



UNIVERSITÀ DI PAVIA  
Dipartimento di Fisica



Istituto Nazionale di Fisica Nucleare

# Introduzione al Machine Learning

Incontri di Fisica Moderna

Dr. Giacomo Polesello, INFN sez. Pavia

Dr. Ian Postuma, INFN sez. Pavia

- \* situazione informatica nel 2006 (quando ho iniziato io fisica) che pc c'erano, quali app e quali telefoni
- \* situazione informatica ora nel 2023
- \* cambia tutto, ma cosa non cambia ?
- \* matematica, analisi, Algebra
- \* qui iniziare a mettere la pulce che algebra e AI sono legate
- \* mostrare cos'è una rete neurale e cosa vuol dire ottimizzarla (come a lezione)
- \* alla fine sono tutte operazioni su matrici che possono essere codificate su GPU
- \* io per cosa le uso ?
- \* ma la mia non è proprio fisica... in fisica le reti neurali dove vengono usate ?

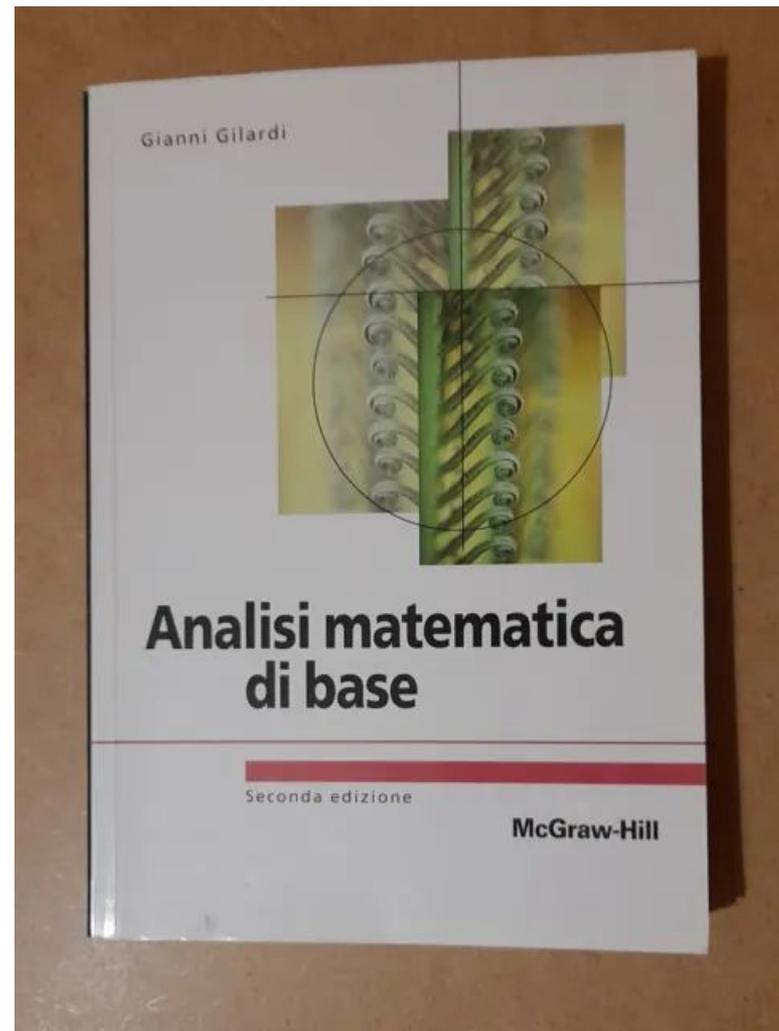
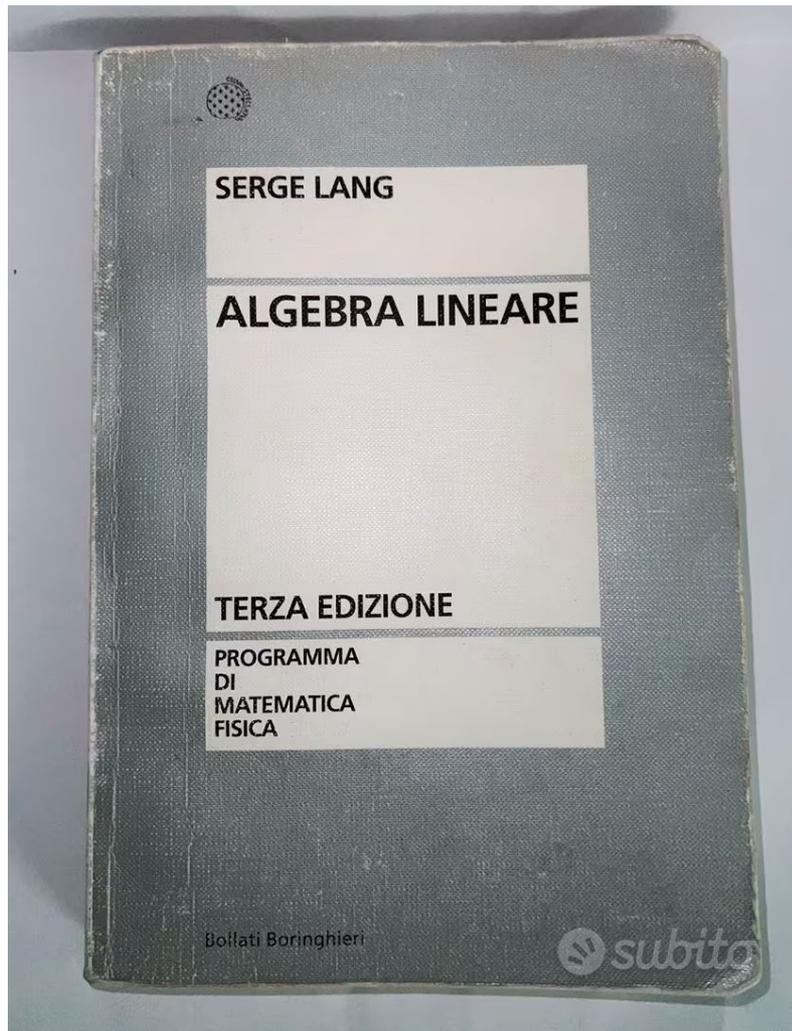
2006





CPU: 2 core  
RAM: 1 GB  
HD: 120 GB  
GPU: 256 MB

OS: ~~windows xp~~  
ubuntu 6.06



2023



CPU: **8** core

RAM: **8** GB

HD: **256** GB

GPU: **8** GB (shared)

