FOOD FOR THOUGHT
MACHINE VS. HUMAN LEARNING
THE STRUCTURE OF THE PROTON

STEFANO FORTE
UNIVERSITÀ DI MILANO & INFN

PLANCKS 2023

MILANO, MAY 14, 2023

A STORY OF DONUTS

José Ignacio Latorre

ARTIFICIAL INTELLIGENCE: PARADIGMS

"KNOWLEDGE BASED" AI

- LEARN AND IMPLEMENT A SET OF RULES
- GOOD FOR CHESS, BAD FOR REAL LIFE

MACHINE LEARNING

- "INTUITIVE"

  REPRESENTATION

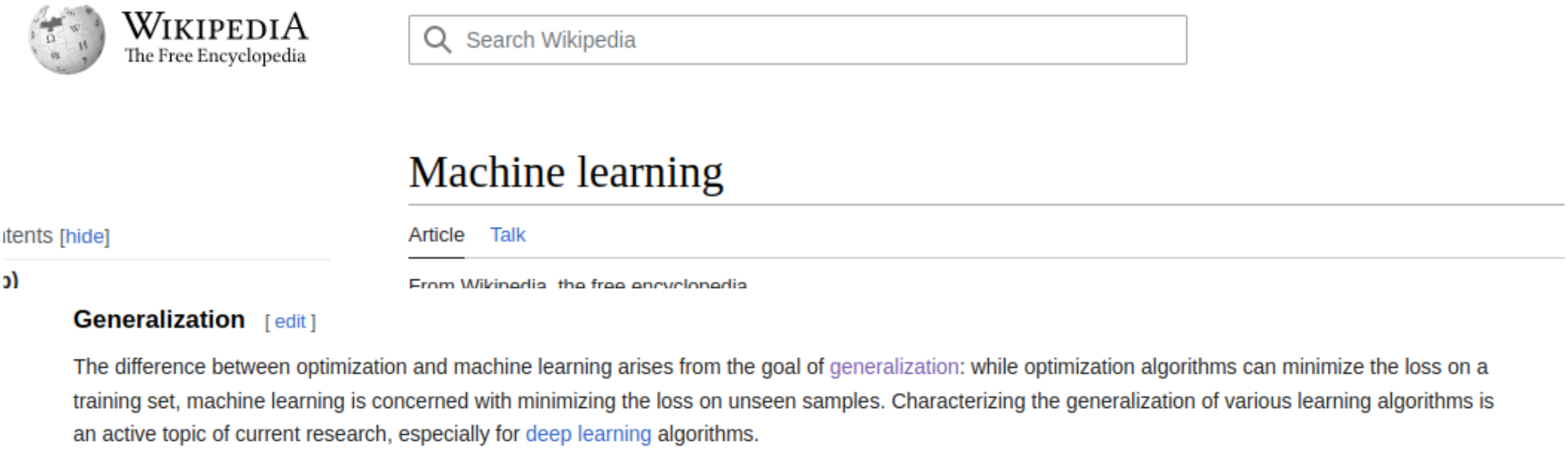
- THE AI MODEL

  BUILID UP

  ITS OWN KNOWLEDGE

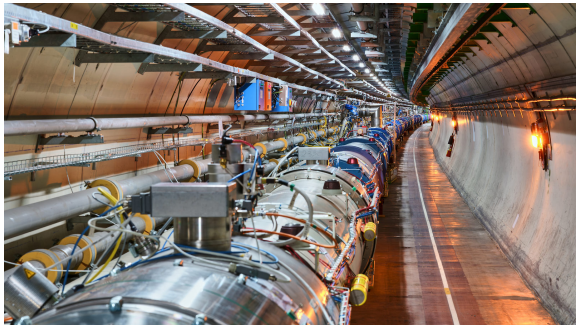
WHAT IS MACHINE LEARNING?

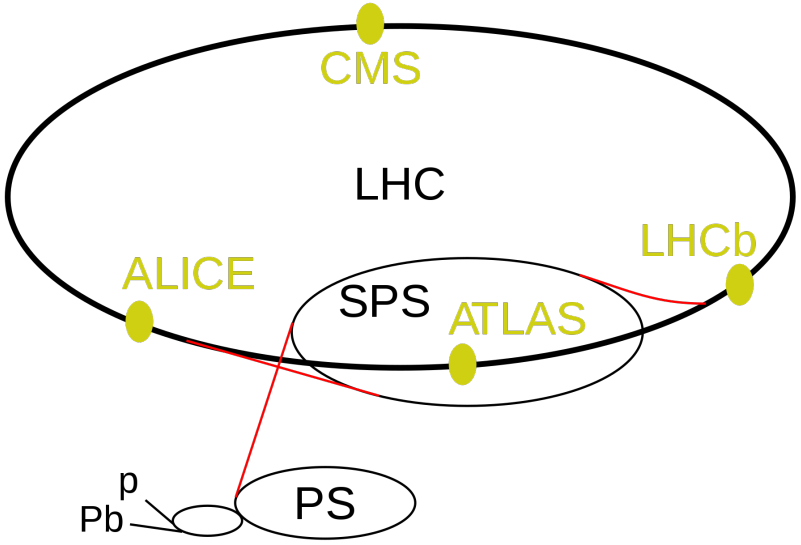
GENERALIZATION

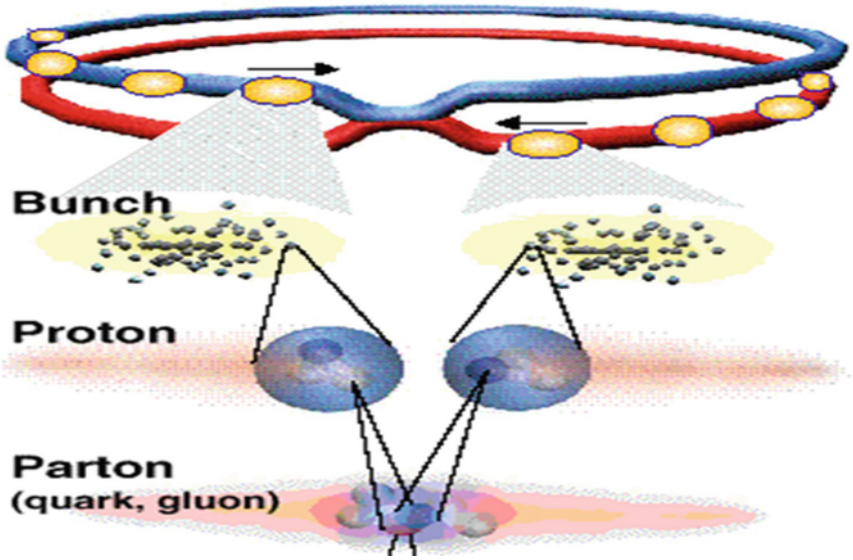


THE PROBLEM:

PROTON COLLISIONS AT THE LHC

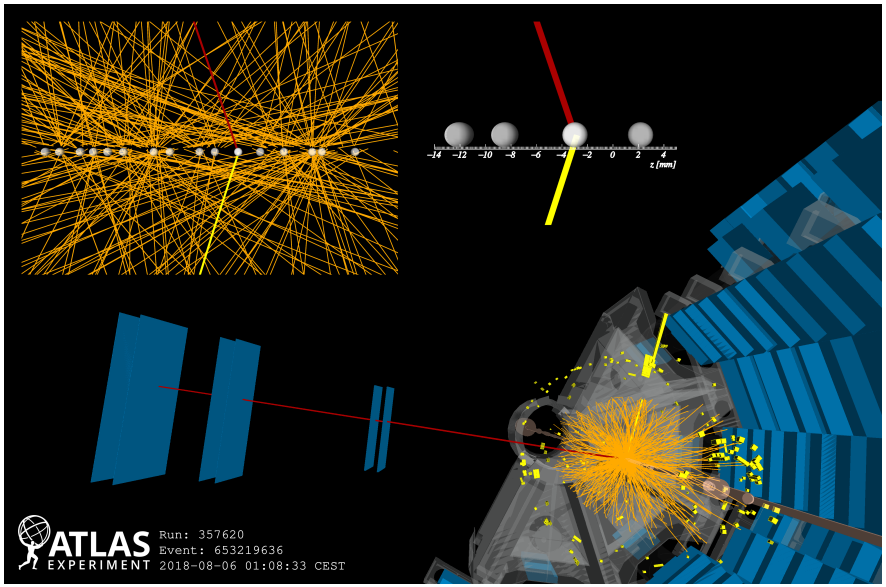




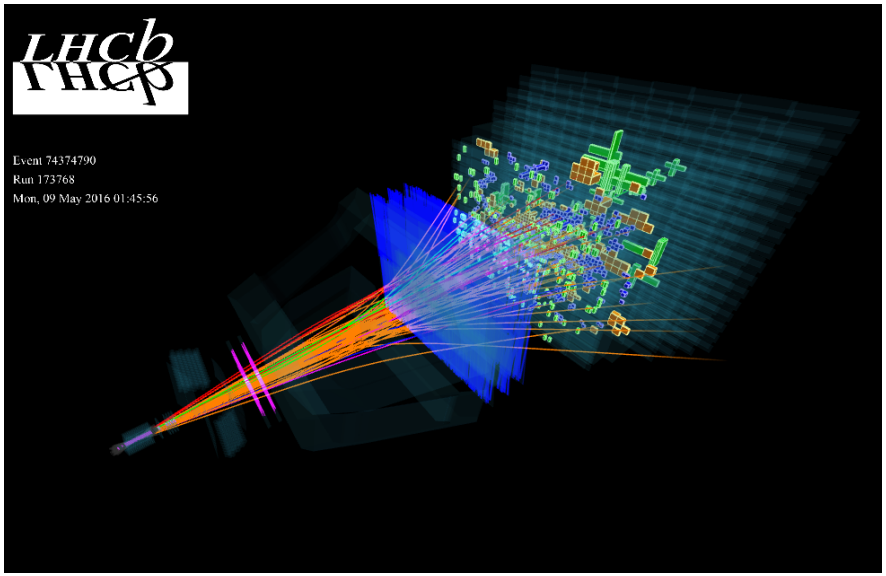


WHAT EXPERIMENTS SEE

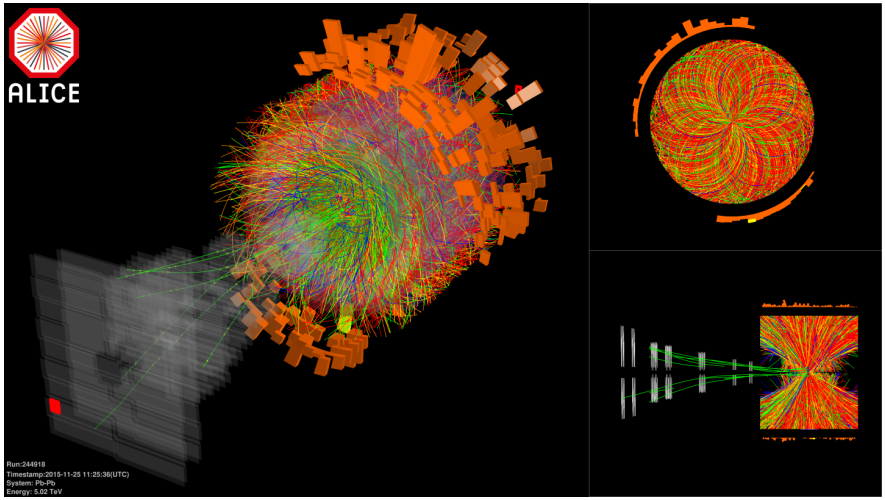
ATLAS



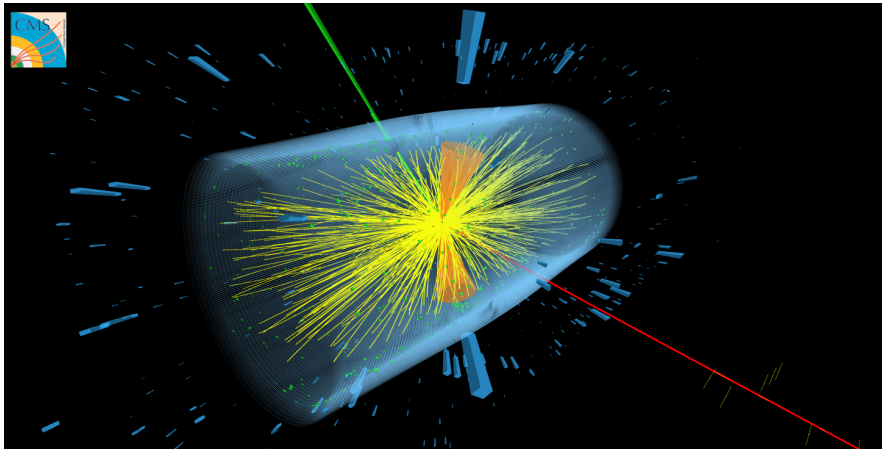
LHCb



ALICE

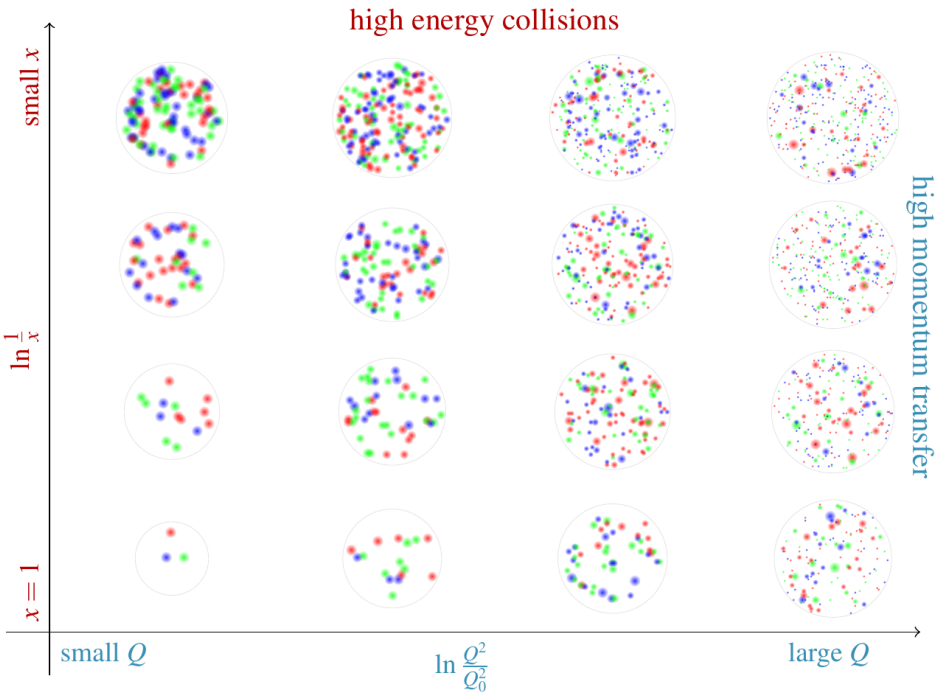


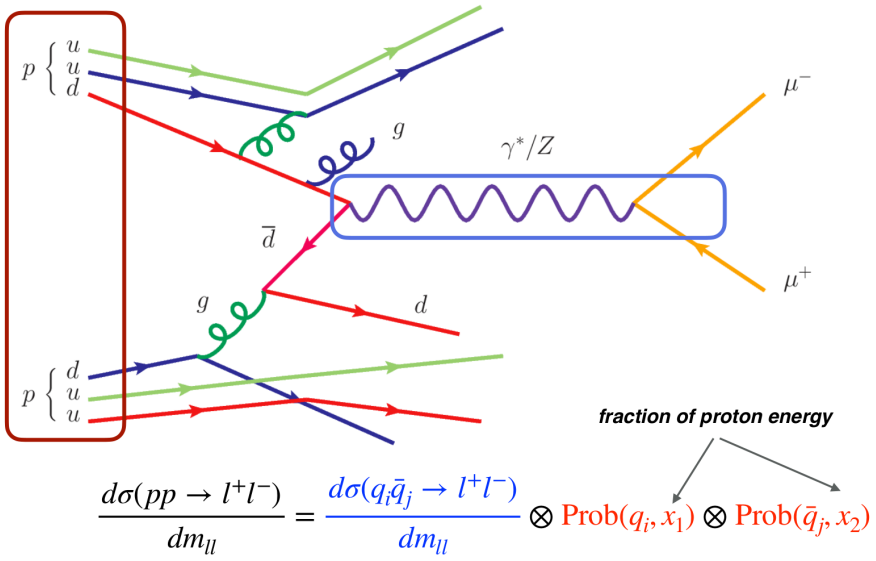
CMS



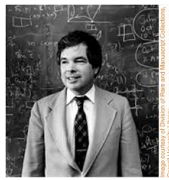
ABOUT 1 BILLION COLLISIONS/SEC; ABOUT 100 PETABYTE/YEAR

