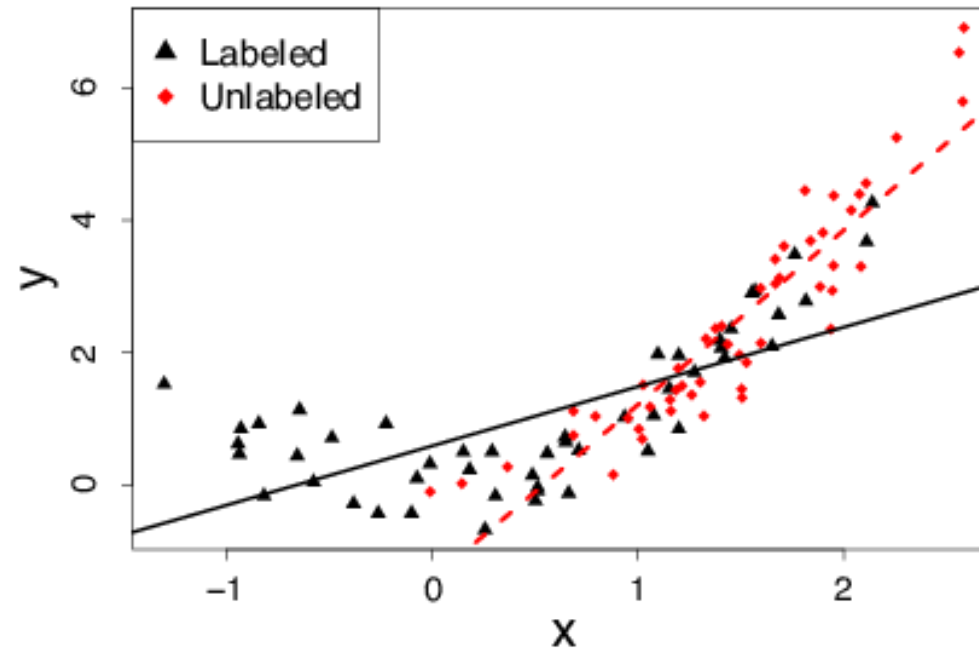


“Dataset Shift in Classification: Approaches and Problems”, [Francisco Herrera \(IWANN\)](#)