OS: android 13

 framework

<https://frame.work/it/en>



CPU: **16 (32)** core

RAM: **64** GB

HD: **1-16** TB

GPU: **8** GB

OS: ~~windows 11~~

ubuntu 23.04



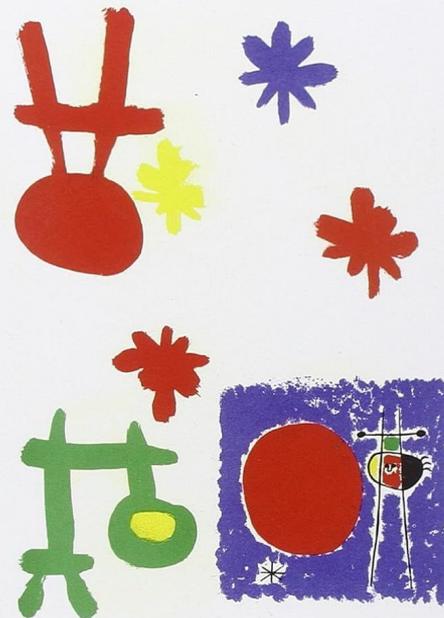
SERGE LANG

# ALGEBRA LINEARE

TERZA EDIZIONE

PROGRAMMA  
DI  
MATEMATICA  
FISICA

Bollati Boringhieri



Carlo Domenico Pagani Sandro Salsa

# Analisi matematica 1

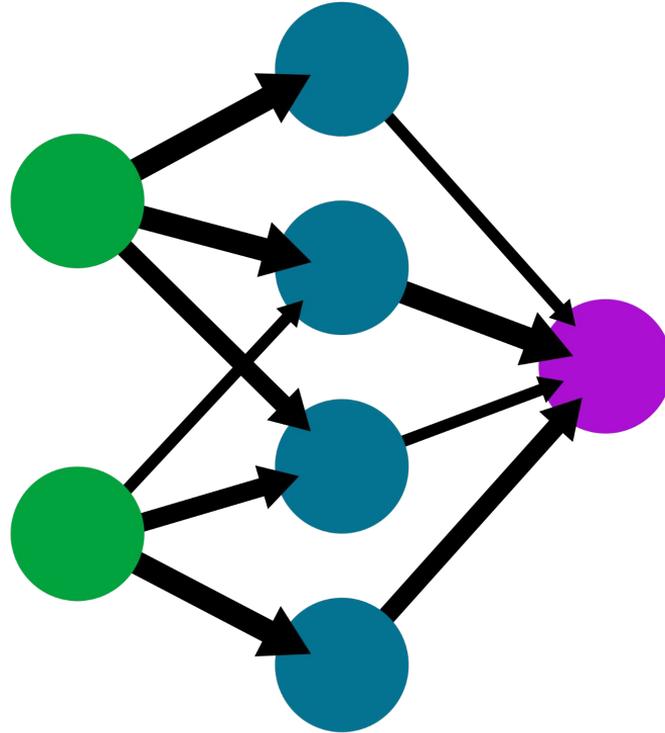
ZANICHELLI

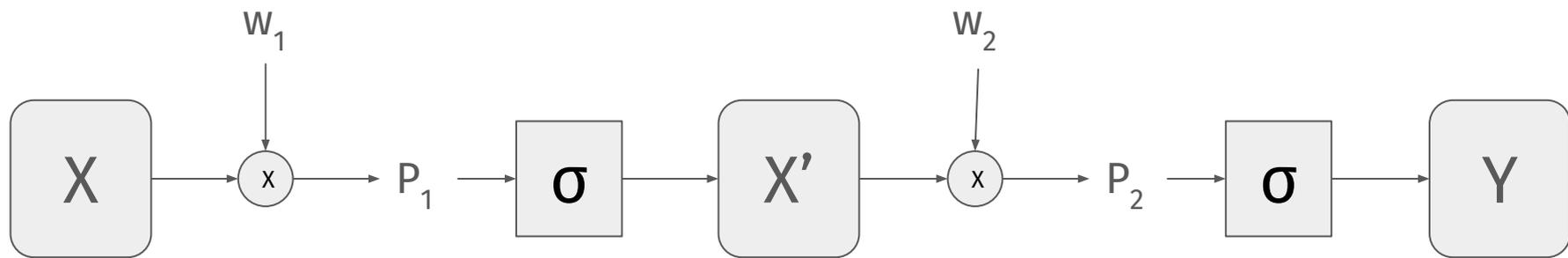
Machine  
Learning

Artificial  
Neural  
Networks

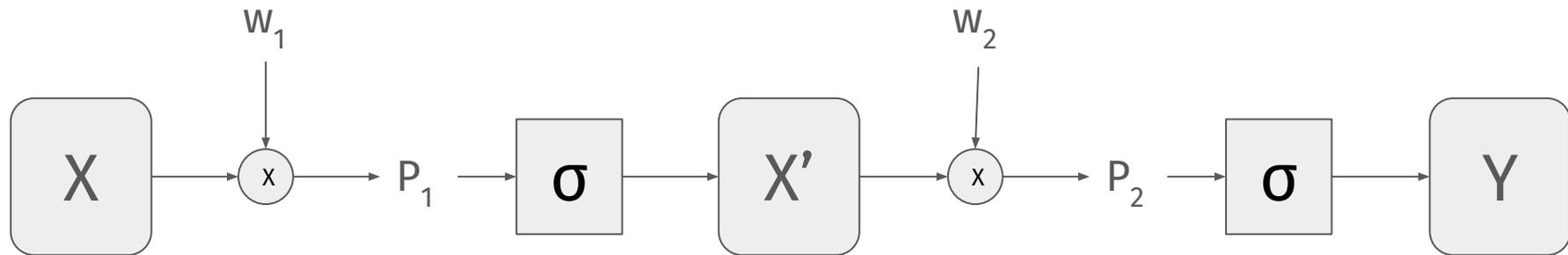
# Semplice rete neurale

strato di    strato    strato di  
ingresso   nascosto   uscita

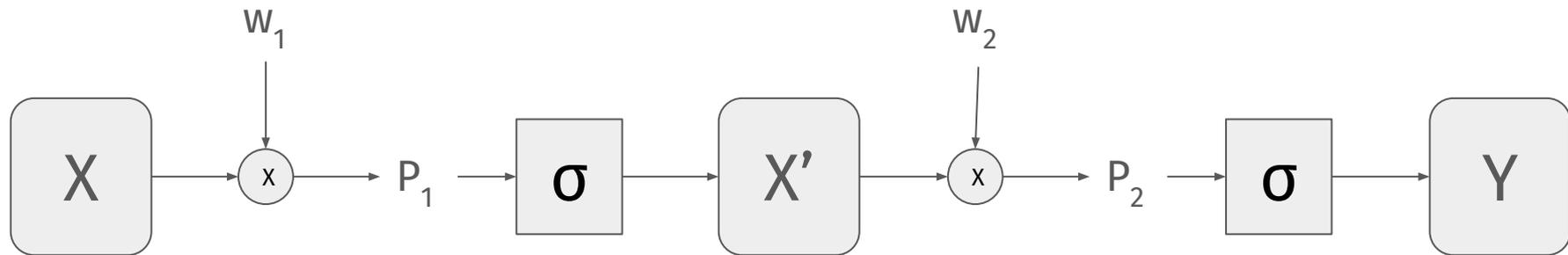




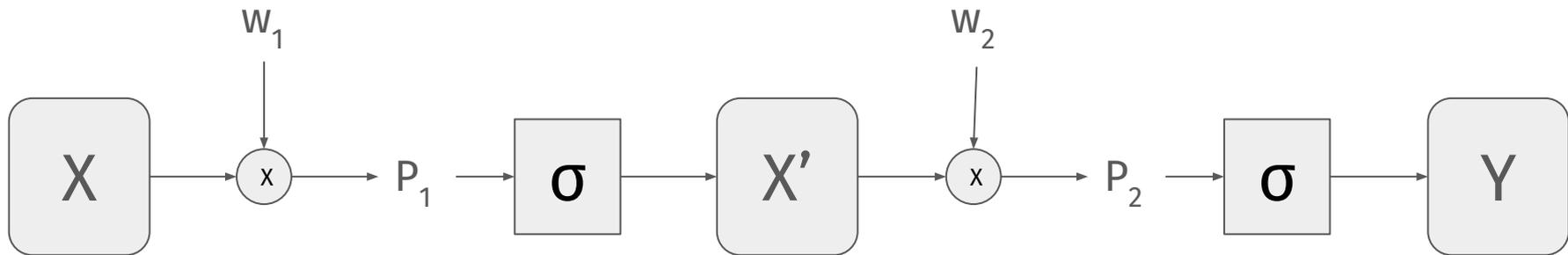
$$\sigma(x) = \frac{1}{1 - e^{-x}}$$

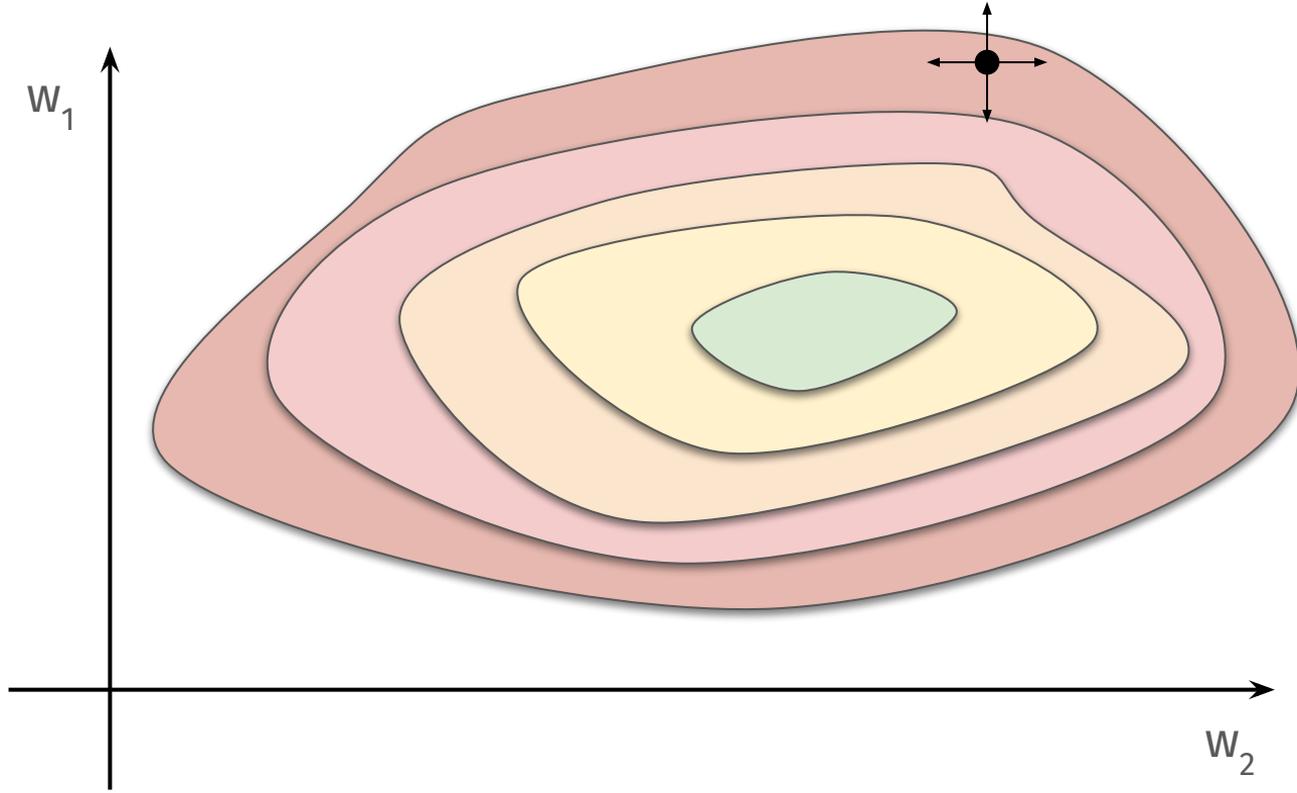


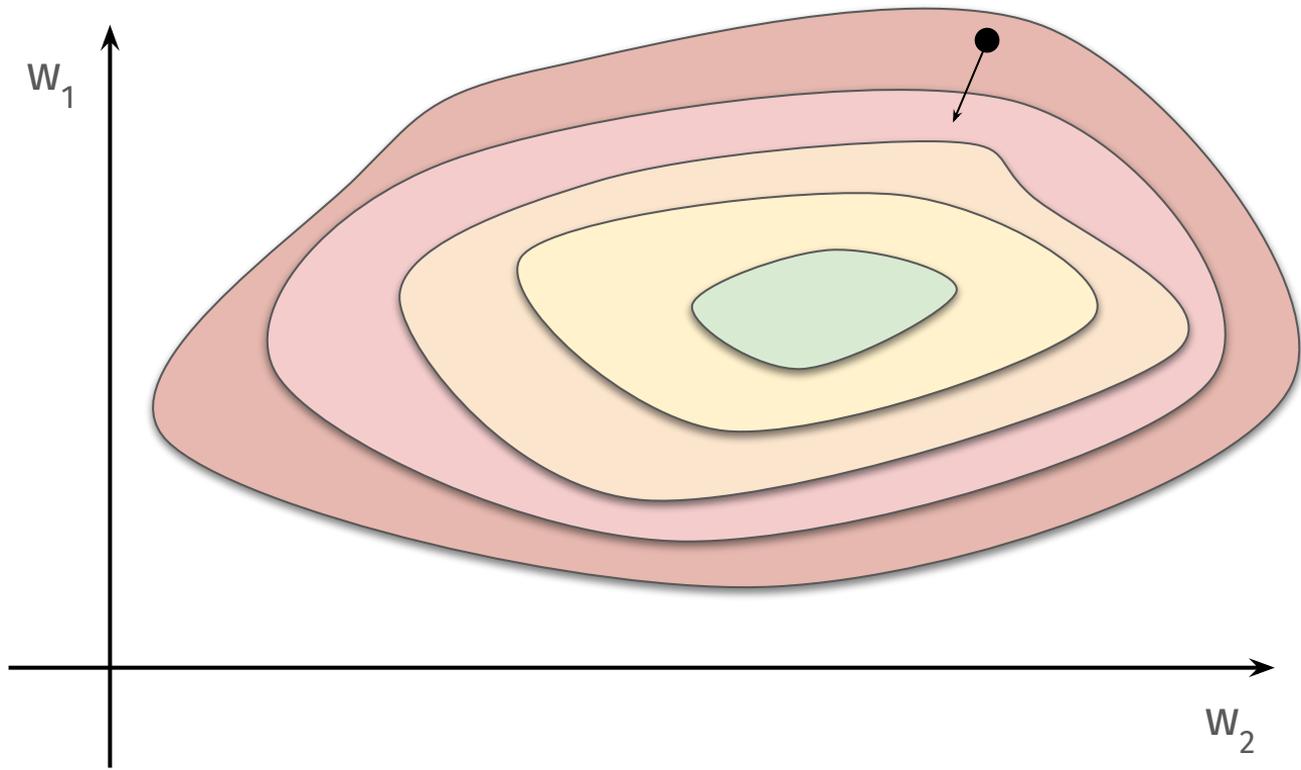
$$Y = \sigma(w_2 \sigma(w_1 X))$$



$$L = \frac{1}{2} (Y - Y_{true})^2$$





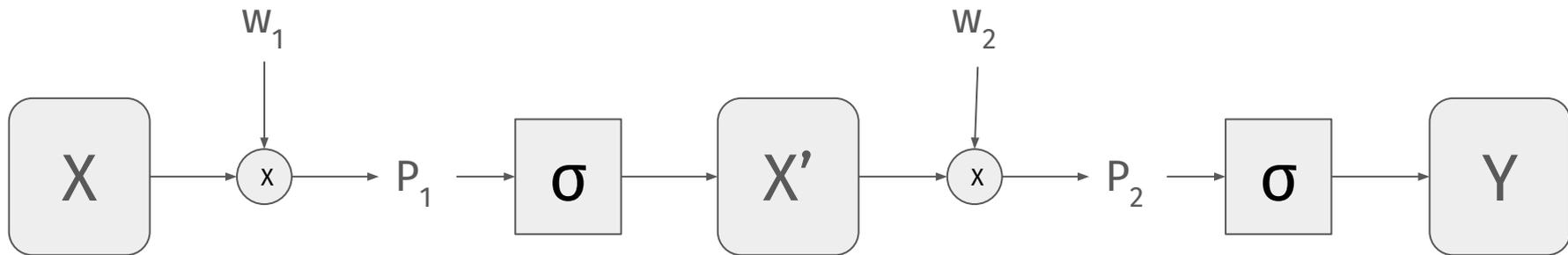


$$\overline{\Delta W} = \frac{\partial L}{\partial w_1} \dot{i} + \frac{\partial L}{\partial w_2} \dot{j}$$

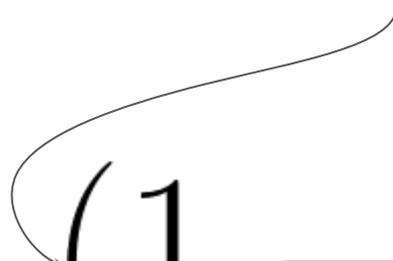
$$\overline{\Delta W} = \left( \frac{\partial L}{\partial w_1} \dot{i} + \frac{\partial L}{\partial w_2} \dot{j} \right) \varepsilon$$

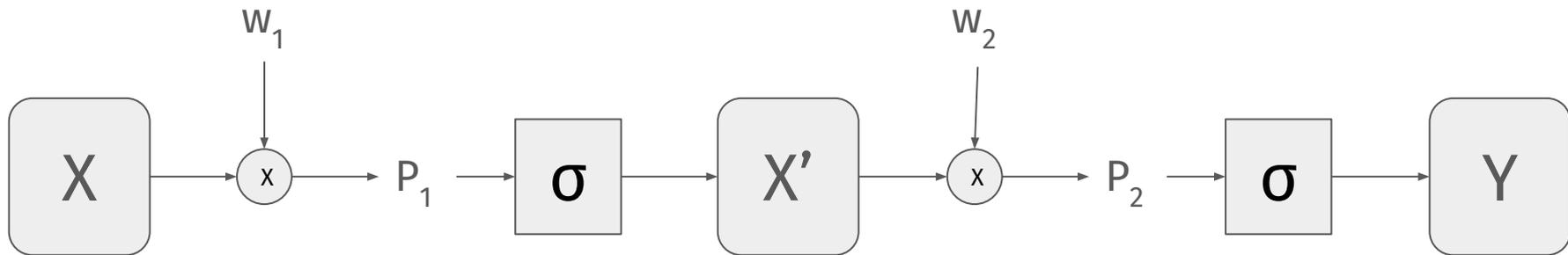
$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial Y} \frac{\partial Y}{\partial P_2} \frac{\partial P_2}{\partial w_2}$$

$$(Y - Y_{true})$$



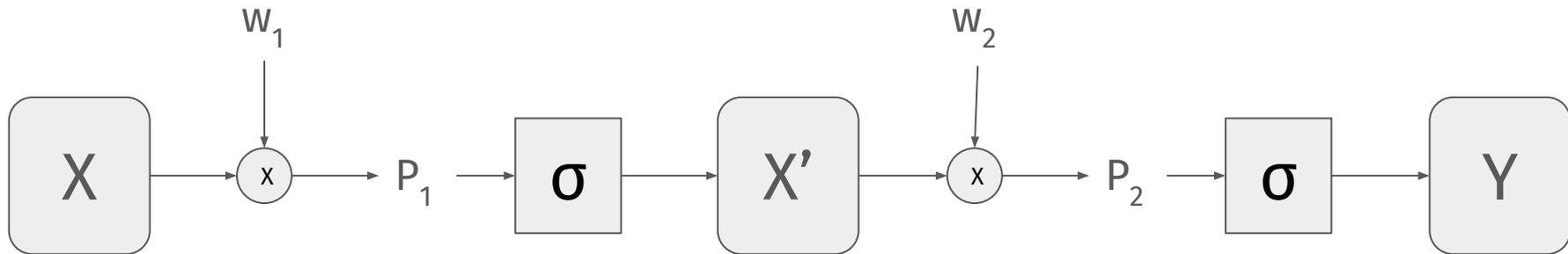
$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial Y} \frac{\partial Y}{\partial P_2} \frac{\partial P_2}{\partial w_2}$$

$$(1 - Y)Y$$




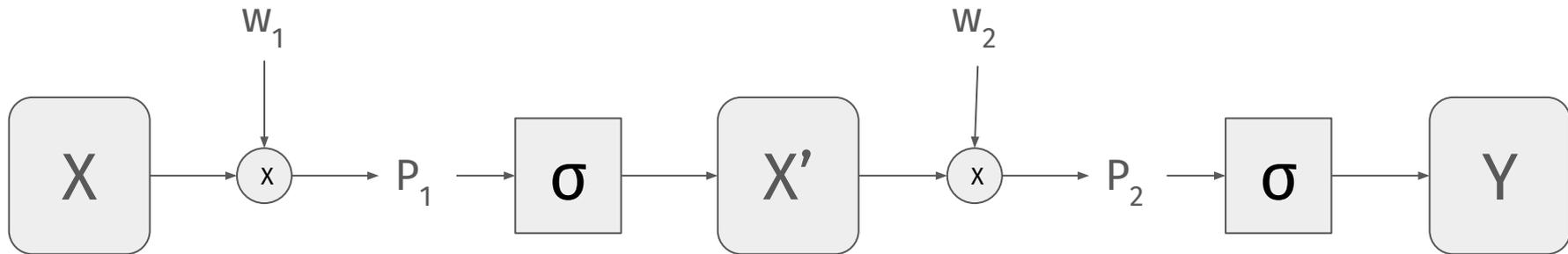
$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial Y} \frac{\partial Y}{\partial P_2} \frac{\partial P_2}{\partial w_2}$$

