NEURAL NETWORKS

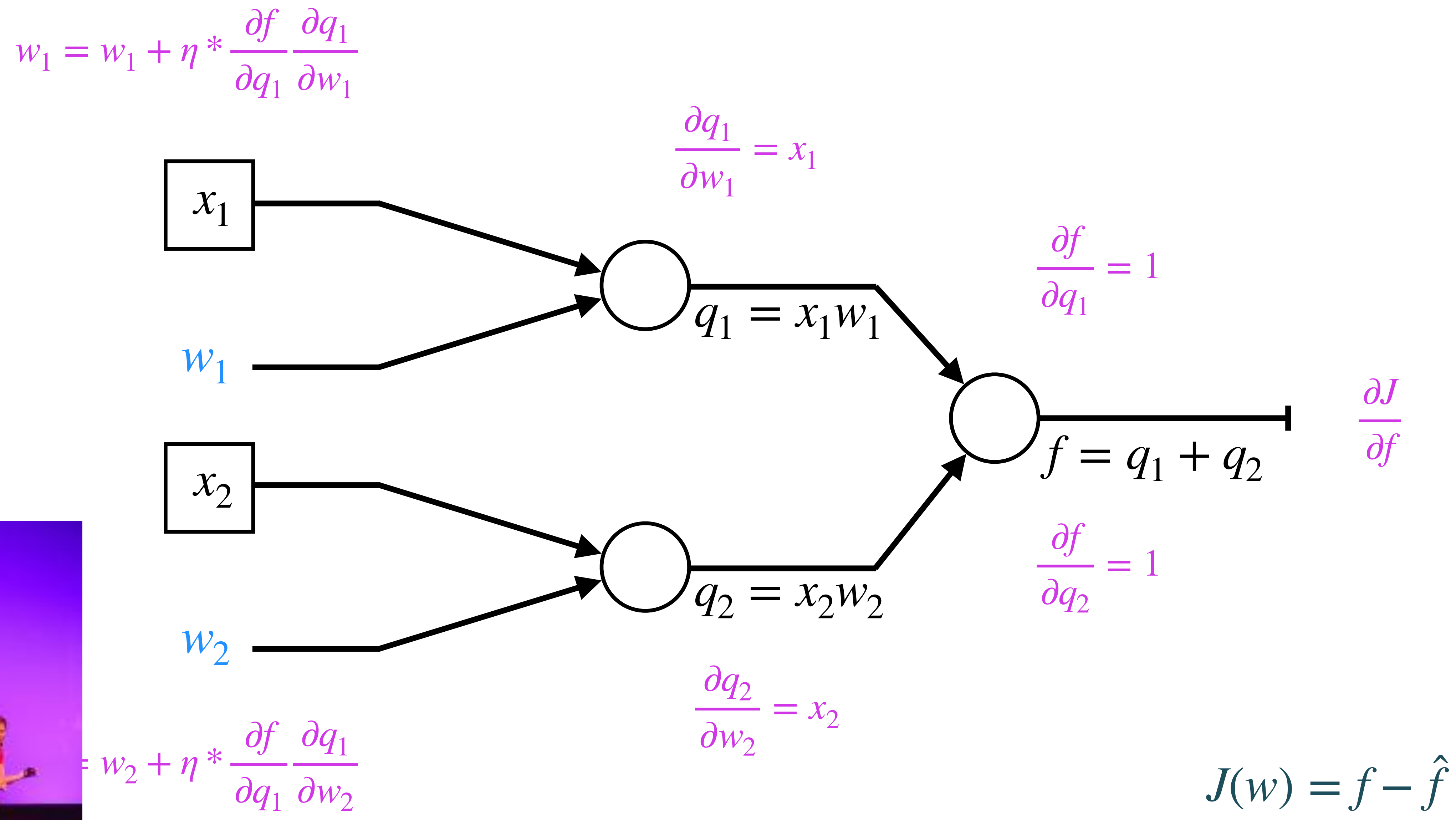
AND DEEP LEARNING

MICHELLE KUCHERA
DAVIDSON COLLEGE

CPS-FR
MIT