INSIDE THE PROTON





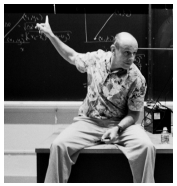
QCD: THE WORK OF MANY PEOPLE



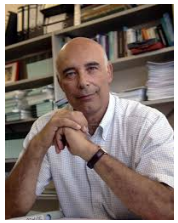
Wilson

Gross

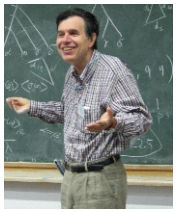
Wilczek



Politzer



Altarelli

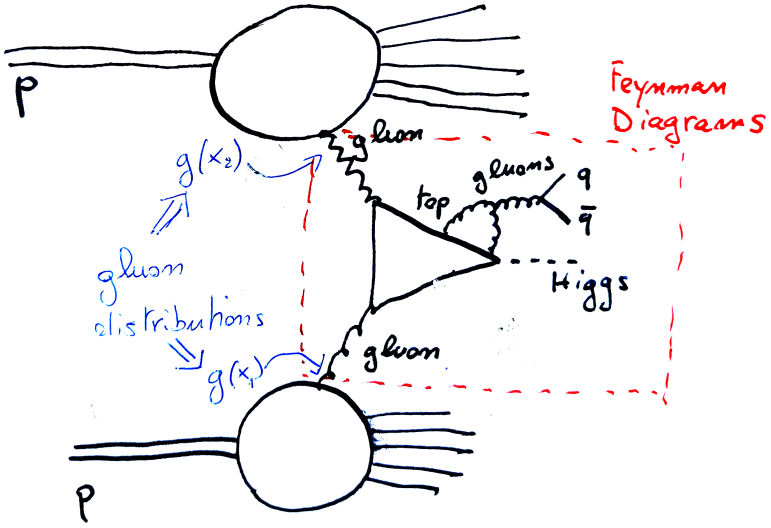


Parisi

Collins

Sterman

HIGGS PRODUCTION AT THE LHC

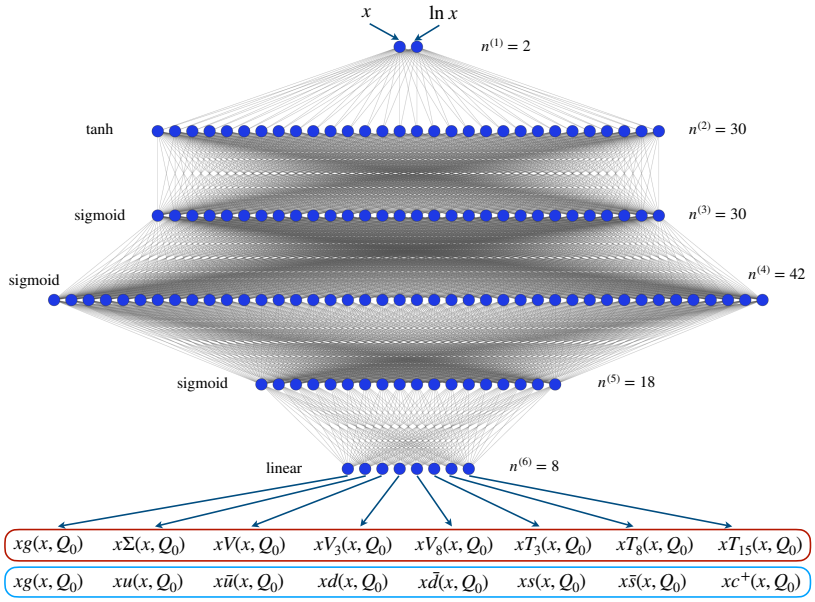


$$\lambda \sim \frac{1}{M_{\text{Higgs}}}$$

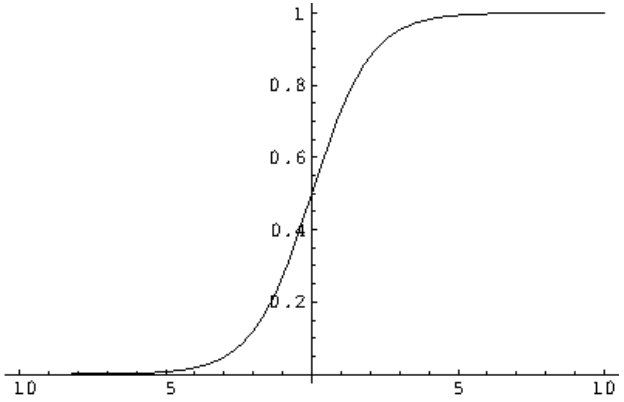
- CAN COMPUTE THE FEYNMAN DIAGRAMS
- CANNOT COMPUTE THE GLUON DISTRIBUTION

SO, YOU WANT TO LEARN THE PROTON?

FIRST IDEA:
NEURAL NETWORKS



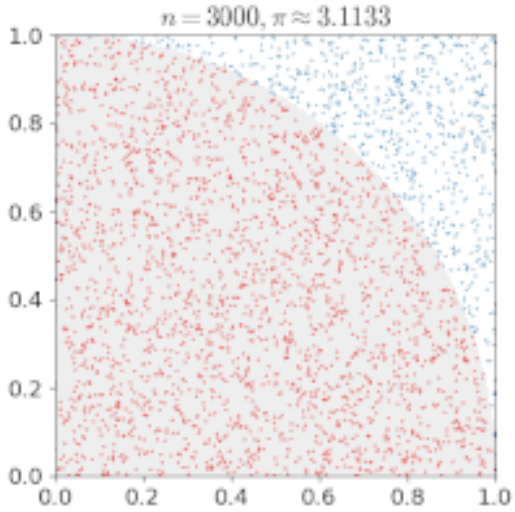
ACTIVATION FUNCTION



$$F^{(i)}_{\rm out}(\vec{x}_{\rm in}) = F\left(\sum_j \omega_{ij} x^j_{\rm in} - \theta_i\right)$$