$X'$



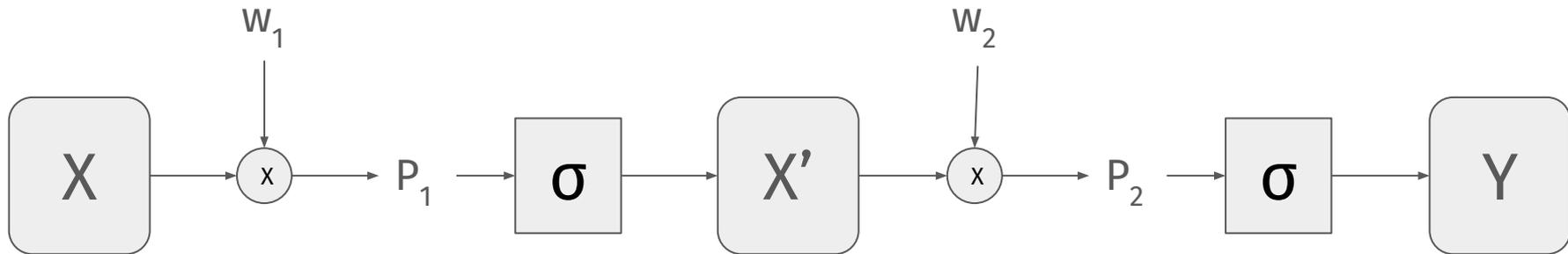
$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial Y} \frac{\partial Y}{\partial P_2} \frac{\partial P_2}{\partial w_2}$$

$$= (Y - Y_{true}) (1 - Y) YX'$$

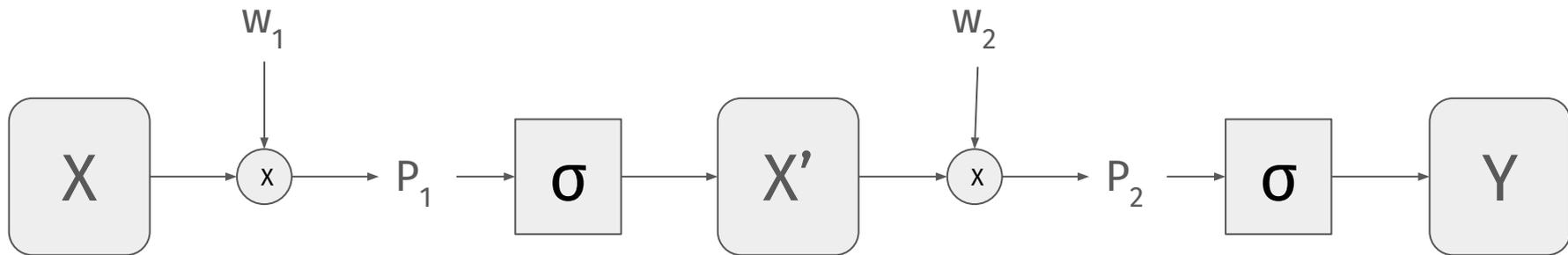


$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial Y} \frac{\partial Y}{\partial P_2} \frac{\partial P_2}{\partial w_2}$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial Y} \frac{\partial Y}{\partial P_2} \frac{\partial P_2}{\partial X'} \frac{\partial X'}{\partial P_1} \frac{\partial P_1}{\partial w_1}$$

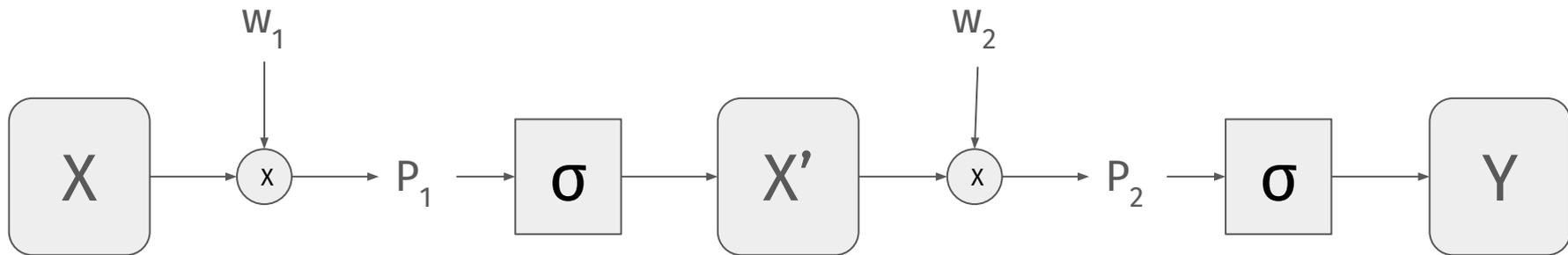


$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial Y} \frac{\partial Y}{\partial P_2} \frac{\partial P_2}{\partial X'} \frac{\partial X'}{\partial P_1} \frac{\partial P_1}{\partial w_1}$$



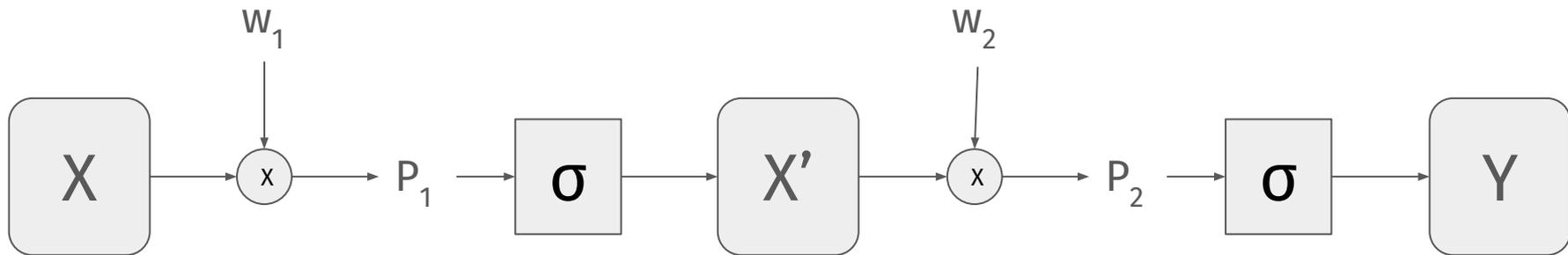
$$(1 - X') X'$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial Y} \frac{\partial Y}{\partial P_2} \frac{\partial P_2}{\partial X'} \frac{\partial X'}{\partial P_1} \frac{\partial P_1}{\partial w_1}$$



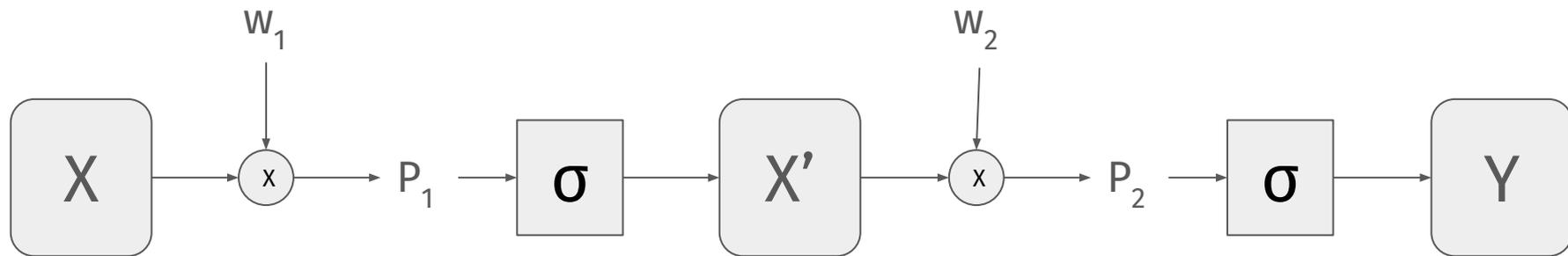
# X

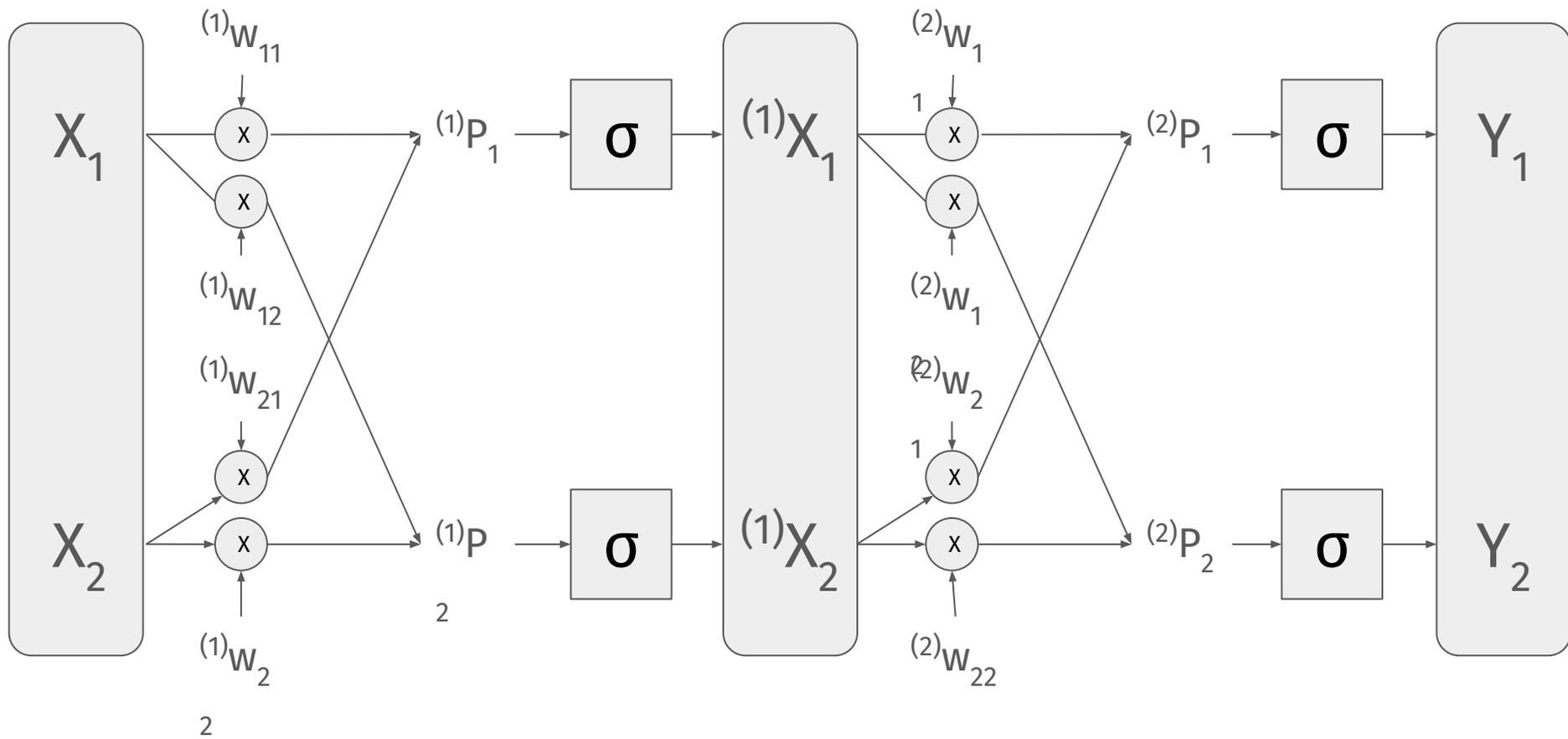
$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial Y} \frac{\partial Y}{\partial P_2} \frac{\partial P_2}{\partial X'} \frac{\partial X'}{\partial P_1} \frac{\partial P_1}{\partial w_1}$$



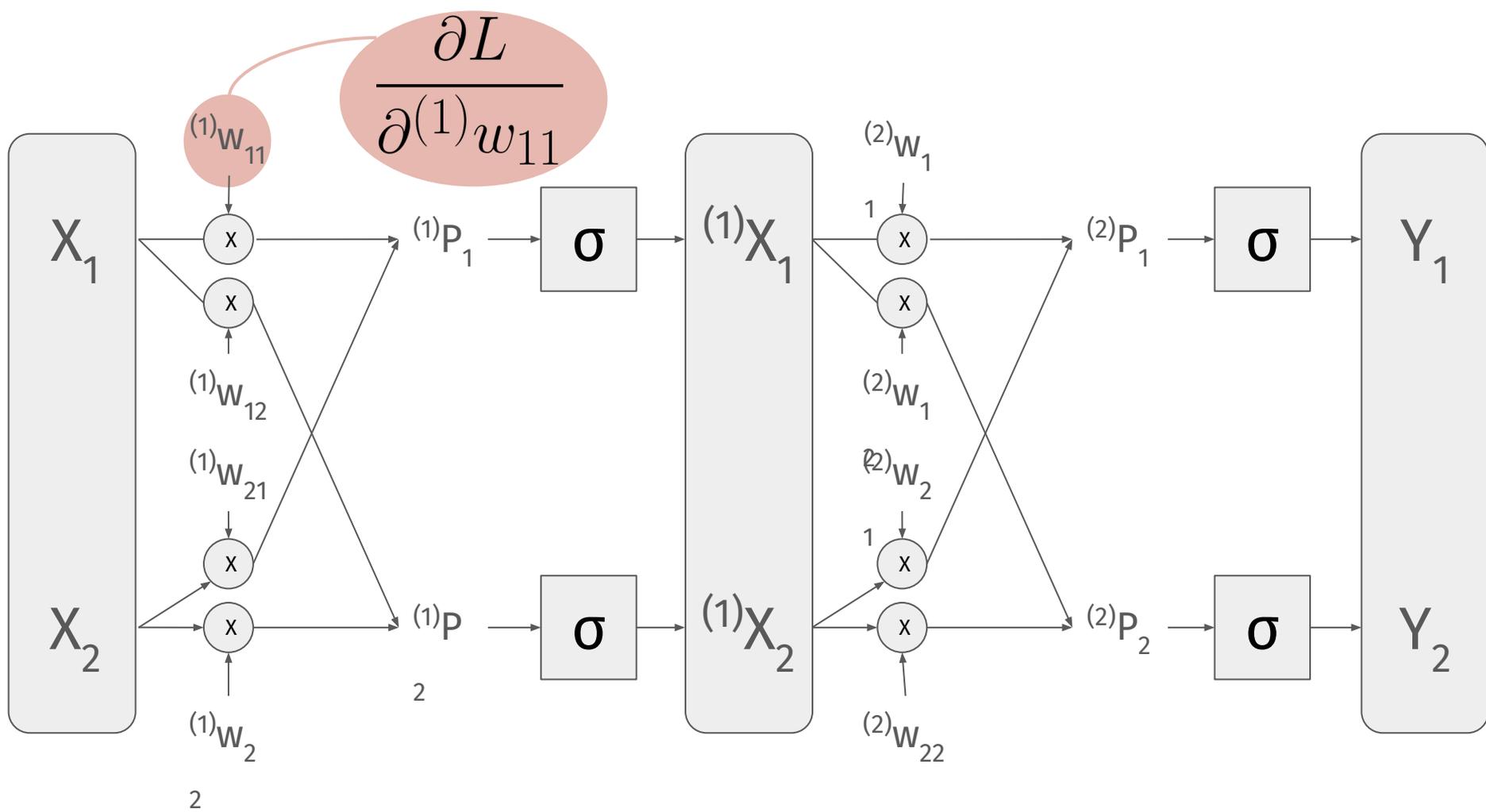
$$= \frac{\partial L}{\partial w_2} \frac{\partial w_2}{\partial X'} (1 - X') X' X$$