Introduzione teorica al Machine Learning

Problematiche nel campionamento

Un esempio di ... **Covariate Shift**

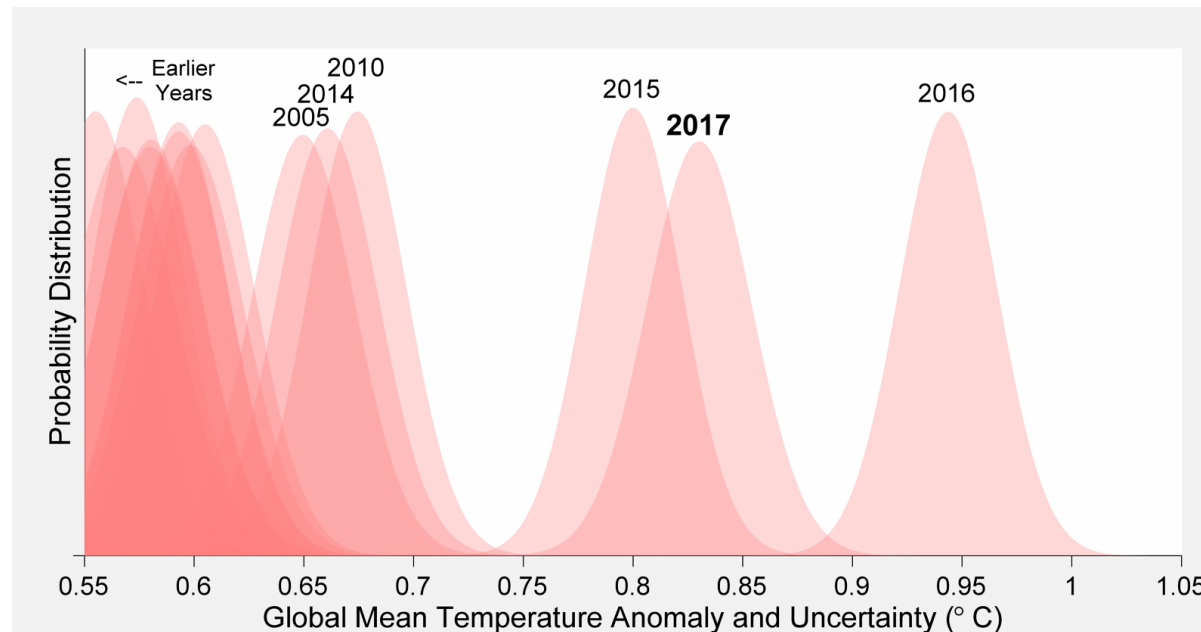


“Toy example for covariate shift in linear regression”,
researchgate.net

Introduzione teorica al Machine Learning

Problematiche nel campionamento

Un esempio di ... ***Probability Shift***



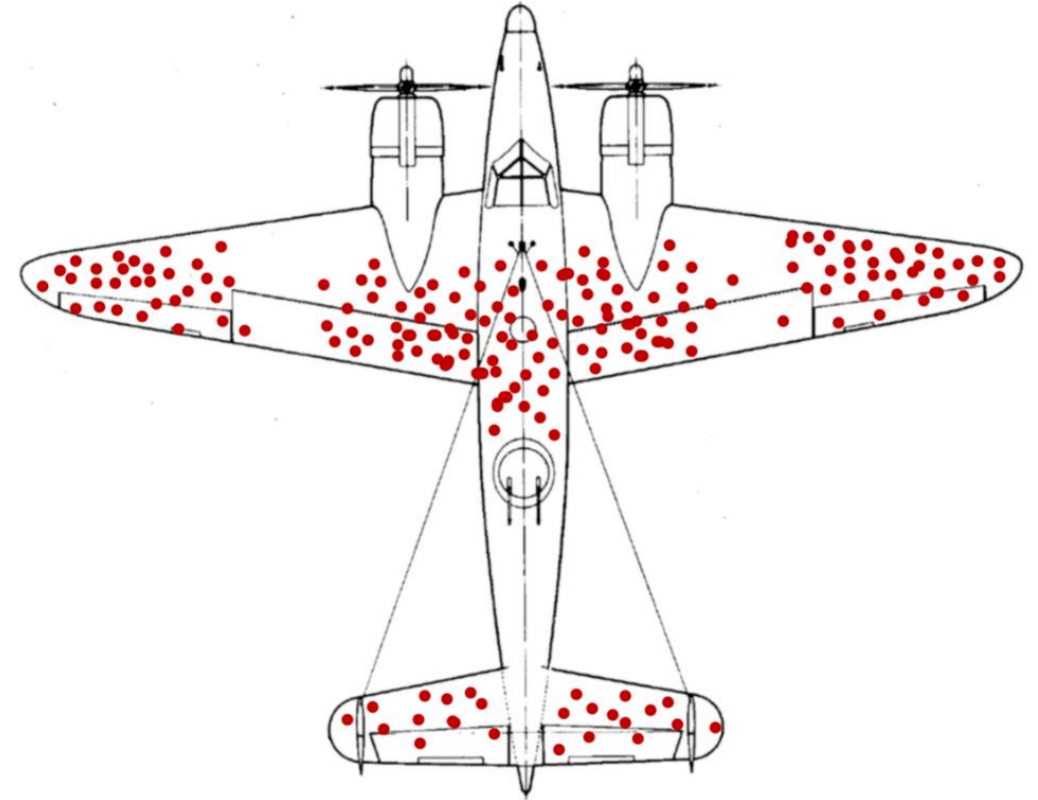
*“Based on Berkeley Earth’s estimates of the global annual average temperature increase relative to 1951-1980.”,
berkeleyearth.org/global-temperatures-2017*

Introduzione teorica al Machine Learning

Problematiche nel campionamento

Un esempio di ... **Selection Bias**
in particolare di ... **Survivorship Bias**

“Planes coming home from battle have bullet holes everywhere but the engine and cockpit, so we should put armor everywhere but the engine and cockpit.”