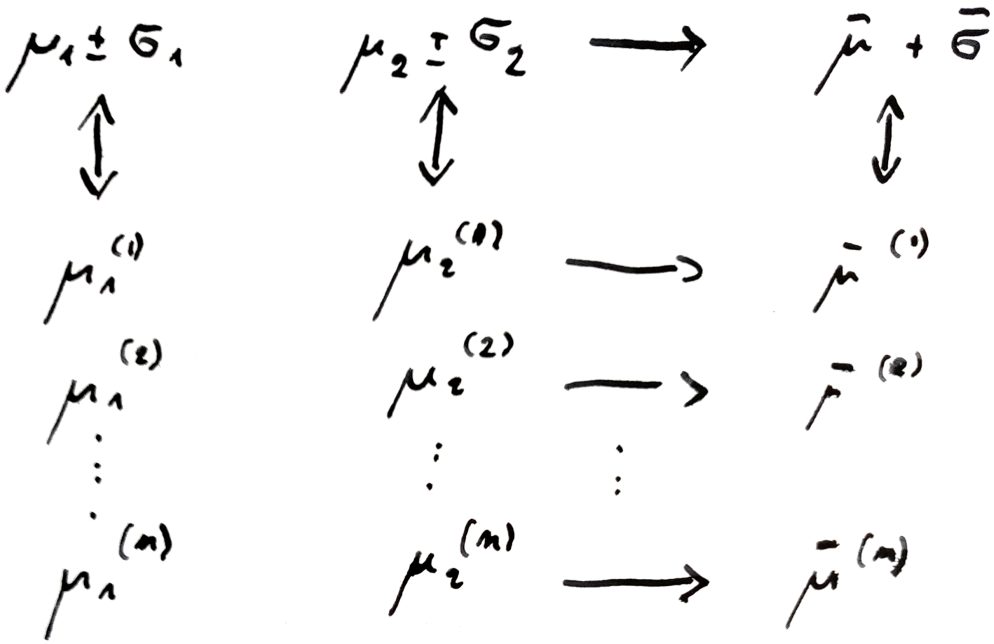
SECOND IDEA:
THE MONTECARLO METHOD

MONTECARLO COMPUTATION OF π



MONTE CARLO COMBINATION
DATA REGRESSION: AVERAGING

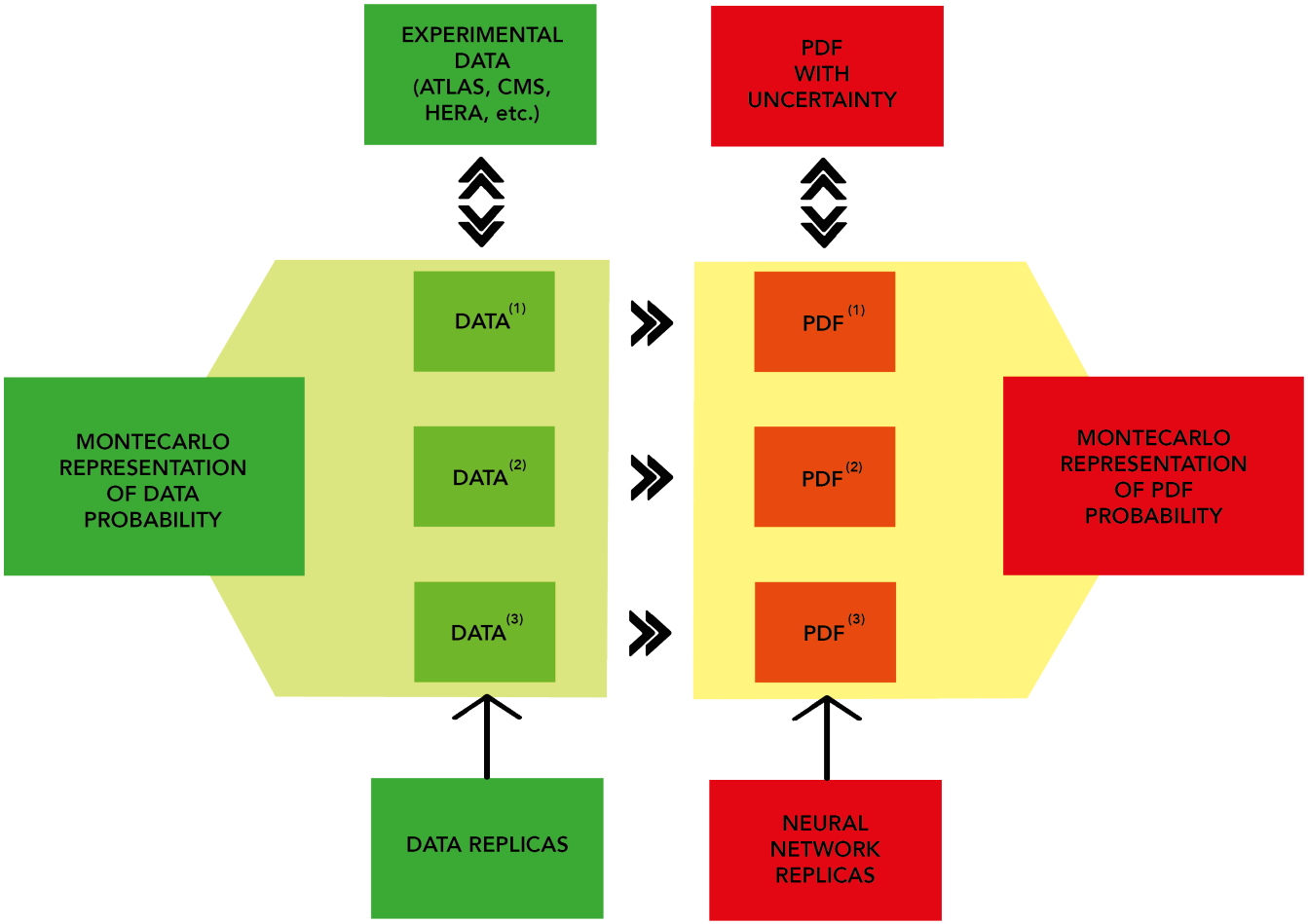
MONTE CARLO REPRESENTATION



MONTE CARLO COMBINATION

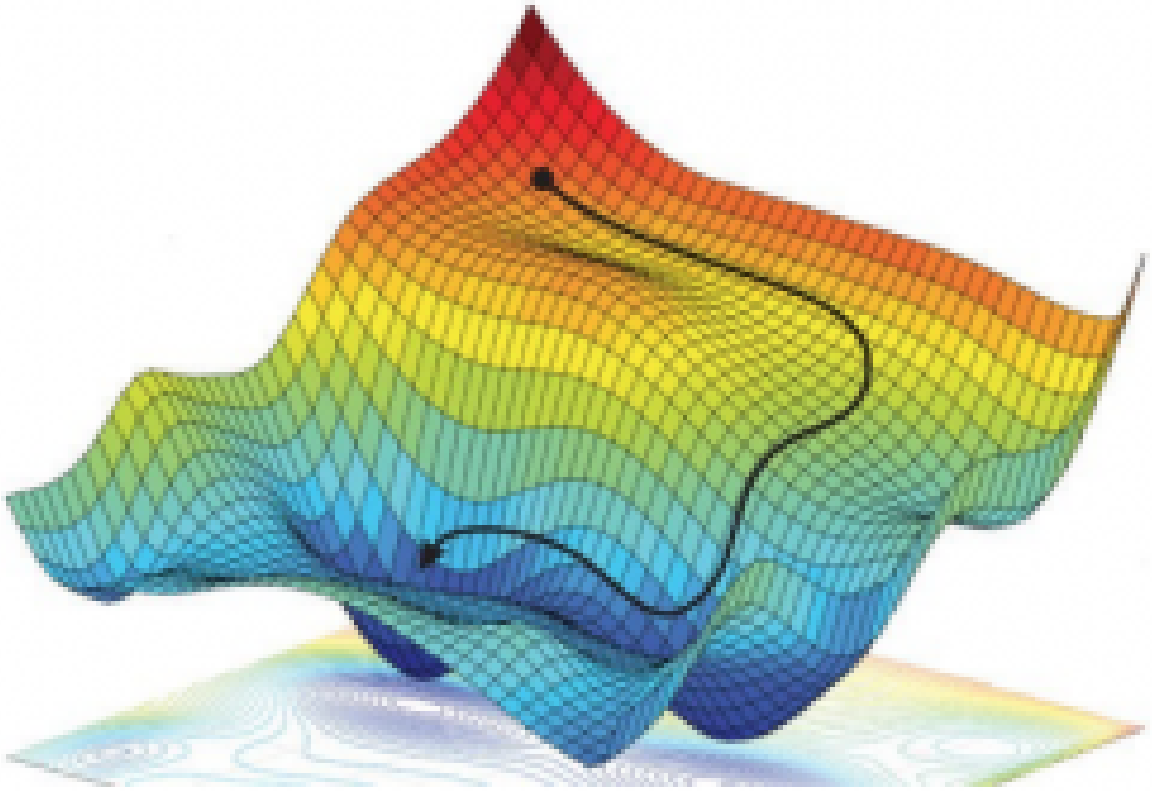
DATA GENERALIZATION: NEURAL NETWORKS