21 AUGUST 2025

MICHELLE KUCHERA

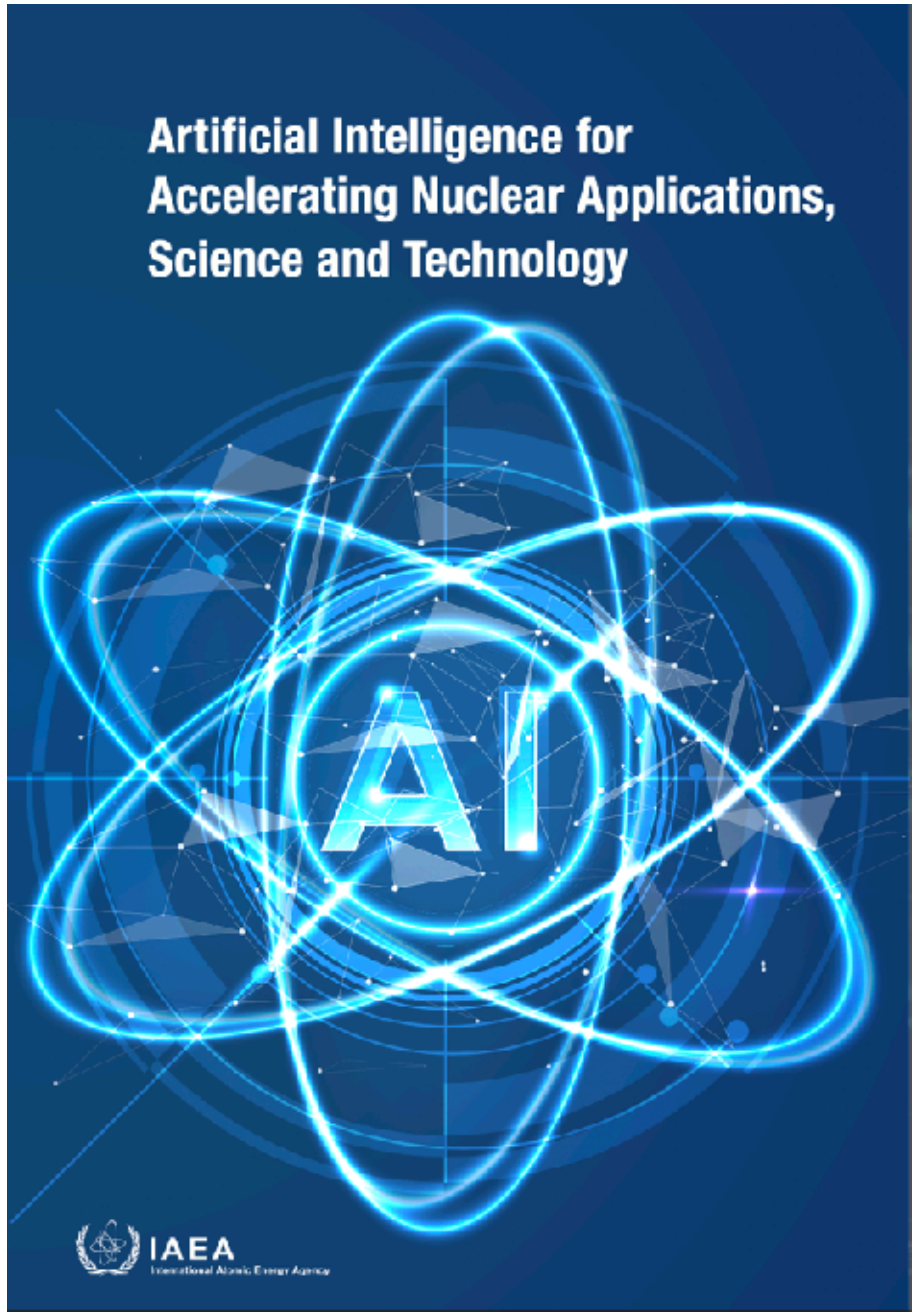
B.S., M.S. PHYSICS
M.S., PH.D. COMPUTATIONAL SCIENCE