“Damage taken by planes able to come back after the fight. Image shows hypothetical data.”,

en.wikipedia.org

Introduzione teorica al Machine Learning

Bias-Variance trade-off

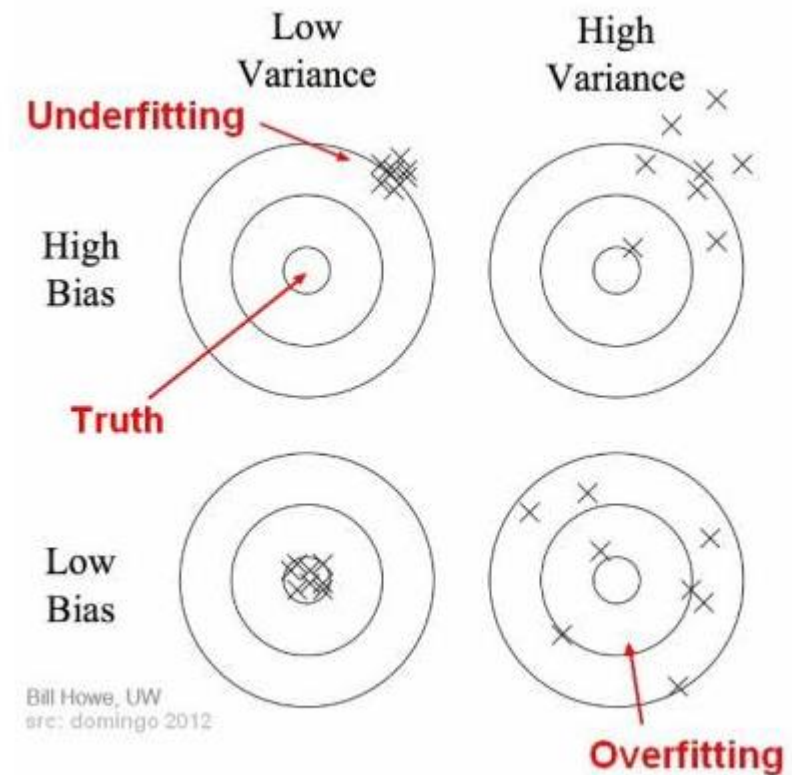
L'errore quadratico medio prodotto dal modello si può scrivere come

$$Err(x) = E \left[(Y - \hat{f}(x))^2 \right]$$

che può essere decomposto

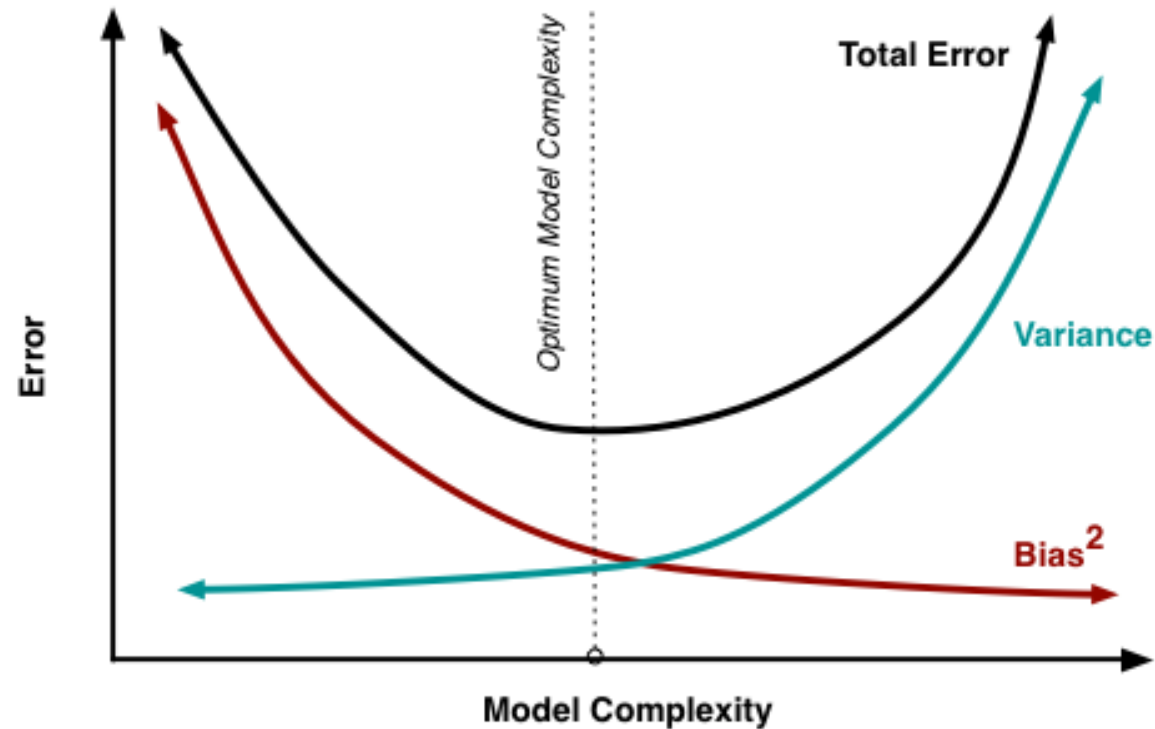
$$Err(x) = \left(E[\hat{f}(x)] - f(x) \right)^2 + E \left[\left(\hat{f}(x) - E[\hat{f}(x)] \right)^2 \right] + \sigma_e^2$$