MONTE CARLO REPRESENTATION



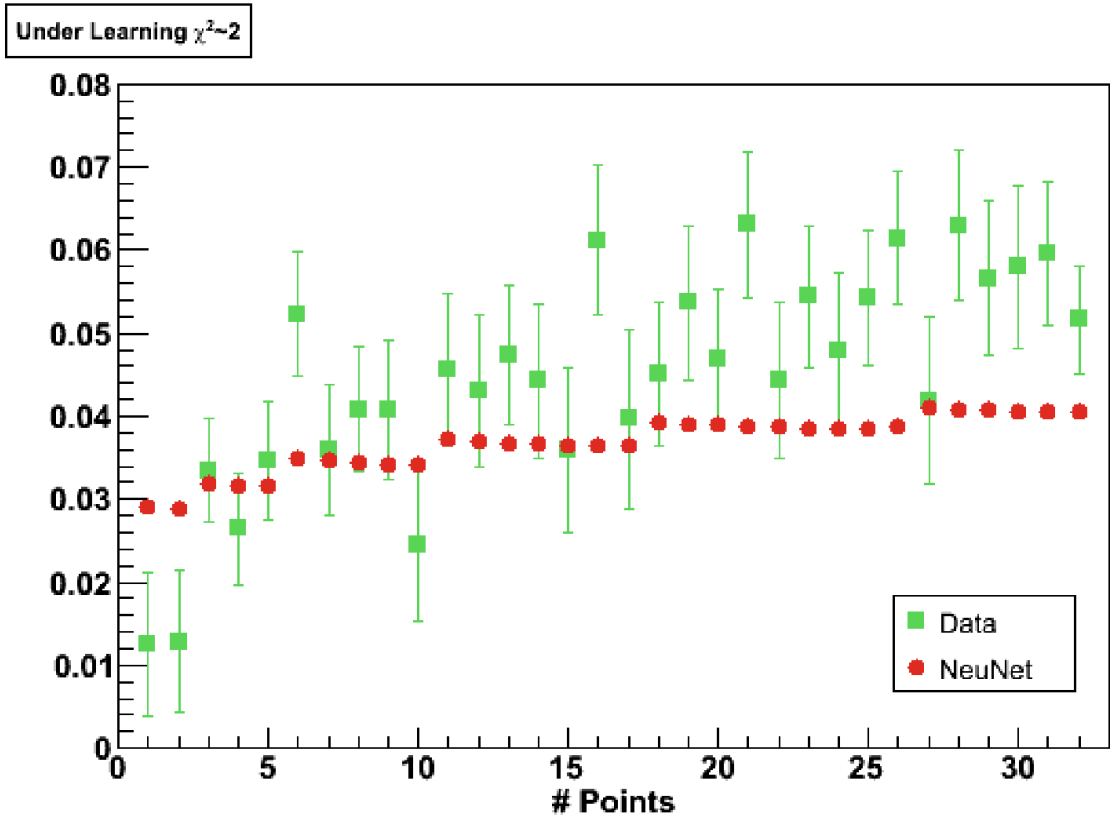
NEURAL NETWORK TRAINING

LOSS MINIMIZATION



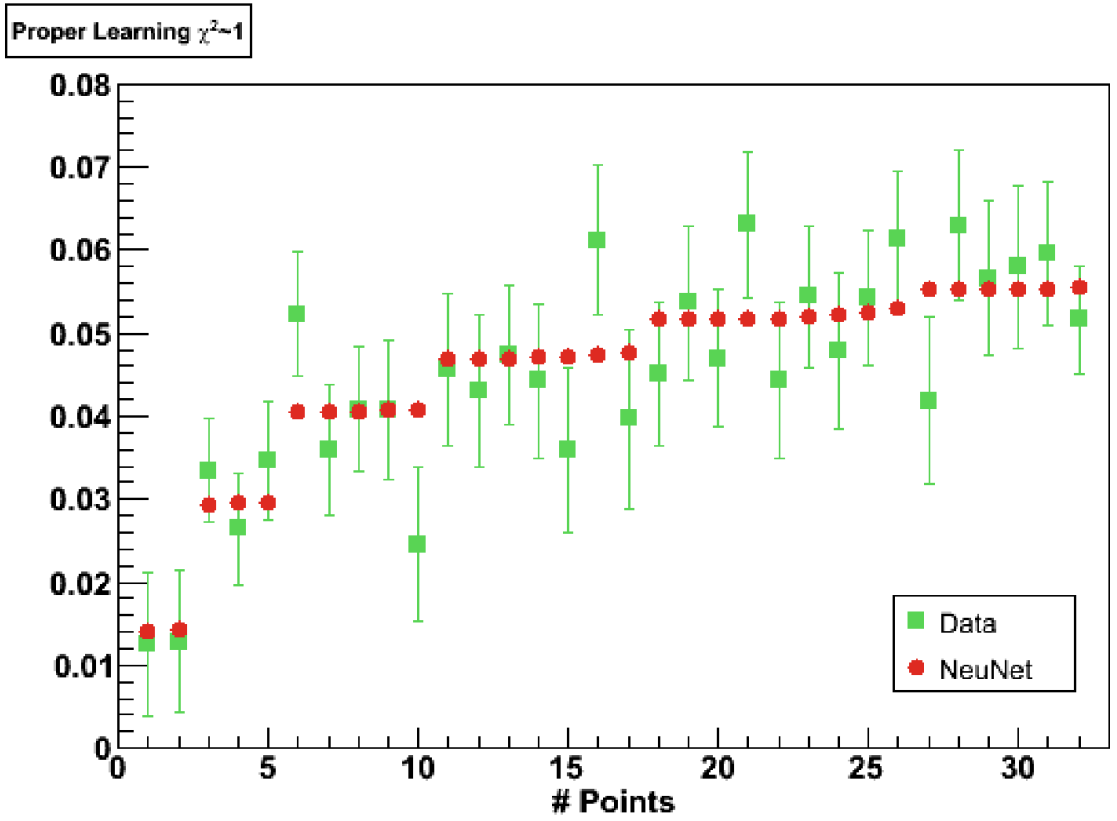
NEURAL NETWORK TRAINING

UNDERLEARNING



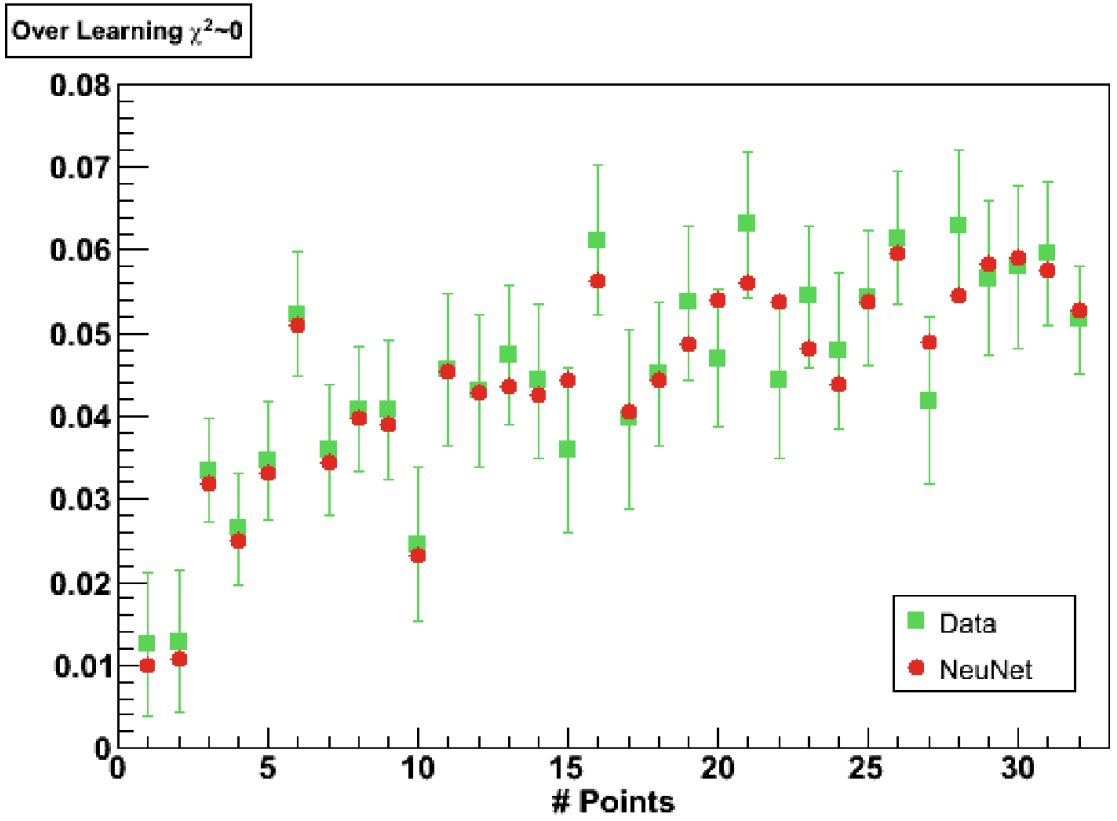
NEURAL NETWORK TRAINING

PROPER LEARNING



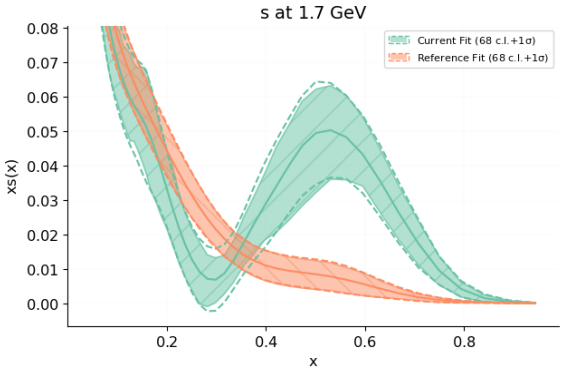