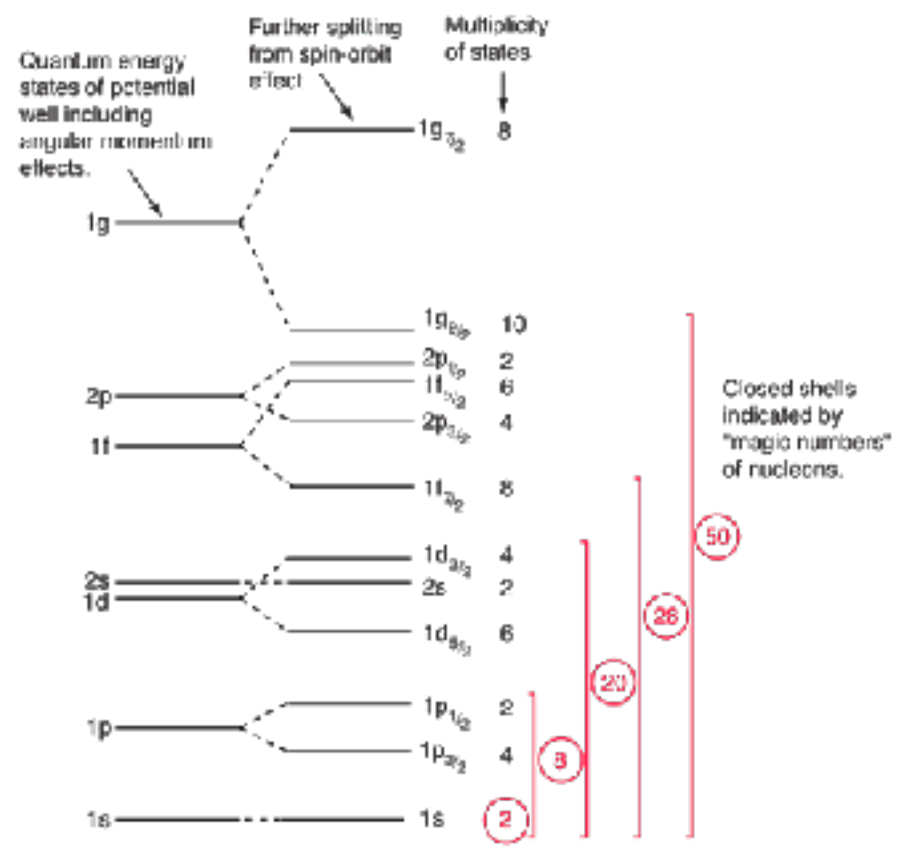


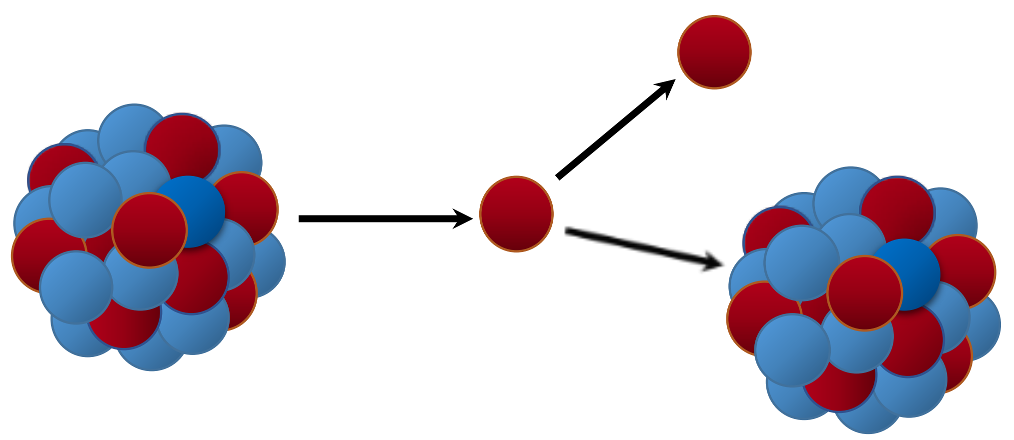
PhD: GPUs for Bayesian Neural Networks (😟)