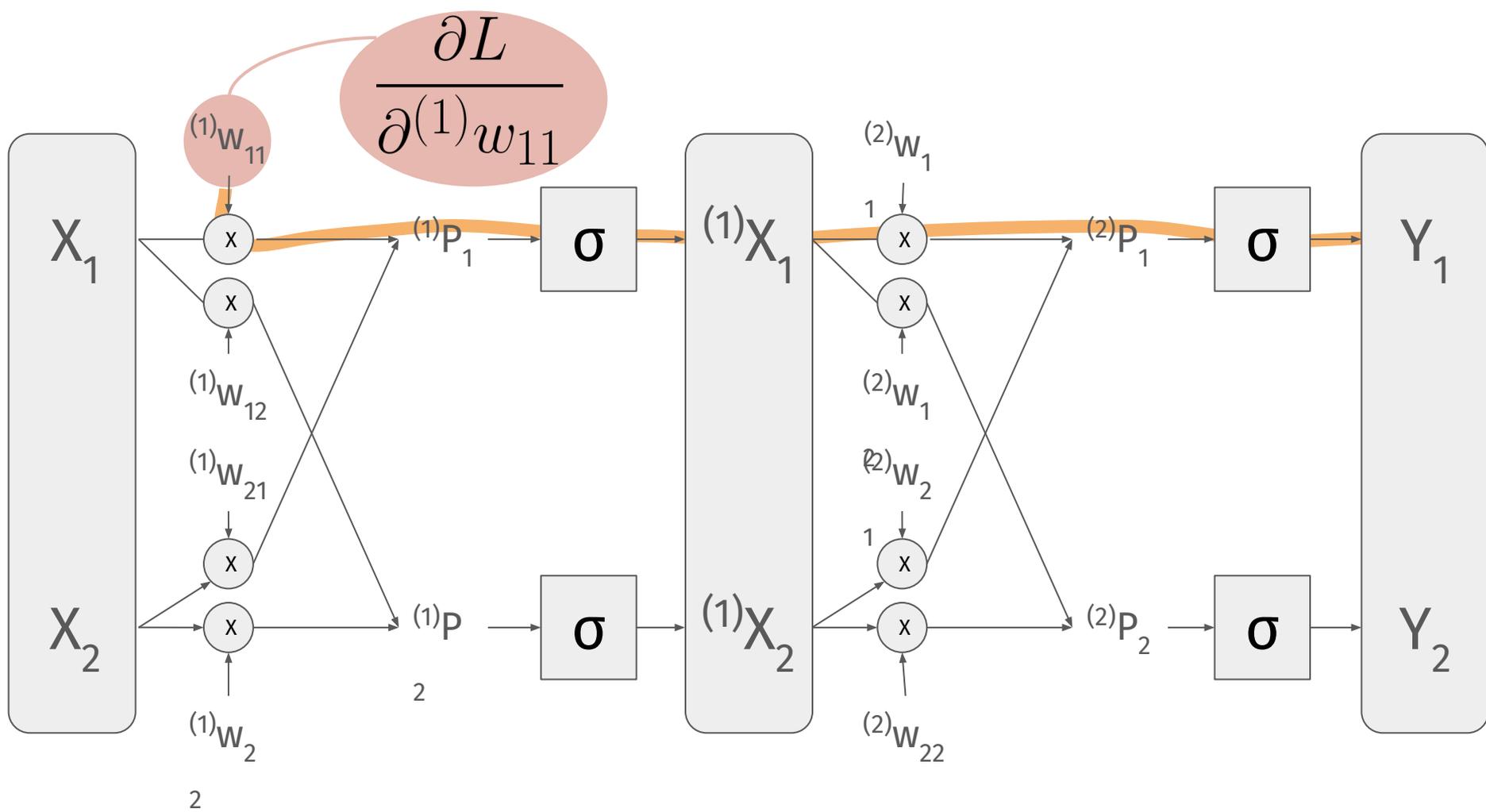
$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial Y} \frac{\partial Y}{\partial P_2} \frac{\partial P_2}{\partial X'} \frac{\partial X'}{\partial P_1} \frac{\partial P_1}{\partial w_1}$$

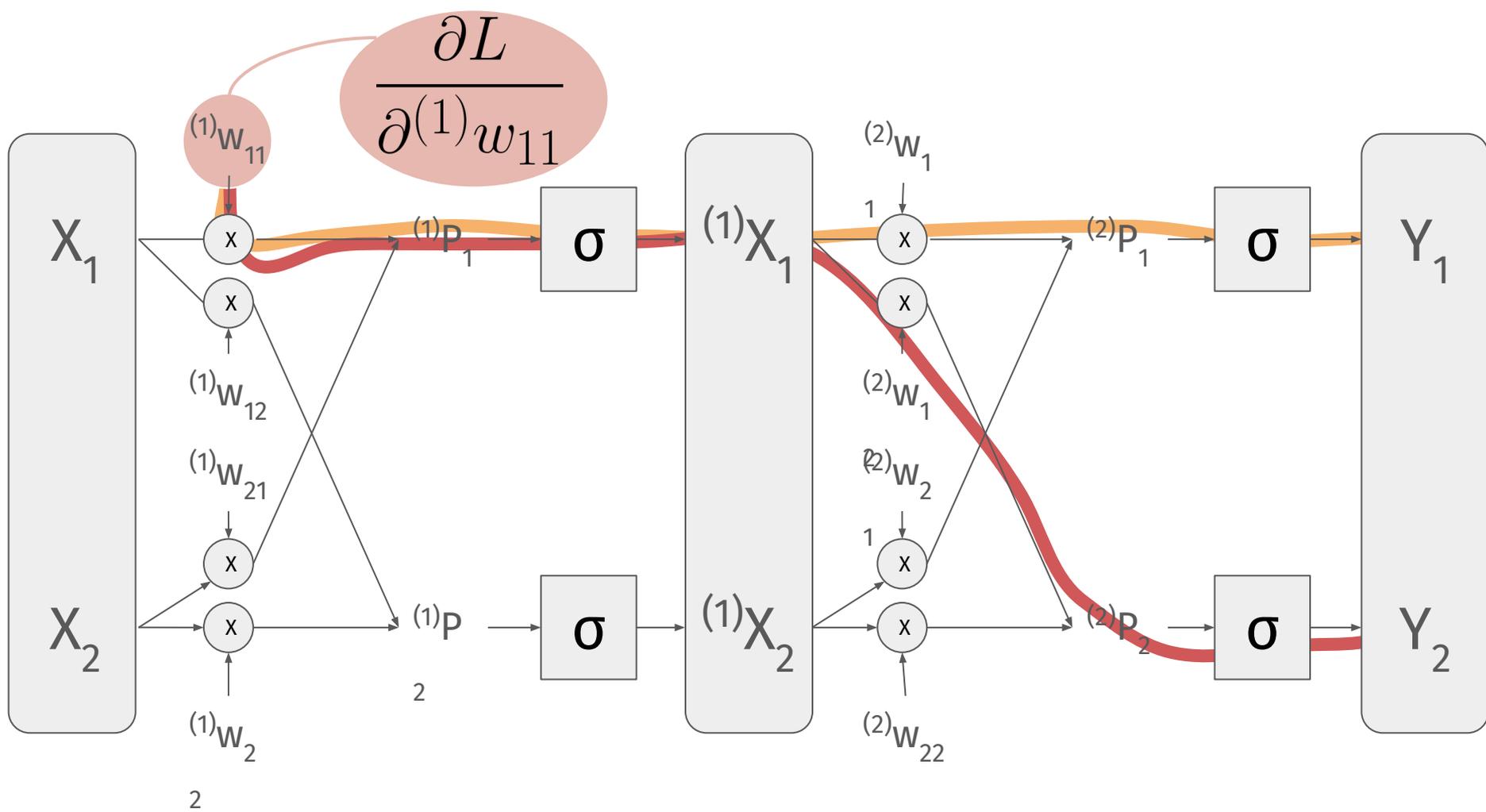




$$\mathbf{Y} = \sigma \left( {}^{(2)}\mathbf{W} \cdot \sigma \left( {}^{(1)}\mathbf{W} \cdot \mathbf{X} \right) \right)$$







$$\frac{\partial L}{\partial^{(1)}w_{11}} = \frac{\partial L}{\partial Y_1} \frac{\partial Y_1}{\partial P_1^{(2)}} \frac{\partial P_1^{(2)}}{\partial X_1^{(1)}} \frac{\partial X_1^{(1)}}{\partial P_1^{(1)}} \frac{\partial P_1^{(1)}}{\partial w_{11}^{(1)}}$$

$$\frac{\partial L}{\partial^{(1)}w_{11}} = \frac{\partial L}{\partial Y_1} \frac{\partial Y_1}{\partial P_1^{(2)}} \frac{\partial P_1^{(2)}}{\partial X_1^{(1)}} \frac{\partial X_1^{(1)}}{\partial P_1^{(1)}} \frac{\partial P_1^{(1)}}{\partial w_{11}^{(1)}}$$

$$+ \frac{\partial L}{\partial Y_2} \frac{\partial Y_2}{\partial P_2^{(2)}} \frac{\partial P_2^{(2)}}{\partial X_1^{(1)}} \frac{\partial X_1^{(1)}}{\partial P_1^{(1)}} \frac{\partial P_1^{(1)}}{\partial w_{11}^{(1)}}$$

$$\frac{\partial L}{\partial^{(1)}w_{11}} = \frac{\partial L}{\partial Y_1} \frac{\partial Y_1}{\partial P_1^{(2)}} \frac{\partial P_1^{(2)}}{\partial X_1^{(1)}} \frac{\partial X_1^{(1)}}{\partial P_1^{(1)}} \frac{\partial P_1^{(1)}}{\partial w_{11}^{(1)}} + \frac{\partial L}{\partial Y_2} \frac{\partial Y_2}{\partial P_2^{(2)}} \frac{\partial P_2^{(2)}}{\partial X_1^{(1)}} \frac{\partial X_1^{(1)}}{\partial P_1^{(1)}} \frac{\partial P_1^{(1)}}{\partial w_{11}^{(1)}}$$

# The importance of an activation function and tensor algebra



<https://l.infn.it/ex7>



[https://commons.wikimedia.org/wiki/File:Sinterklaas\\_zwarte\\_piet.jpg](https://commons.wikimedia.org/wiki/File:Sinterklaas_zwarte_piet.jpg)

Prof. Silva Bortolussi, Dr. Valerio Vercesi, Dr. Ricardo Ramos, Dr. Setareh Fatemi, PhD Barbara Marcaccio

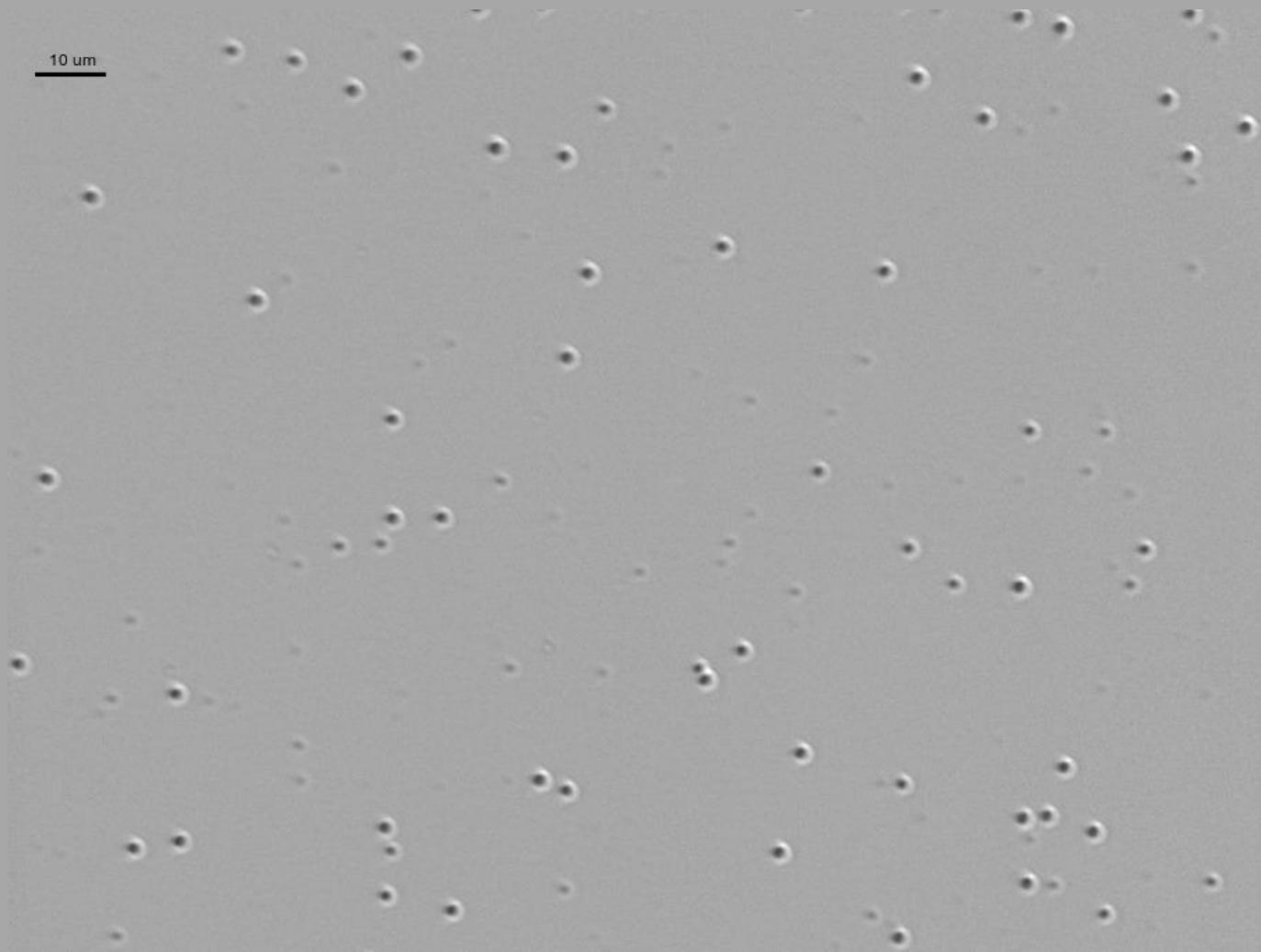
Boron Neutron Capture Therapy

&

AI in Fisica Medica

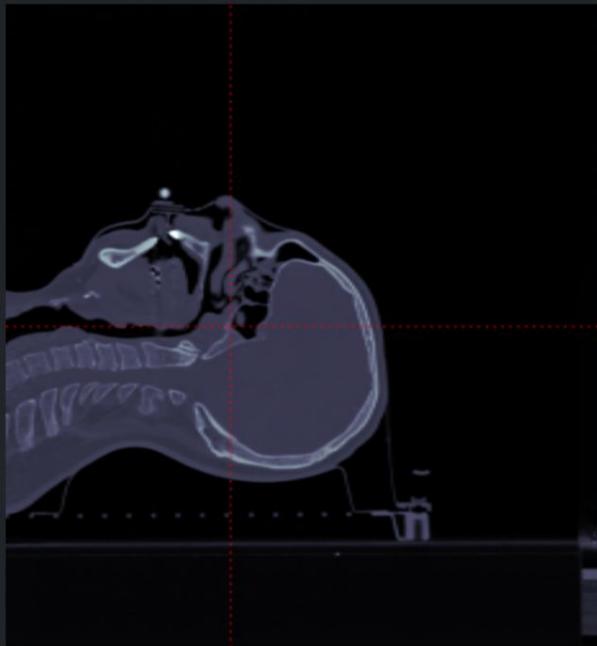
Prof. Alessandro Lascialfari, Dr. Francesca Brero,  
PhD Agnese Robustelli Test