NEURAL NETWORK TRAINING

OVERLEARNING

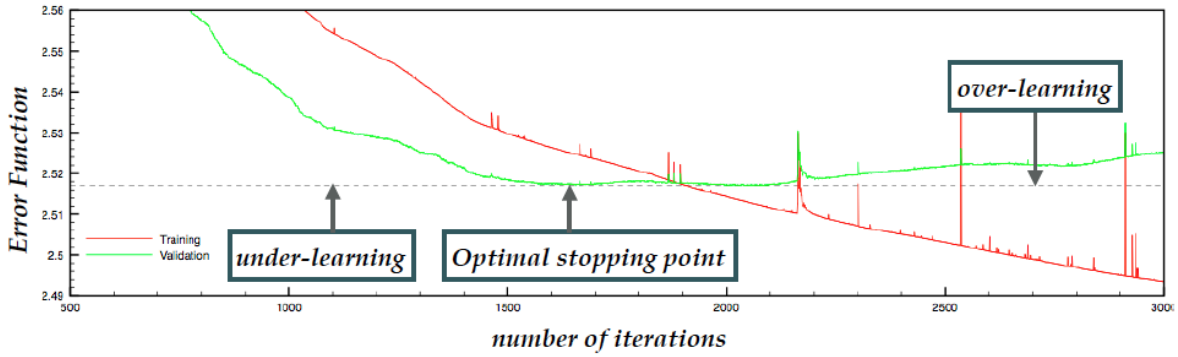


OPTIMAL LEARNING
CROSS-VALIDATION
AN OVERLEARNING SOLUTION

THE STRANGE QUARK DISTRIBUTION OF PROTON MOMENTUM FRACTIONS



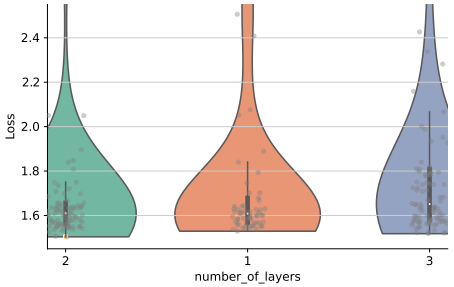
REMOVED
BY THE CROSS-VALIDATION METHOD



WHAT IS THE MACHINE GOOD FOR?