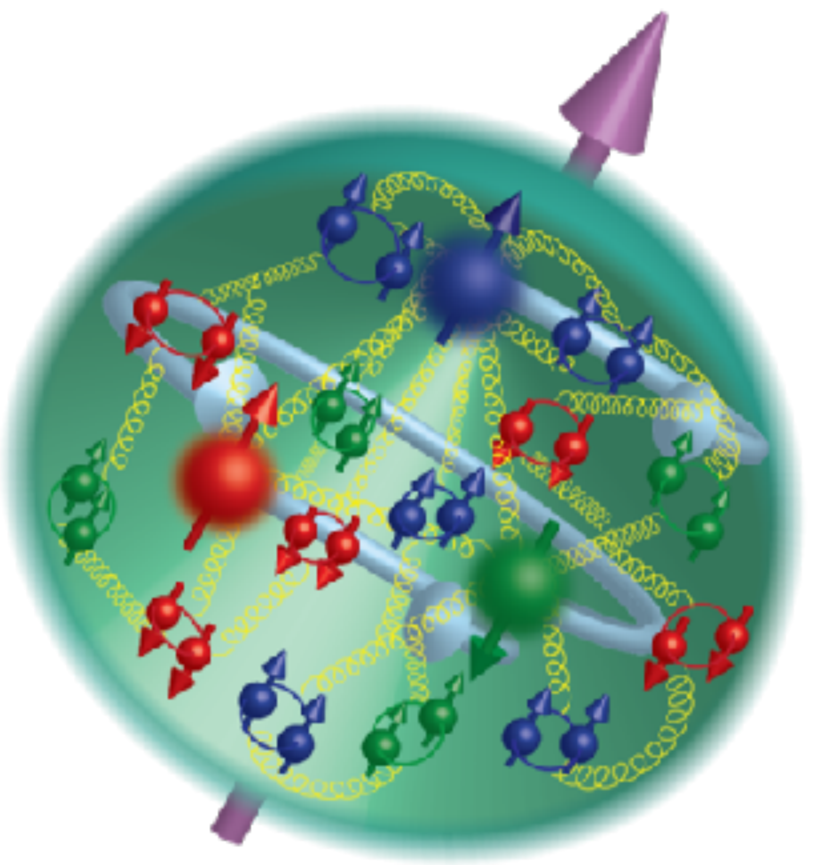
MICHELLE KUCHERA

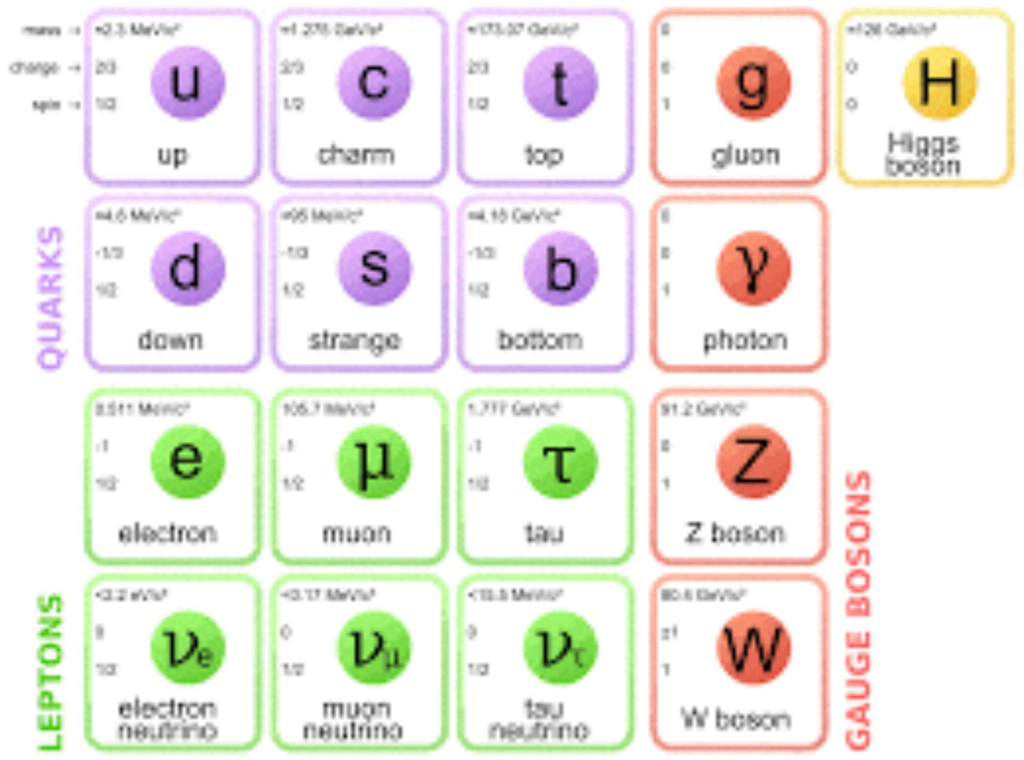
B.S., M.S. PHYSICS
M.S., PH.D. COMPUTATIONAL SCIENCE