10  $\mu$ m

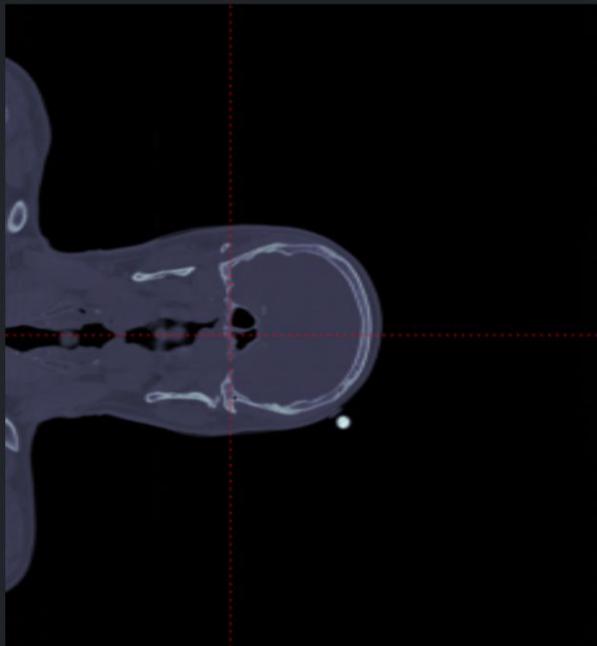




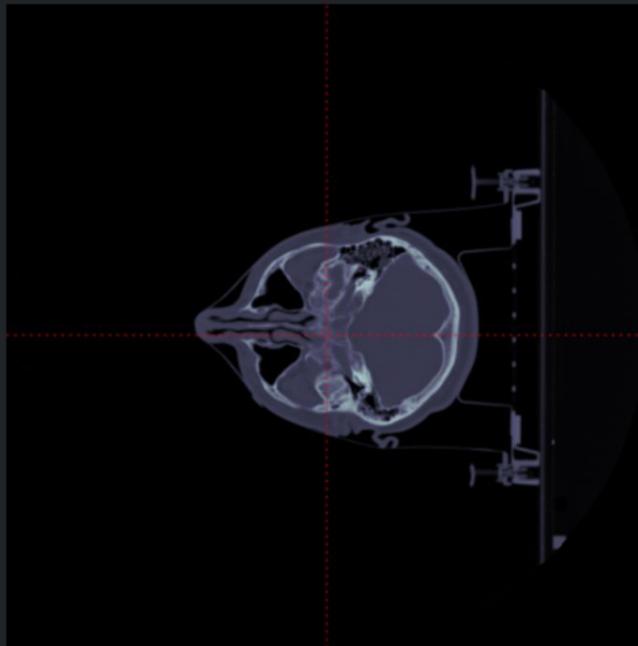
File "ct.nii.gz" uploaded and opened successfully!



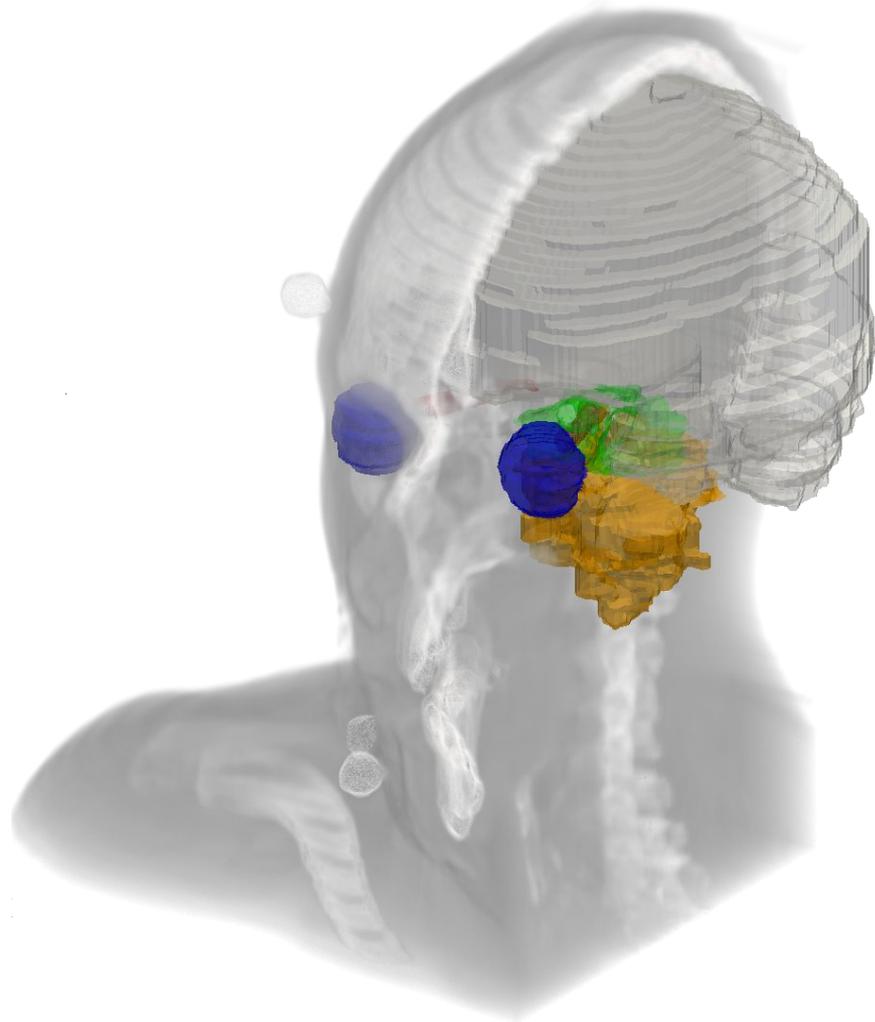
Y - 262

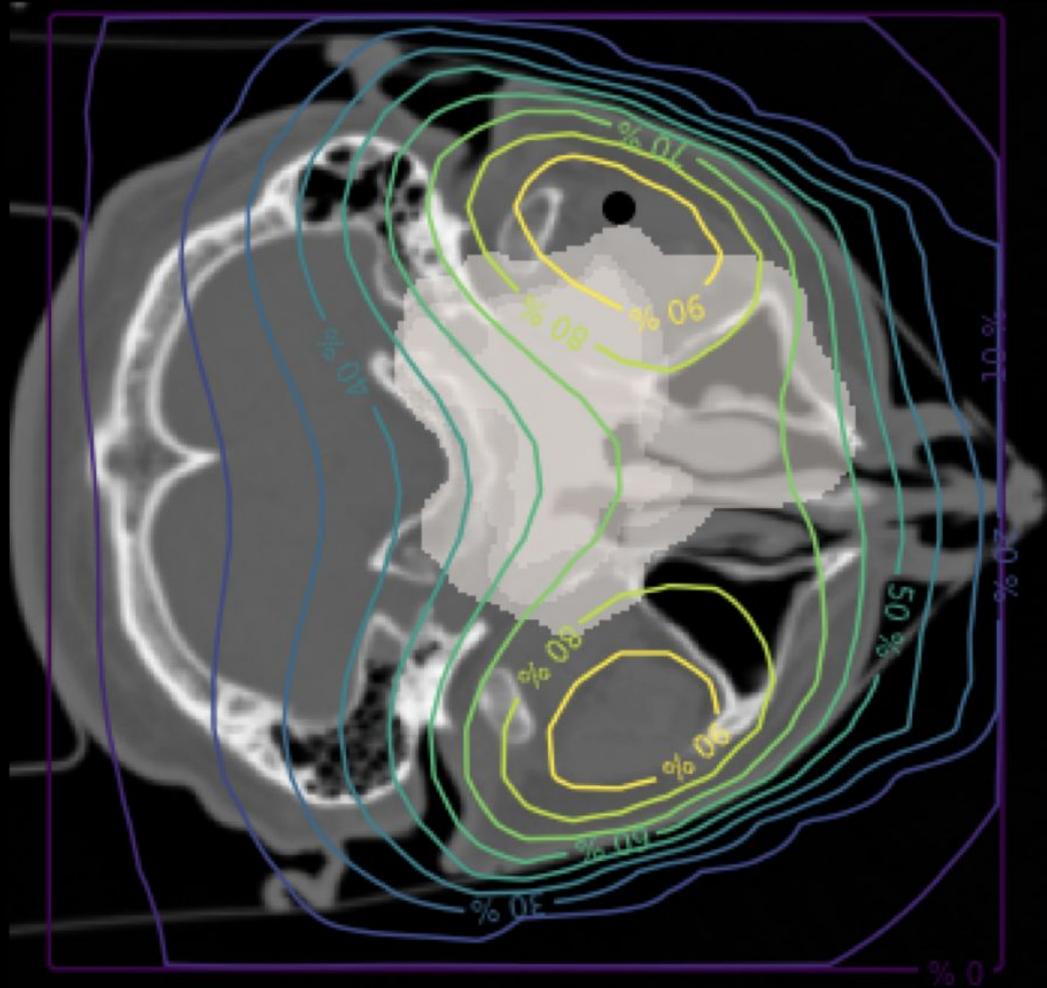


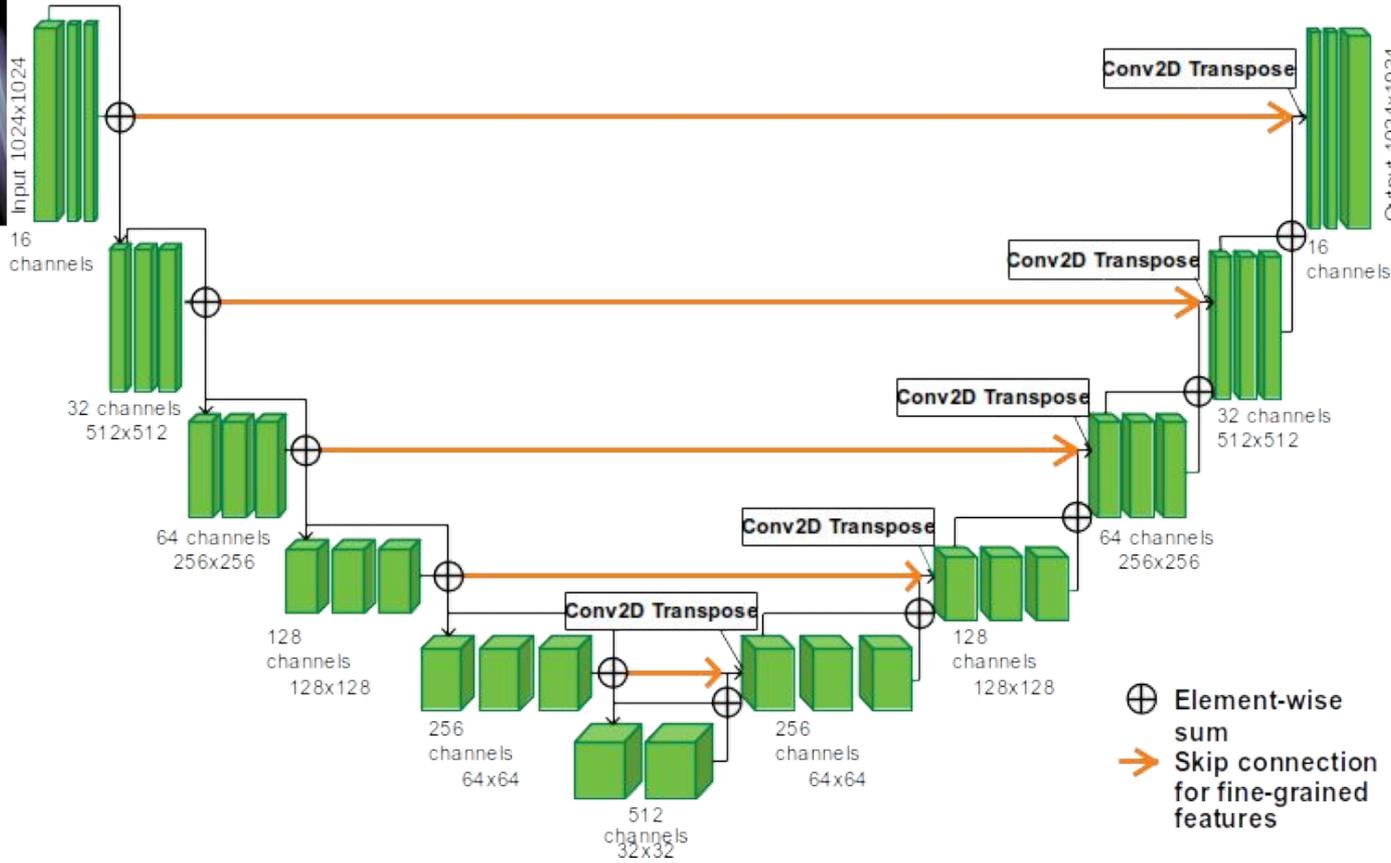
Y - 256



Z - 87







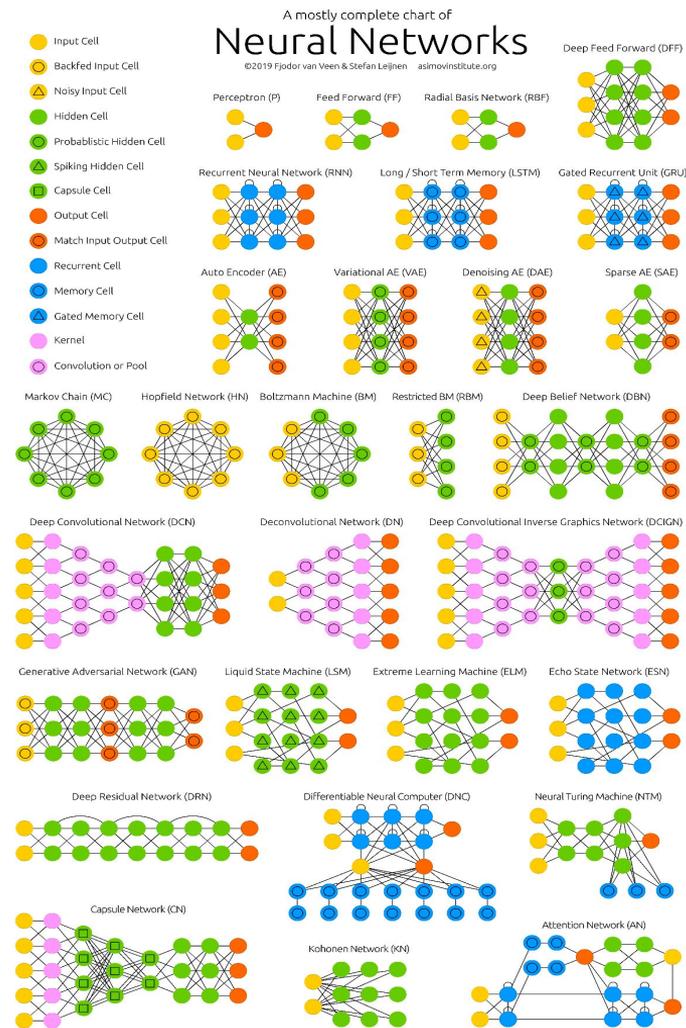


**e ora qualcosa  
di più fisico**

# L'intelligenza artificiale e' un coltellino svizzero



Molte topologie di reti sviluppate per risolvere problemi diversi: ognuna è uno strumento diverso