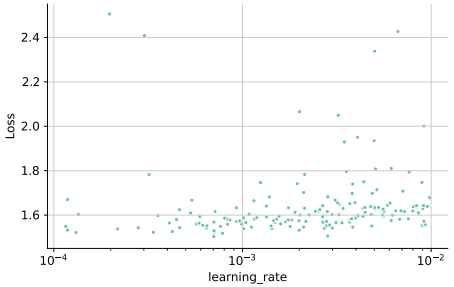
HYPEROPTIMIZATION

number of layers

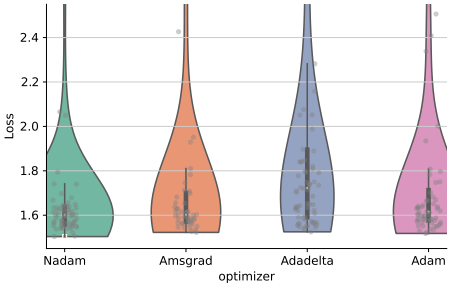


initialization

learning



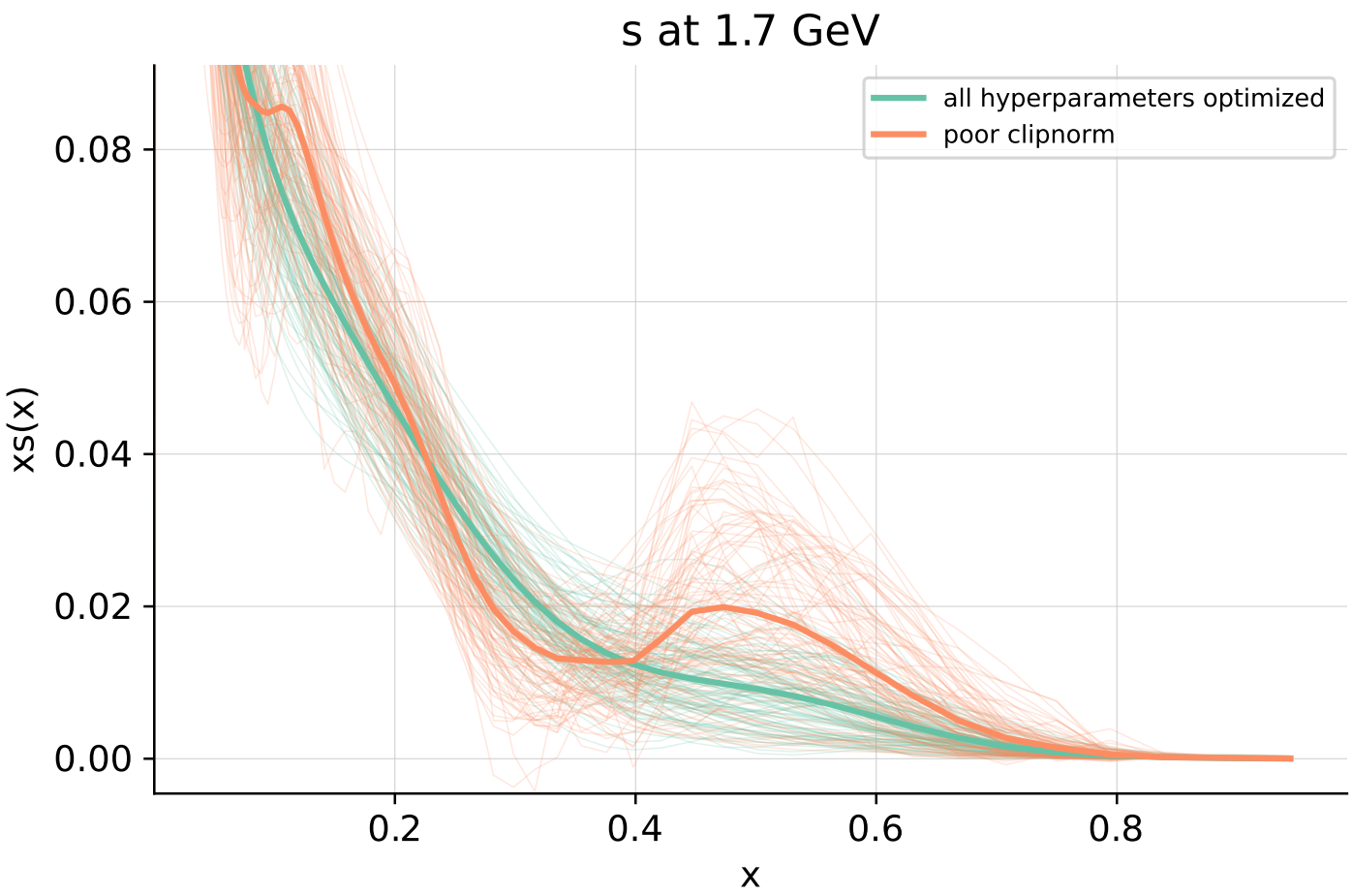
optimization



HYPEROPT PARAMETERS

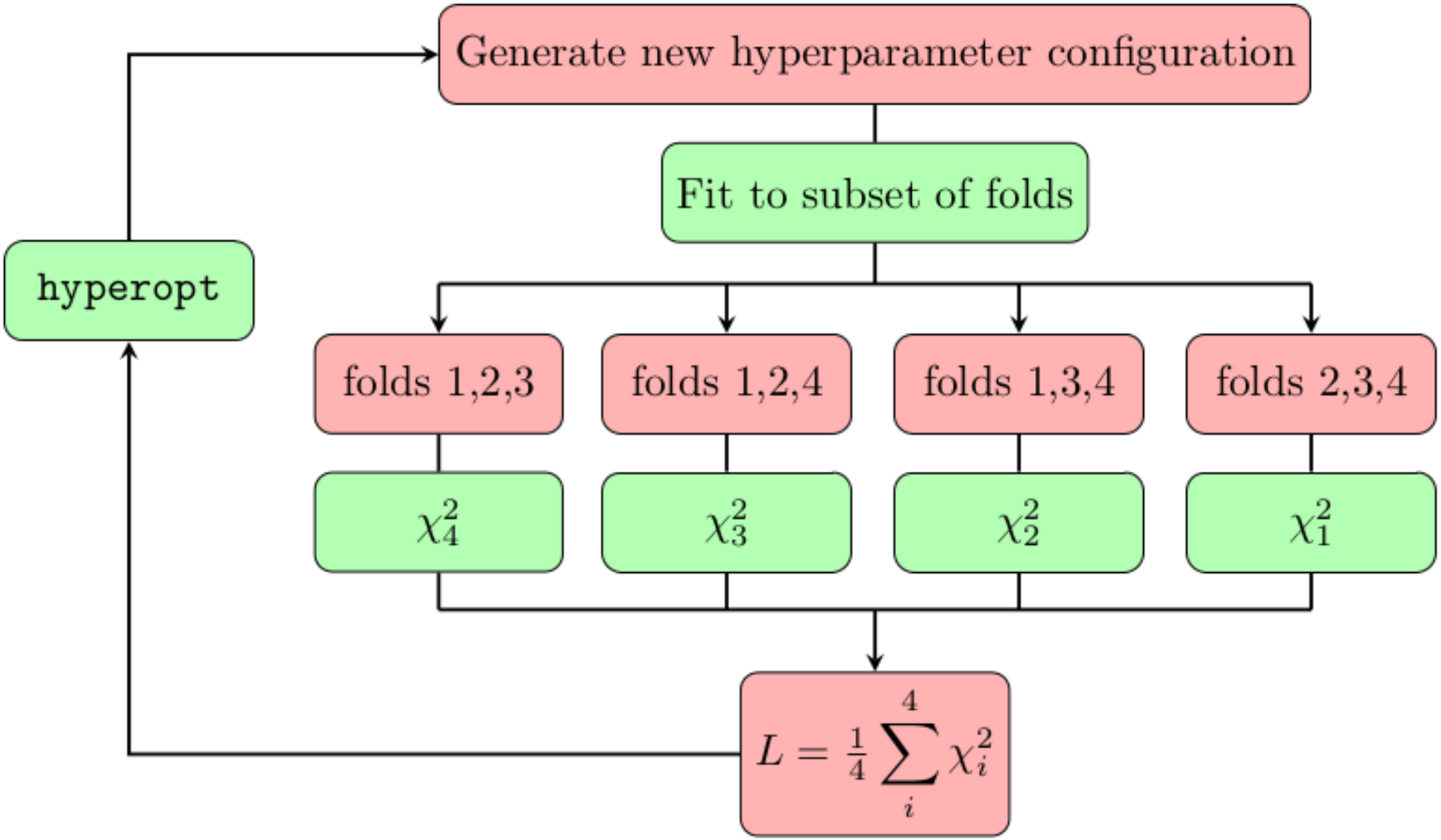
Neural Network	Fit options
Number of layers	Optimizer
Size of each layer	Initial learning rate
Dropout	Maximum number of epochs
Activation functions	Stopping Patience
Initialization functions	Positivity& integrability multiplier

OVERFITTING AGAIN!



GENERALIZATION!

K-FOLDING

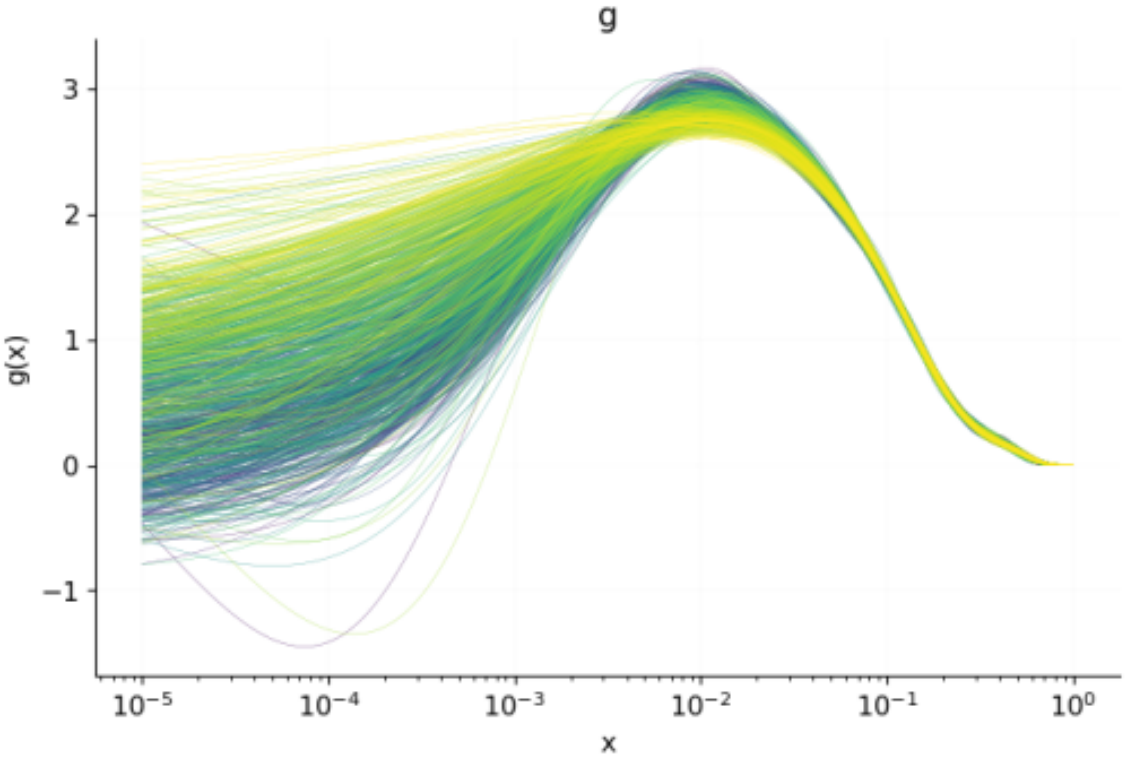


THE RESULT

DO WE REALLY NEED THE MACHINE?