$$Err(x) = \text{Bias}^2 + \text{Variance} + \text{Irreducible Error}$$



Introduzione teorica al Machine Learning

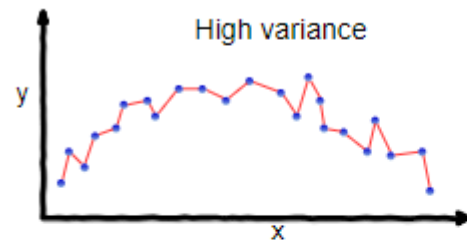
Bias-Variance trade-off



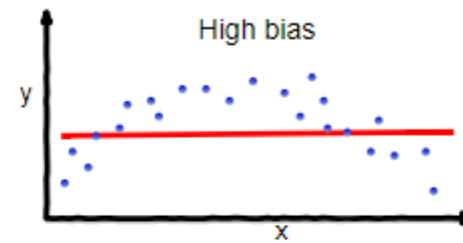
"Understanding the Bias-Variance Tradeoff", Bryan White

Introduzione teorica al Machine Learning

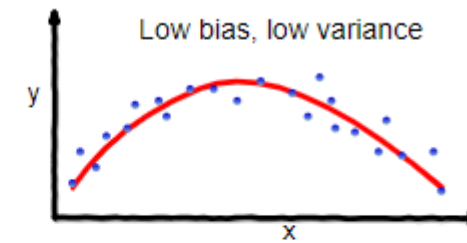
Bias-Variance trade-off



overfitting



underfitting



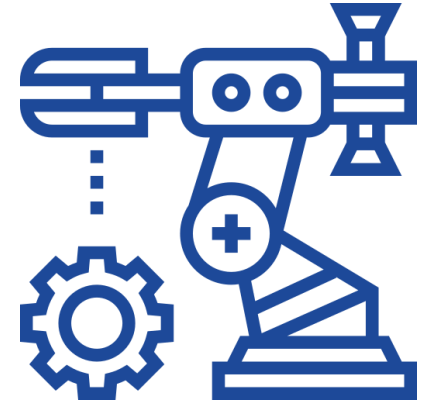
Good balance

Introduzione teorica al Machine Learning

Bias-Variance trade-off

Da Wikipedia:

- **Linear and Generalized linear** models can be regularized to decrease their variance at the cost of increasing their bias
- In **artificial neural networks**, the variance increases and the bias decreases as the number of hidden units increase (regularization applied)
- In ***k*-nearest neighbor** models, a high value of *k* leads to high bias and low variance
- In **decision trees** the depth of the tree determines the variance (pruned to control variance)



Grazie dell'attenzione

Fabio Mardero

fabio.mardero@gmail.com