# Quali strumenti? Come usarli?

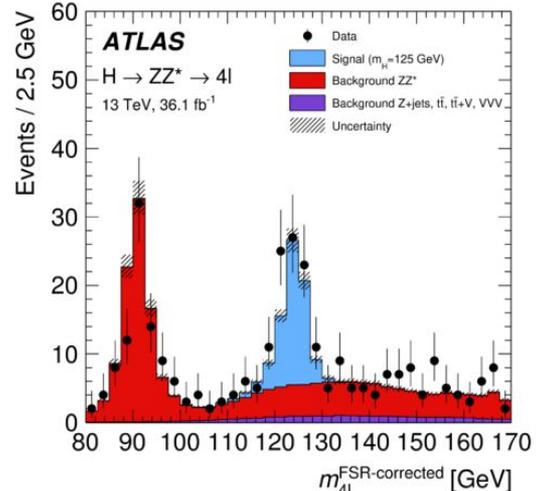
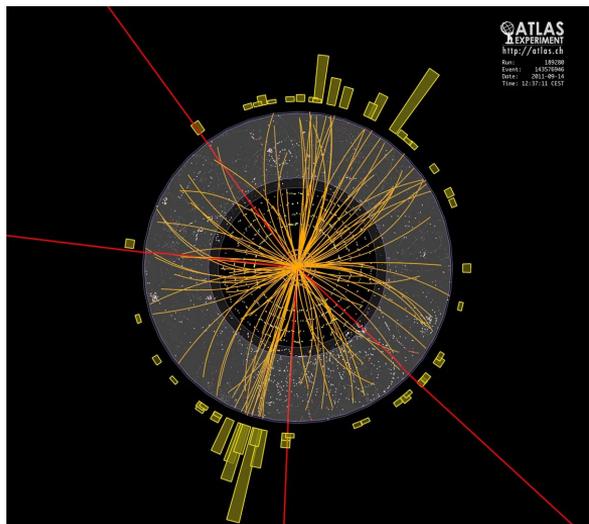
Gli strumenti vanno usati nel modo corretto:

- Definire per ogni problema lo strumento adatto
- Preparare i dati in modo da ottimizzare la prestazione dello strumento
- Se è il caso, adattare lo strumento alle nostre necessità

Nel seguito: un paio di esempi di questo approccio all'uso di AI in fisica delle particelle



# Esperimenti in fisica delle particelle



Si fanno collidere  
due particelle



Si prende una "foto" dei  
prodotti della collisione



Si elabora la foto  
e si scopre una  
nuova particella

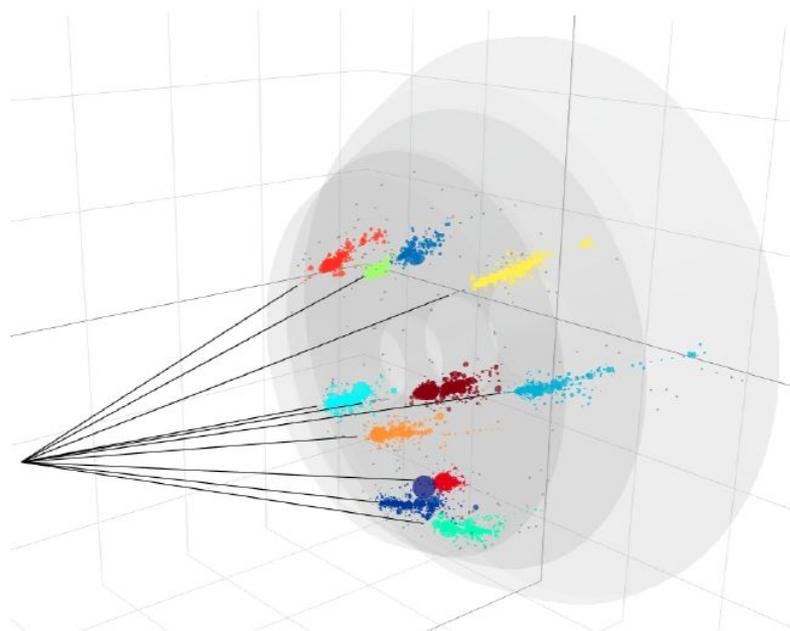
# Eventi di fisica particelle verso immagini convenzionali

Differenze principali:

**Dimensionalità:** le nostre immagini sono 3D (coordinate spaziali) o 4D (spaziali+temporali)

**Big Data:** migliaia di pixel in ogni dimensione: da 10 a 1M più grande di un'immagine 2D (memoria, spazio disco...)

**Sparsità:** La maggior parte dell'azione concentrata in piccole regioni del rivelatore



Possiamo usare le idee sviluppate per applicazioni non scientifiche ma dobbiamo metterci le mani per farle funzionare per i nostri obiettivi!

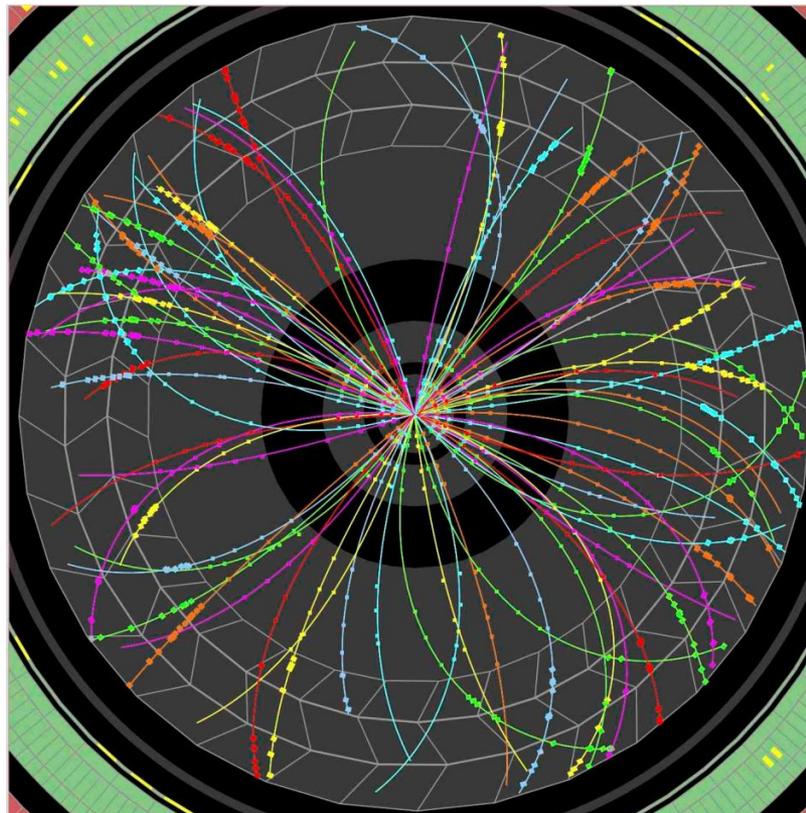
# Problema di analisi di immagine: Ricostruzione di tracce

Centinaia di migliaia di punti spaziali nel rivelatore (“hit”)

Vogliamo raggrupparli in ‘tracce’, gruppi di punti che si allineano su un arco di cerchio

**Approcci possibili:**

Hit come coordinate di un’immagine: reti convoluzionali



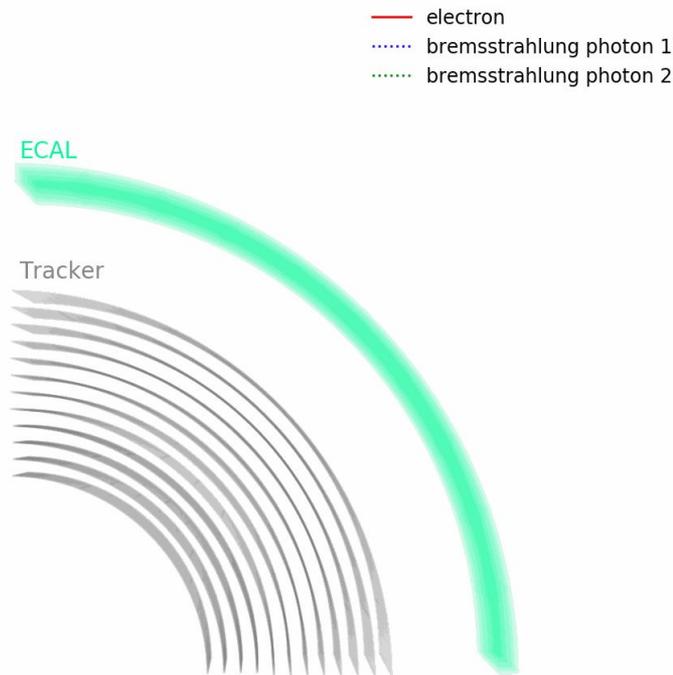
# Problema di analisi di immagine: Ricostruzione di tracce

Centinaia di migliaia di punti spaziali nel rivelatore (“hit”)

Vogliamo raggrupparli in ‘tracce’, gruppi di punti che si allineano su un arco di cerchio

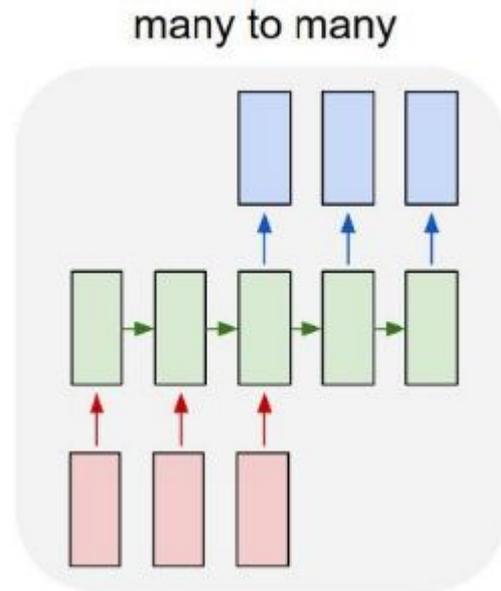
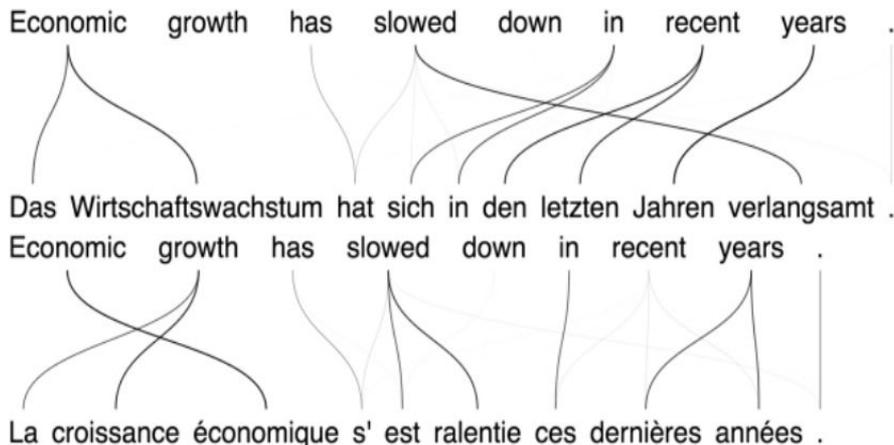
**Approcci possibili:**

Ogni strato del rivelatore come fase successive dell’evoluzione della traccia influenzata da fasi precedenti: reti ricorrenti



# Uso di reti ricorrenti: un esempio

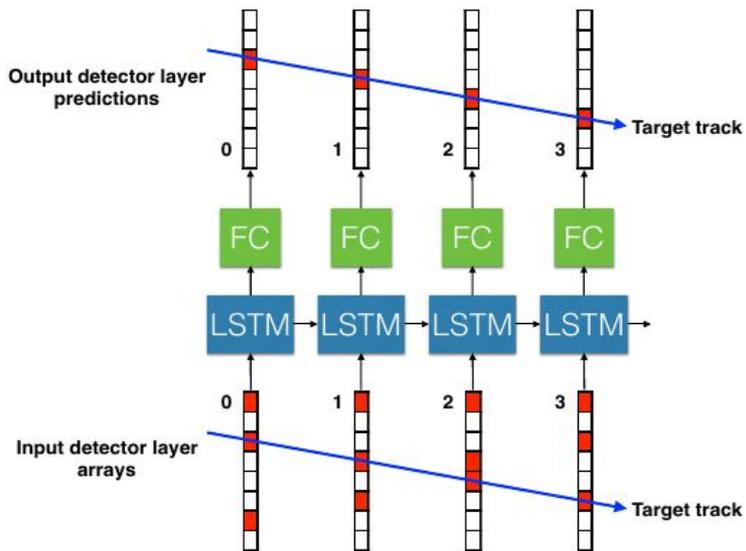
Traduzione automatica: il significato di ogni parola dipende dalla posizione di tutte le altre parole nella frase



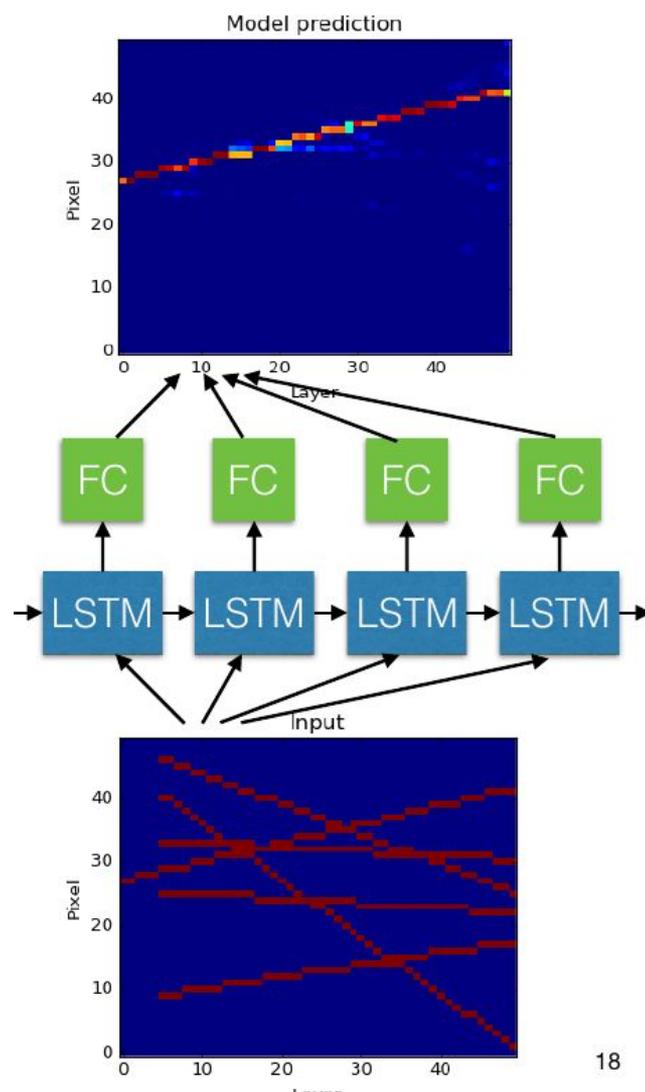
Ingressi organizzati in sequenze  
A ogni elemento associato nodo  
**Addestramento di ogni nodo influenzato da stato dei nodi precedenti**

