



# **Classificazione dei bosoni elettrodeboli con una rete neurale al Large Hadron Collider**

**7 Novembre 2024**

Candidato:  
**Jacopo Lancione**

Relatore:  
Prof. **Emanuele Roberto Nocera**

- Introduzione
  - LHC
  - Machine Learning
  
- Dataset
  - Decadimenti dei bosoni
  - Preprocessing
  
- Reti Neurali
  - Architettura e principi
  - Risultati dell'allenamento

L'anello di accelerazione + grande del cern (foto del logo), parliamo del CMS da cui arrivano i miei dati, in cui si producono particelle in abbondanza,

(foto del cern e dl cms)

Del CMS mi interessa giusto introdurre il fatto che ci siano dei calorimetri perché alcuni loro parametri sono tra le feature

Si produce un enorme mole di dati e per trattarli si utilizzano anche tecniche di machine learning (e così passo alla prox slide)

E posso accennare molto rapidamente al Nobel di quest'anno

# Machine Learning



In questa slide devo far passare il concetto di labeled e unlabeled data (posso anche metterle come item) e la separazione del dataset in train test e validation (per il caso supervised)

## Unsupervised Learning

- Clustering
- Riduzione dimensionale

## Supervised Learning

- Classificazione
- Riconoscimento di immagini e testo

Qua dicendo che il mio progetto ruota attorno ad un problema di classificazione passo alla prox slide

# Machine Learning



In questa slide devo far passare il concetto di labeled e unlabeled data (posso anche metterle come item) e la separazione del dataset in train test e validation (per il caso supervised)

## Unsupervised Learning

- Clustering
- Riduzione dimensionale

## Supervised Learning

- **CLASSIFICAZIONE**
- Riconoscimento di immagini e testo

Qua dicendo che il mio progetto ruota attorno ad un problema di classificazione passo alla prox slide

Che sia chiaro dove si colloca il mio progetto: affrontare un problema di classificazione binaria (logistic regression), nell'ambito della Fisica delle alte energie

Immagine classica del modello std e 1 di 1 rete neurale giusto per dire rapidamente il Cosa e il Come

il mio obiettivo: allenare 1 rete che distingua al meglio tra i 2 canali di decadimento



# Il Dataset

Diciamo subito qlche dettaglio in + sul dataset (e magari mettiamolo anche a fondo slide, il riferimento a dove ho scaricato i dati)  
qua posso mettere i diagrammi di Feynman dei decadimenti, giusto per mettere qualcosa sotto gli occhi al pubblico  
Elencare le features ie la cinematica di interesse e anche le variabili lasciate da parte in riferimento al rivelatore



Racconto di come ho trattato i dati: la storia del  $\chi^2$  (vogliamo dirla? nn ne conosco i dettagli purtroppo), la qstione dgli outlier  
Qua mostriamo sicuramente i pairplot che sono la cosa più indicativa, magari anche i boxplot? -> in questa maniera escono + slides (e qua posso sprecarmi con il logaritmo e lo questione del'approccio scartato con le sigma)  
Raccontiamo la storia di correlazioni evidenti che permetterebbero una facile classificazione, nel caso + semplice attraverso 1 appl lineare  
I concetti da far passare sono 2: è meglio è avere 1 dataset uniforme, quindi scaliamo tutto e ci sbarazziamo degli outlier, evitare di introdurre ridondanze (ie guardare in faccia i dati con pairplot)



# Reti Neurali

Qua bisogna introdurre i parametri su cui ho agito: numero di layer, nodi, attivazione, algoritmo di ottimizzazione (questo nella slide successiva)

E devo spiegare come funziona la questione dei parametri (pesi e bias) e dove si introduce la non linearità (attivazione)

Questa slide la organizzerei come un elenco .ntato a sx e una bella immagine con cui io riesca a spiegare tutto



L'idea di 1 loss function da minimizzare (come se fosse un'energia), questa non è semplice da spiegare visivamente, di questa metterei proprio la formula così la commento un attimo

Algoritmo (sarebbe carino accennare al learning rate e al momento e all'adattività degli algoritmi)

Immagini di allenamenti significativi, magari anche in cui si veda l'overfitting

E qua ci va una carrellata di rock curves che può tranquillamente occupare + slides, quali voglio scegliere come significative? Con algoritmi diversi e mostrando bene il test point

- Le reti neurali si prestano molto bene a compiti di particle identification
- Sono degli strumenti molto flessibili e quindi il mio progetto è facilmente generalizzabile ad altre necessità/misure
- Ho identificato una classe di modelli equivalenti

studi futuri: provare a combinare i dataset e allenare una rete su quelli per distinguere i bosoni uno dall'altro

- Le reti neurali si prestano molto bene a compiti di particle identification
- Sono degli strumenti molto flessibili e quindi il mio progetto è facilmente generalizzabile ad altre necessità/misure
- Ho identificato una classe di modelli equivalenti

studi futuri: provare a combinare i dataset e allenare una rete su quelli per distinguere i bosoni uno dall'altro

**Grazie a tutti!**