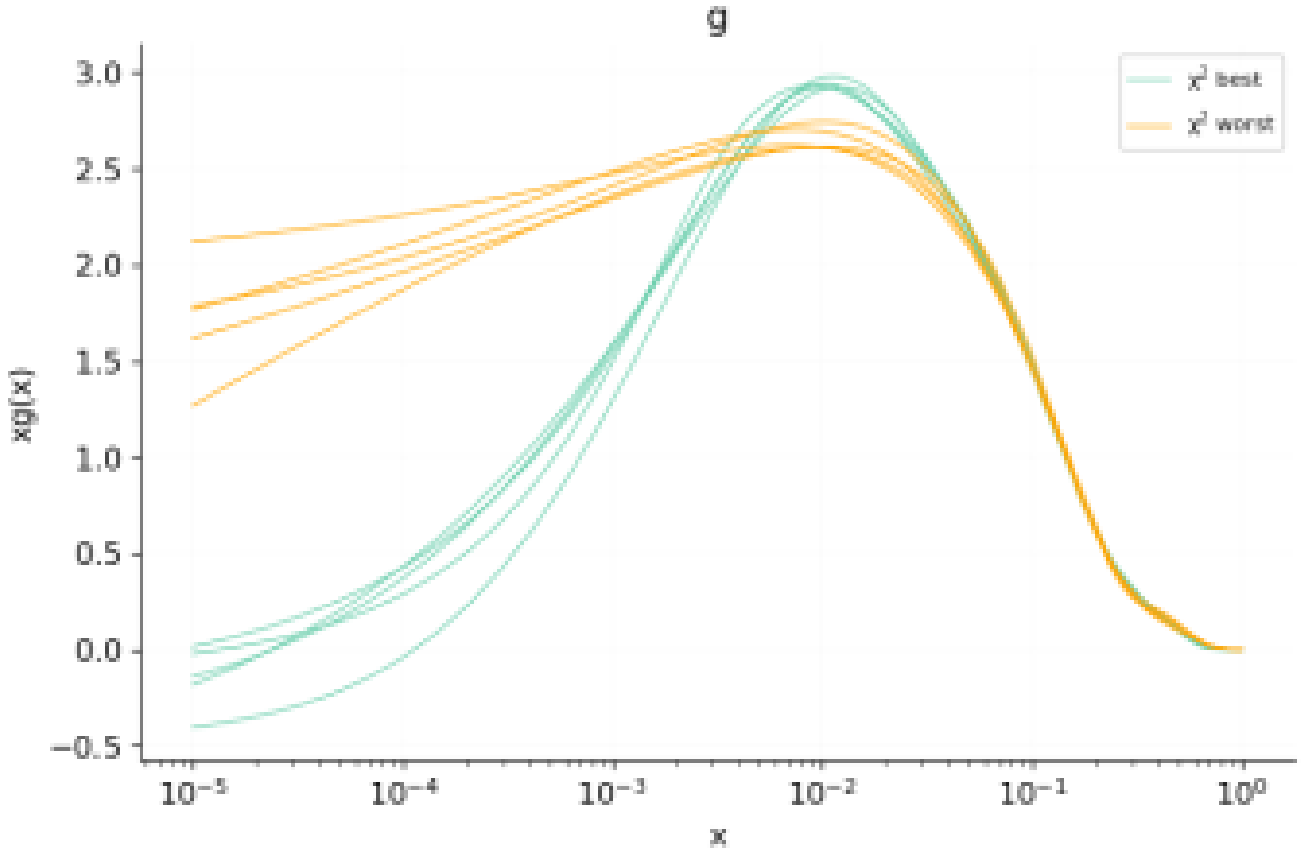
A CLOSER LOOK AT THE GLUON

WORST VS BEST AGREEMENT WITH DATA



YET CLOSER....

WORST VS BEST AGREEMENT WITH DATA

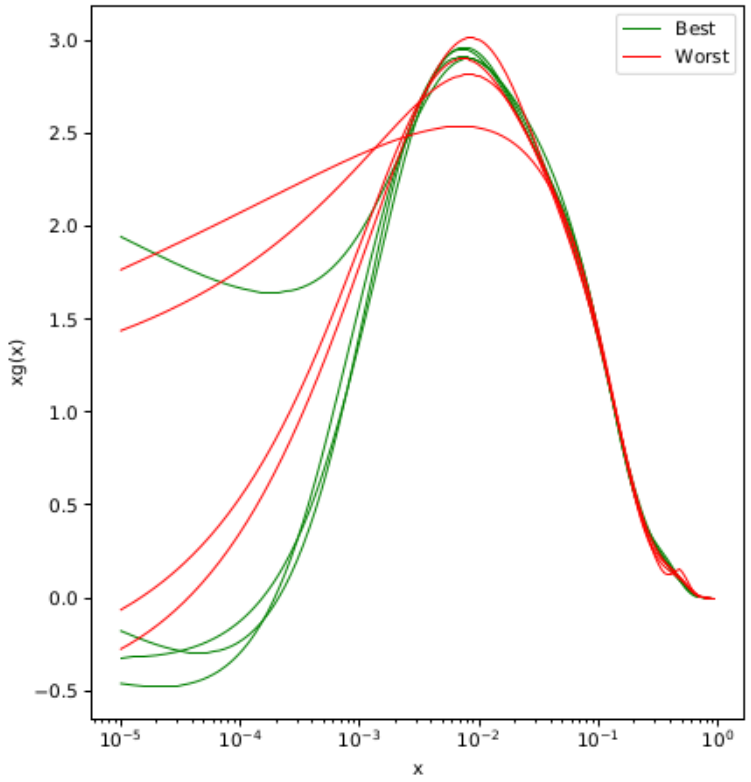


WHAT IS GOING ON?

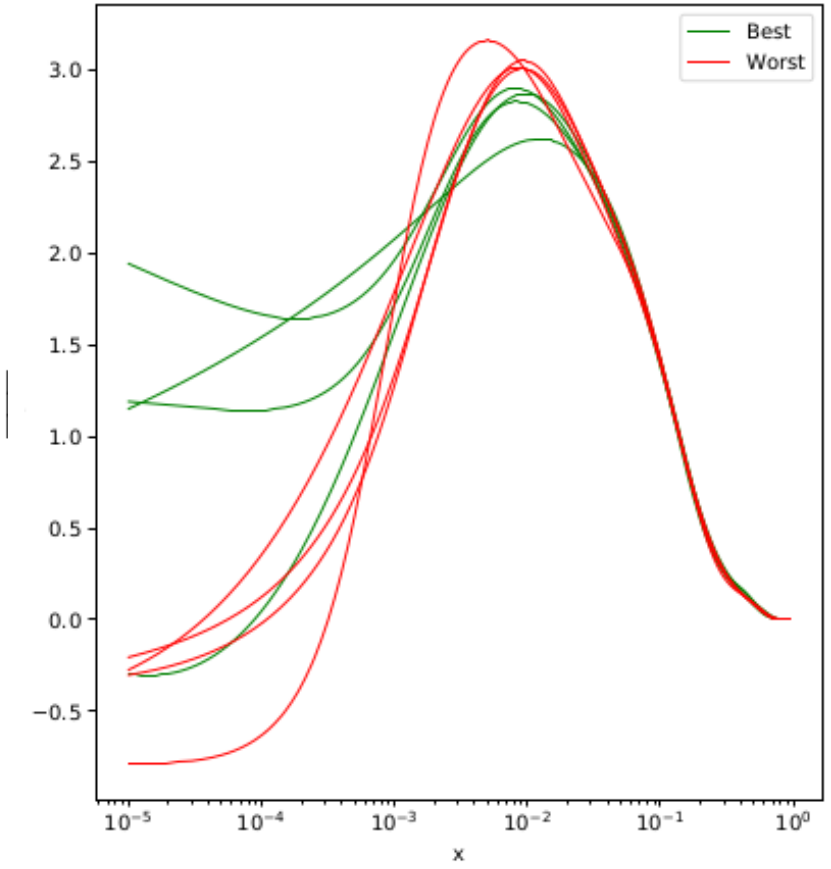
EXPLANATION

GENERALIZATION!

FITTED FOLDS



EXCLUDED FOLD



- BEST VS WORST REVERSED
- "BEST" DO NOT GENERALIZE

SAYS THE MACHINE!

SO, DO WE REALLY

NEED THE MACHINE?

LET'S ASK CHATGPT!