# Approcci AI al tracciamento: traccia come serie ordinata

Ogni strato del rivelatore e' un vettore  
In ingresso tutti gli 'hit'  
In uscita solo gli hit appartenenti a una  
stessa traccia



HEP.TrkX project

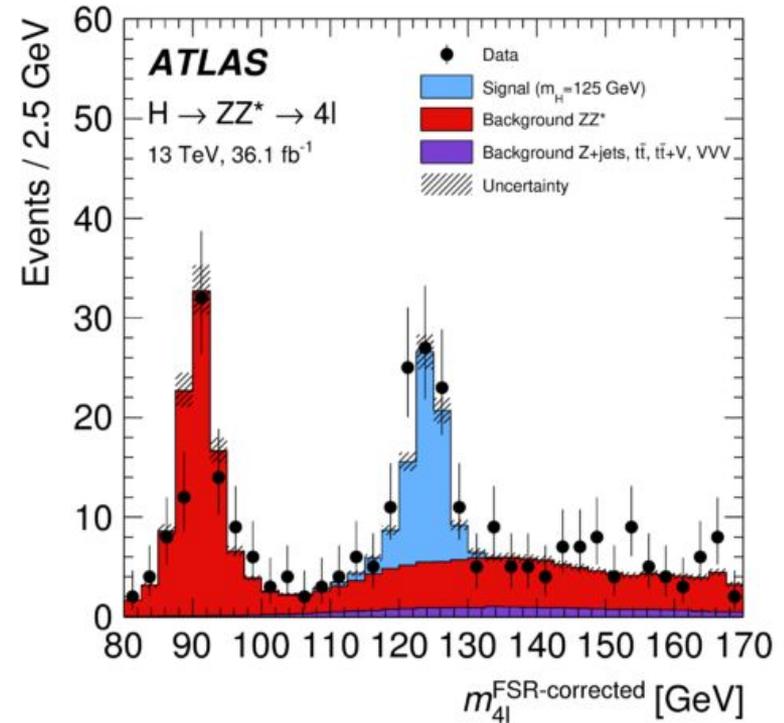


# Problema: la ricerca di nuova fisica

Sappiamo come si comporta la fisica 'nota' nel rivelatore

- **Simuliamo** eventi di fisica nota con il metodo Monte Carlo (Background rosso nella figura)
- Confrontiamo con i dati (pallini neri)
- Se sono diversi vuol dire che c'è della fisica nuova (il segnale di Higgs in azzurro)

Caso "facile": sappiamo cosa stiamo cercando **ricerca supervised**



# Problema: La ricerca di nuova fisica

- La fisica “nuova” è per definizione sconosciuta
- Molte proposte teoriche:
  - Qual è quella vera?
  - E se nessuno ha pensato a quella vera?

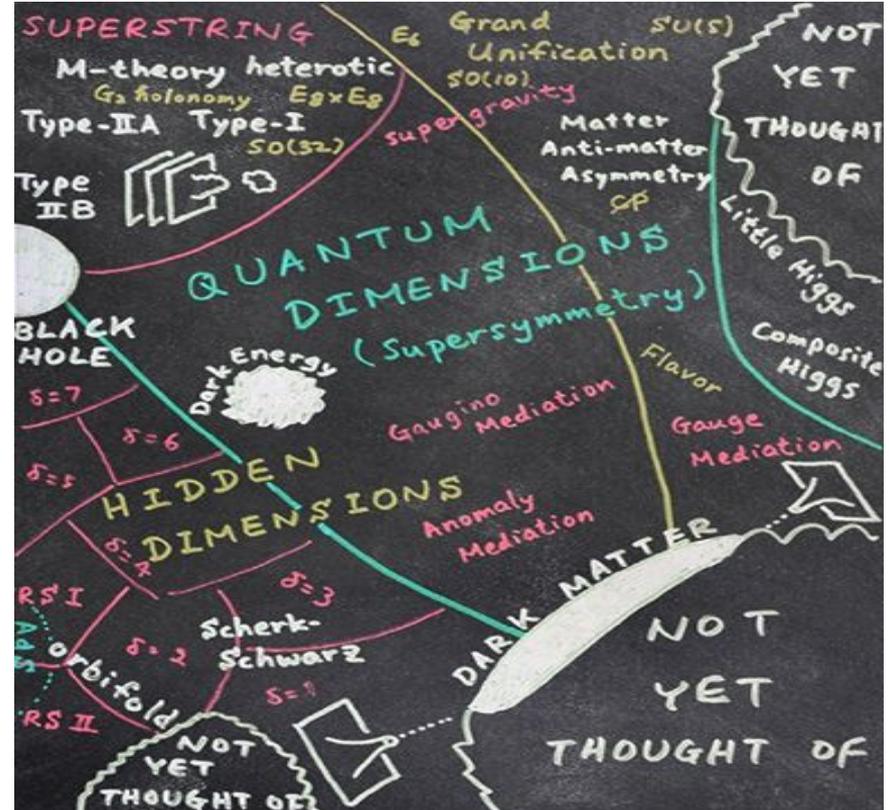


**Ricerca unsupervised**

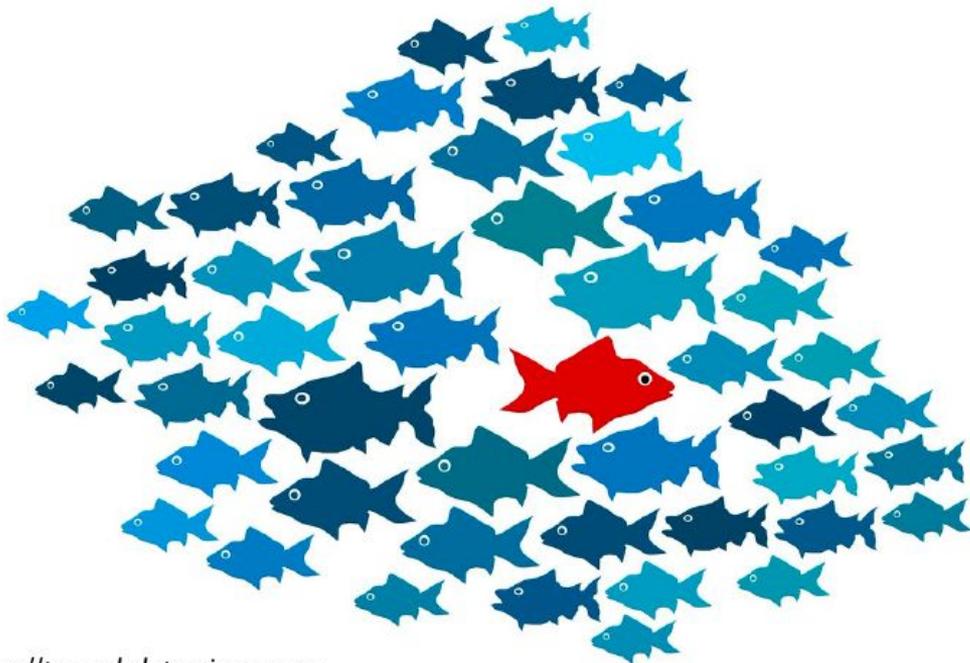
- Sappiamo che eventi si producono in assenza di nuova fisica (Standard Model - SM)
- Vogliamo che tra tutti gli eventi trovi quelli “anomali” rispetto a SM



**Ricerca di anomalie**



# Ricerca di anomalie



<https://towardsdatascience.com>

Insieme di oggetti: qual è quello  
“strano”?

**In vita reale:** Controllo carte di  
credito: se un’operazione è  
“anomala” rispetto alle nostre  
abitudini viene bloccata



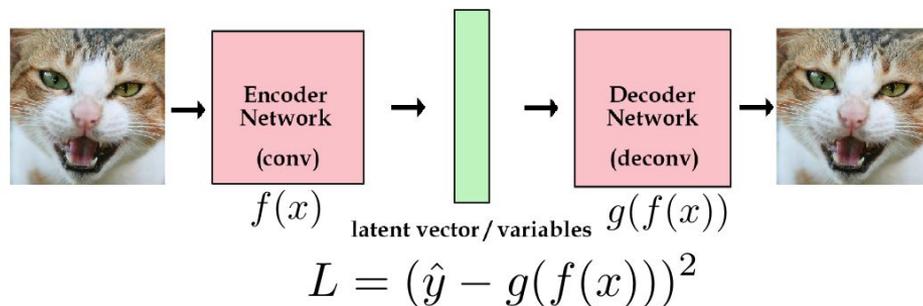
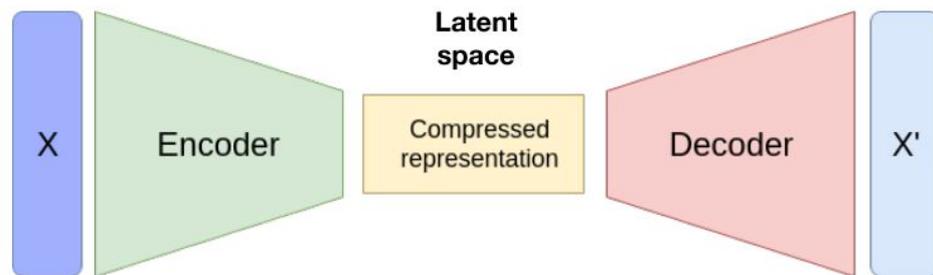
# Esempio di algoritmo per anomalie: autoencoder

Rete 1 (encoder) codifica immagine in spazio latente a dimensionalità inferiore:

Rete 2 (decoder): dall'immagine latente costruisce nuova immagine

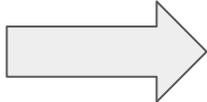
Si addestra la rete minimizzando differenza tra immagine in entrata e quella in uscita (**Loss  $L$** )

Se si addestra l'algo con foto di gatti, una foto di cane in ingresso darà  **$L$**  molto più grande che una nuova foto di un gatto

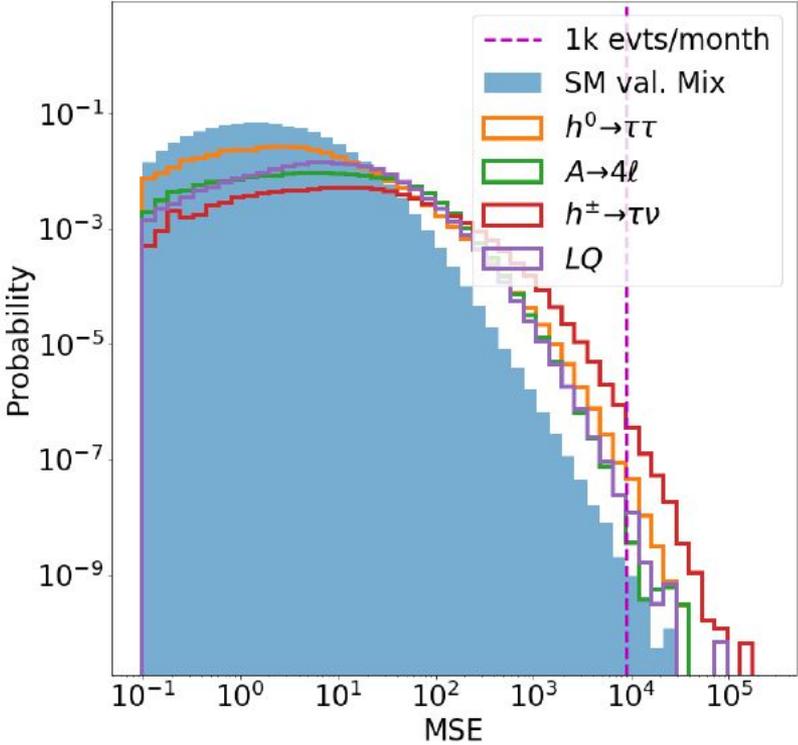


# Esempio di ricerca anomalia in HEP

- Codifica ogni evento in 21 variabili
- Addestra la rete su cocktail di eventi simulati SM
- Calcola distribuzione di funzione di "Loss" per eventi SM
- Manda in ingresso eventi di nuova fisica (NP) e calcola funzione di "Loss"
- Distribuzione "Loss" per gli eventi NP ha un valore piu' alto che per eventi SM
- In esperimento reale: si predice a partire dalla simulazione il numero di eventi aspettati con  $Loss > X$ . C'è un eccesso di eventi in questa regione?



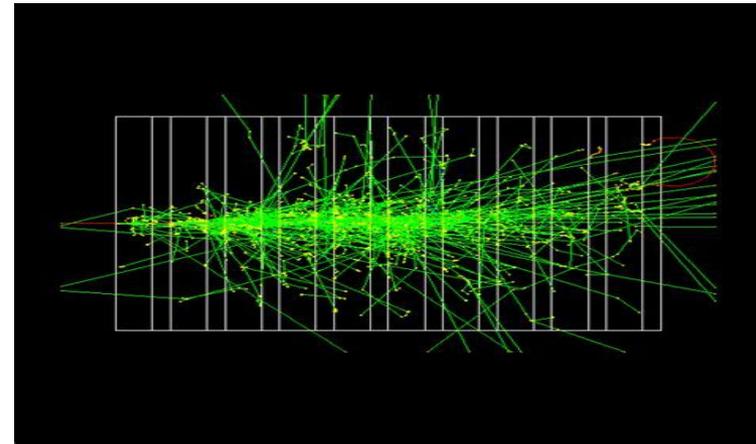
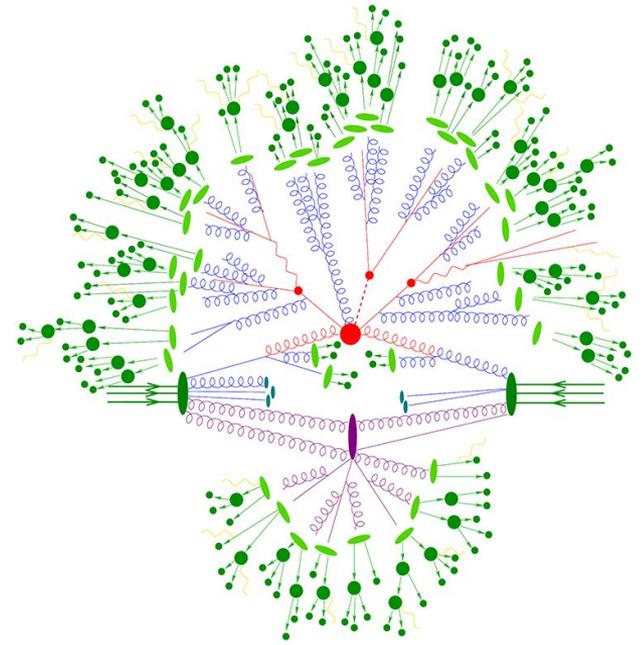
**Scoperta!**



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

# Problema: simulazione in HEP

- Ogni esperimento produce  $\sim 10^{10}$  eventi all'anno
- Per analizzarli serve ~stesso numero di eventi simulati
- Simulazione: emulare per ogni evento l'interazione di migliaia di particelle in passi di frazioni di mm nel rivelatore
- **Incredibilmente lento e costoso!**
- **Possiamo usare AI per sveltire il processo?:** Si da in pasto a un algoritmo un campione di eventi simulati e gli si chiede di produrre eventi diversi con le stesse caratteristiche



# AI generativa

ChatGPT



GitHub Copilot

AI genera immagini/testi che non esistevano prima, ma che sono “plausibili” come immagini reali: deepfake

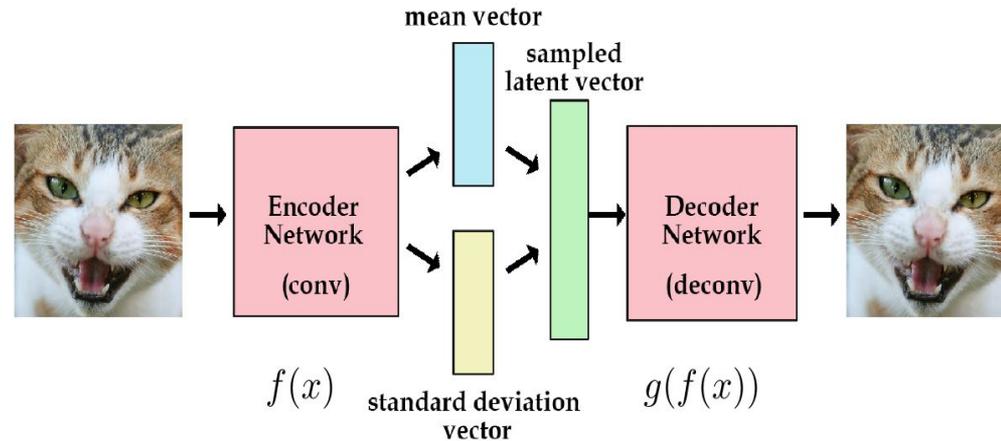


# Esempio di AI generativa: Variational autoencoder

Funziona come autoencoder, ma invece di un vettore, codifica per ogni 'feature' media e deviazione standard

Il decoder e' allenato a riprodurre l'immagine iniziale da immagini campionate dalla distribuzione gaussiana

Le immagini in uscita non esistevano prima, ma la rete (e noi!) le riconosce come omogenee a quelle di addestramento

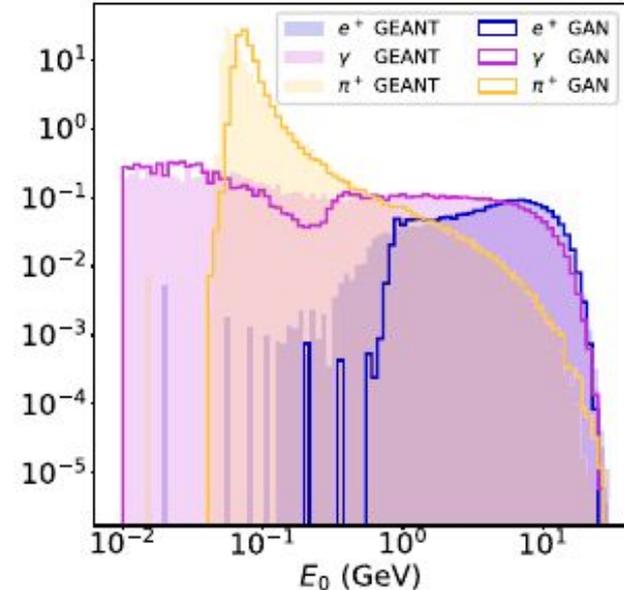
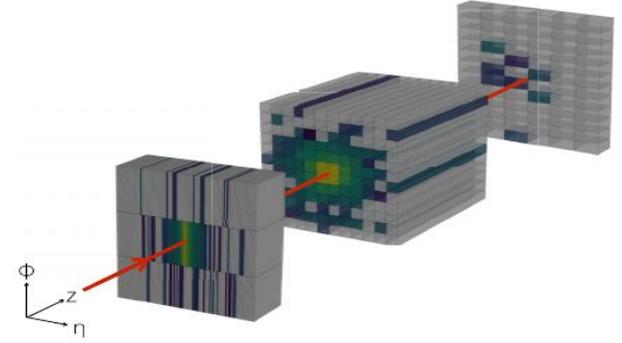


# Esempio: simulazione deposito energia in rivelatore

Si modella il deposito di energia nel rivelatore come una matrice 3-dimensionale

Si utilizza una rete generativa avversariale per stabilizzare la simulazione

Buona riproduzione del deposito di energia per diverse particelle





[giacomo.polesello@pv.infn.it](mailto:giacomo.polesello@pv.infn.it)

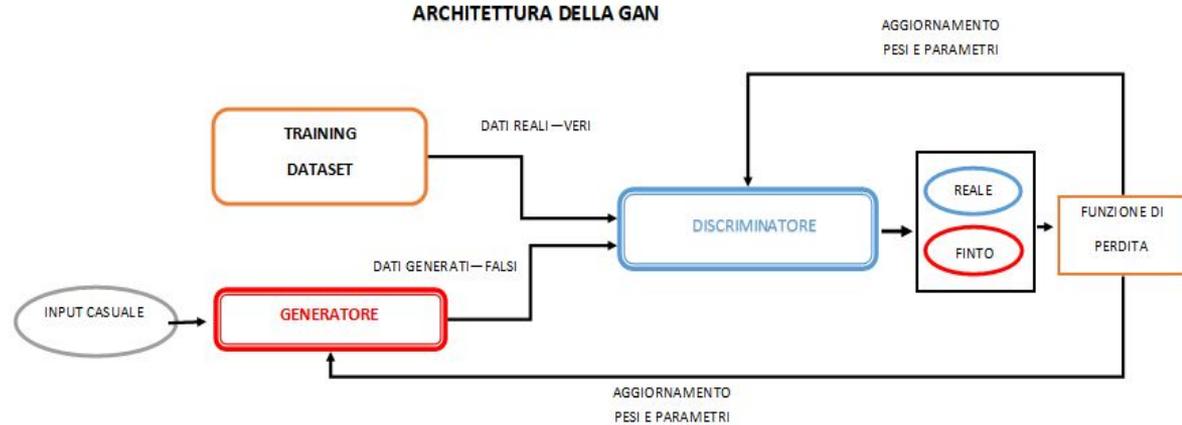
[ian.postuma@pv.infn.it](mailto:ian.postuma@pv.infn.it)

Backup

# Rete generativa avversariale

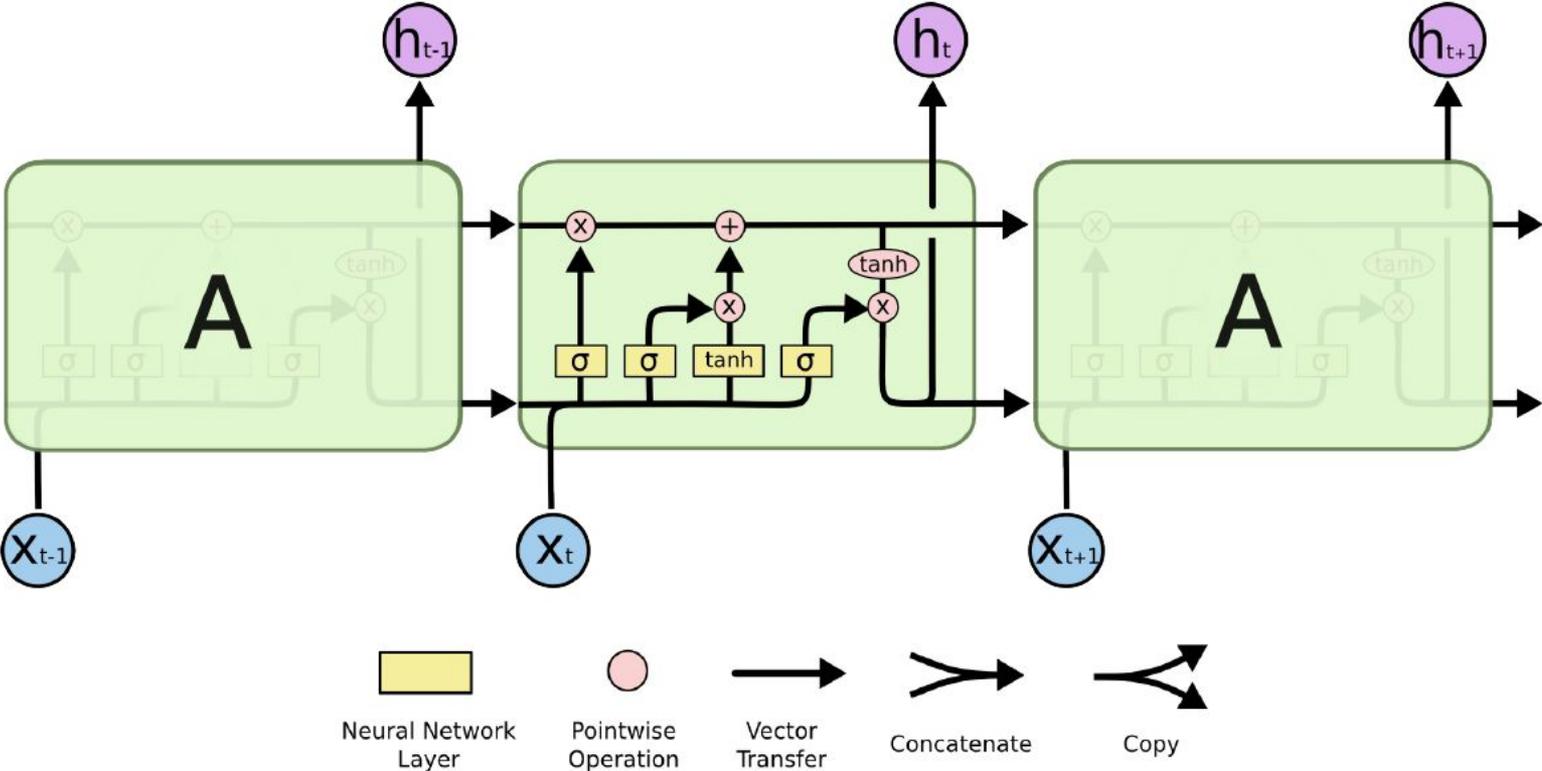
Si pongono in competizione due reti:

- Il generatore: scopo è di produrre eventi che imitino quelli di addestramento
- Il discriminatore: scopo è di trovare quali eventi sono finti, prodotti dal generatore

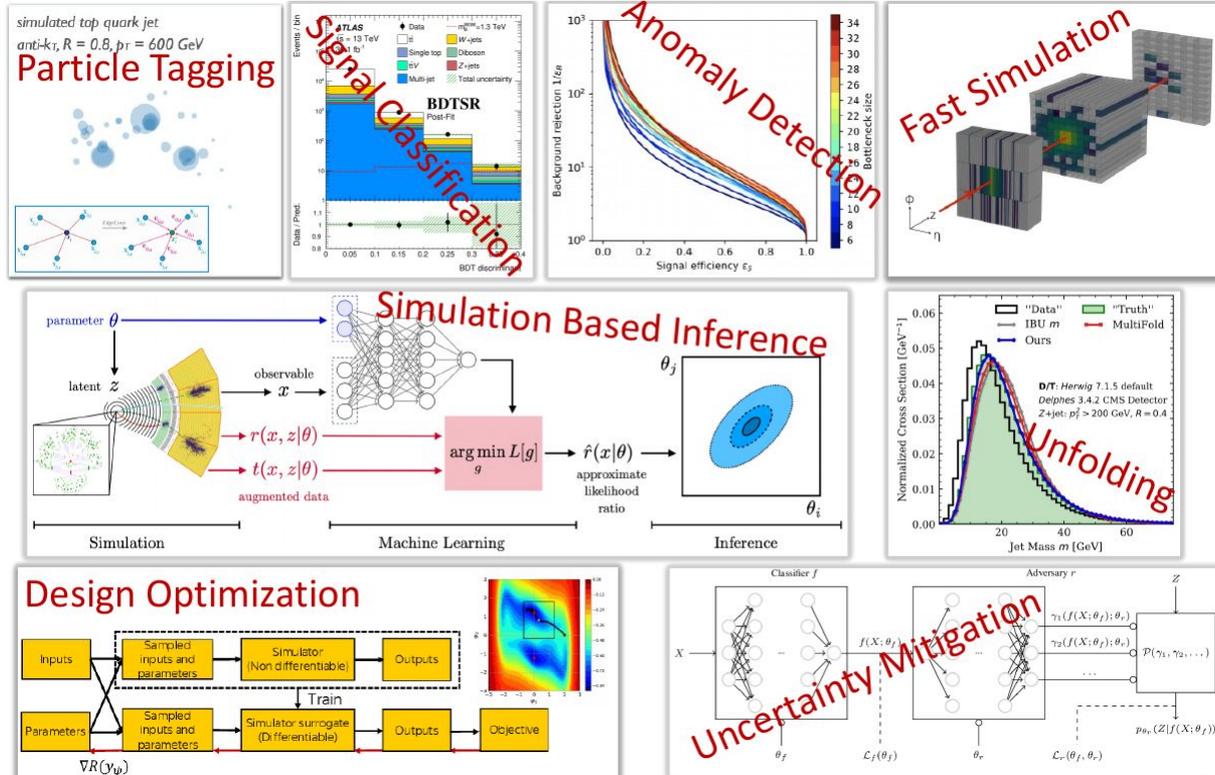


Tipicamente più facile da addestrare di un autoencoder variazionale

# Long-short term memory recurrent cell



# Usi dell'intelligenza artificiale in HEP



Negli ultimi anni sempre più problemi in HEP affrontati con strumenti AI