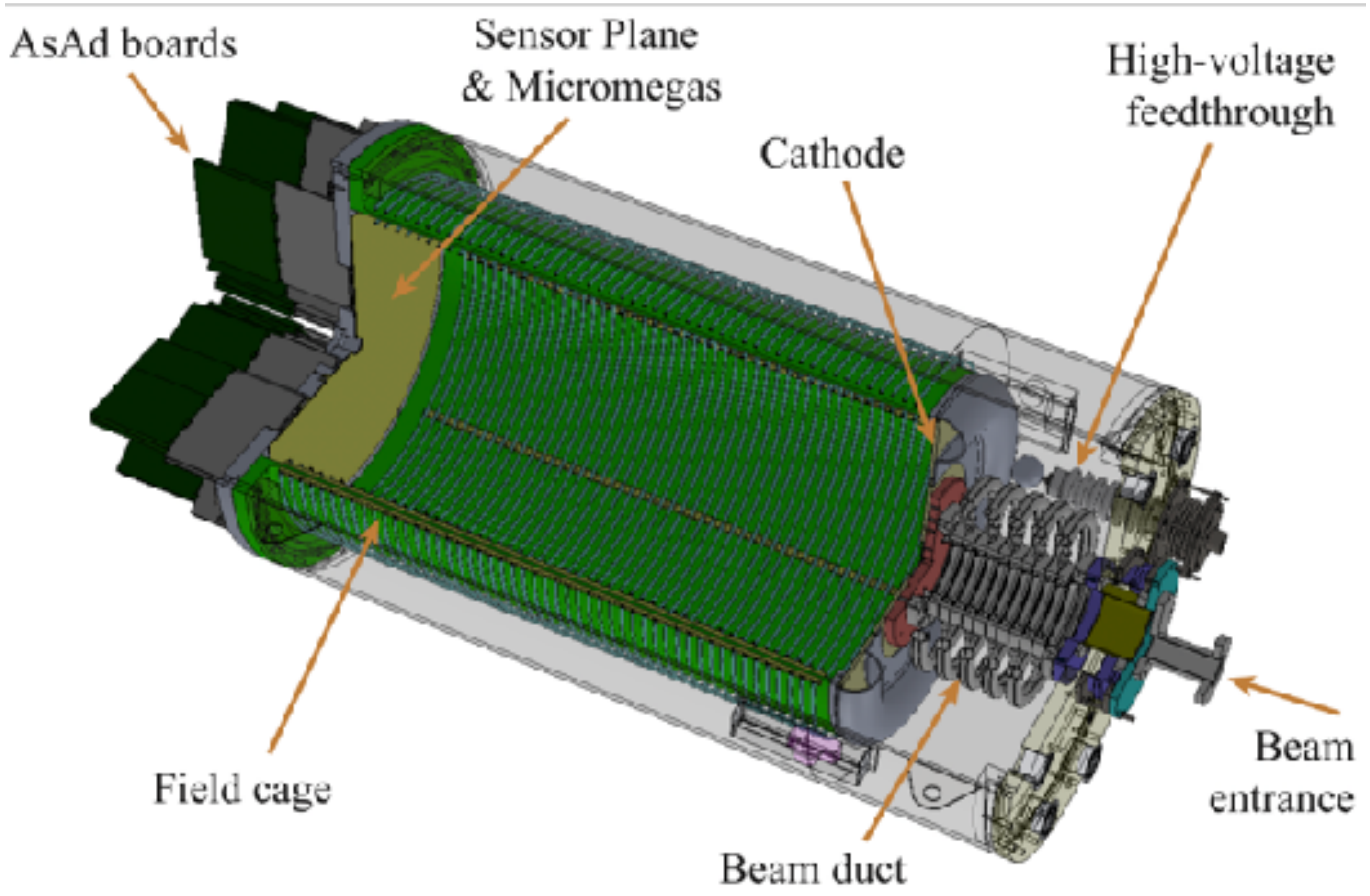


FRIB

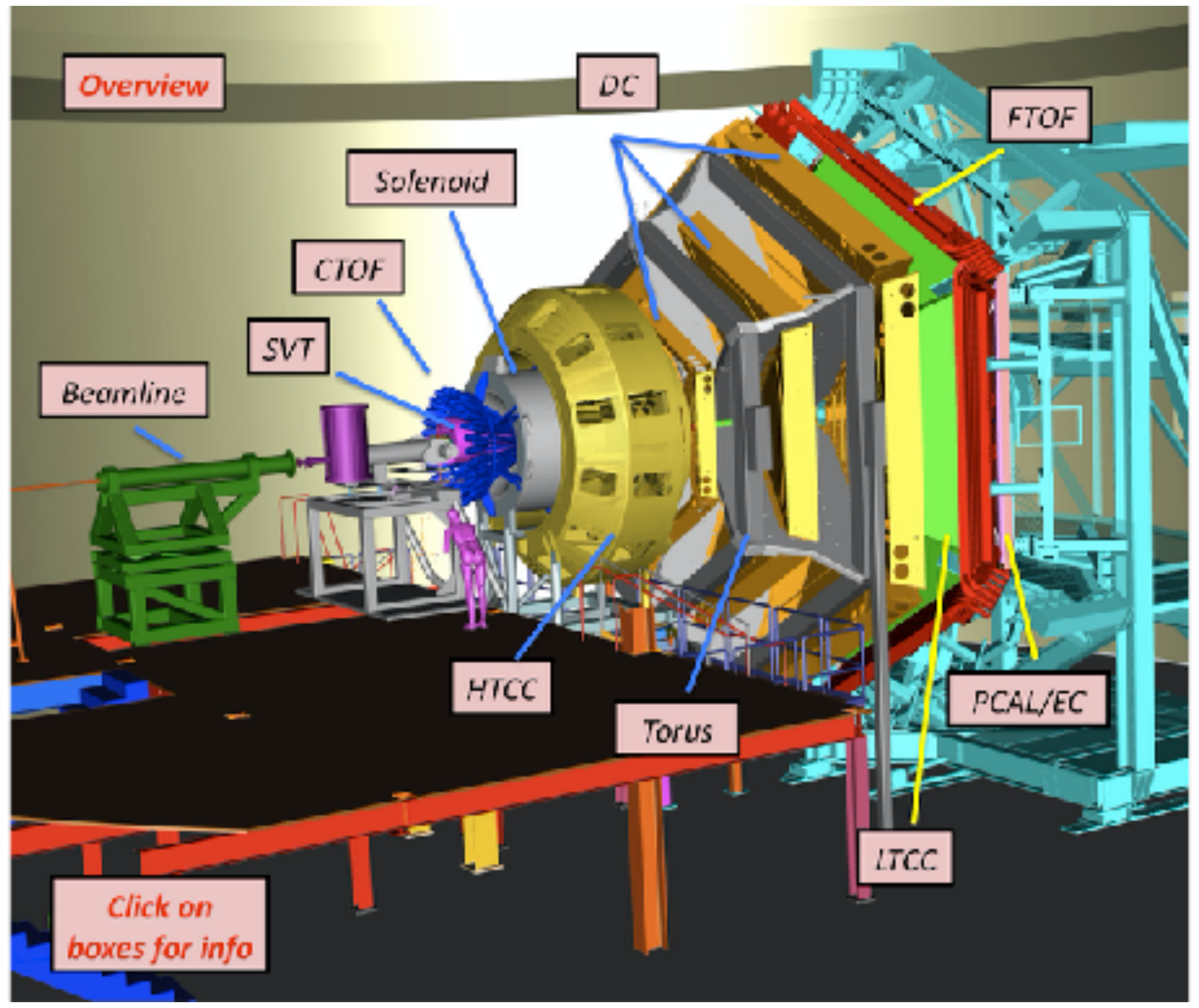
Jefferson Lab

CERN

EXPERIMENTAL DATA



J. BRADT ET. AL., *NUCLEAR INSTRUMENTS AND METHODS*, 2017.



FRIB

AT-TPC

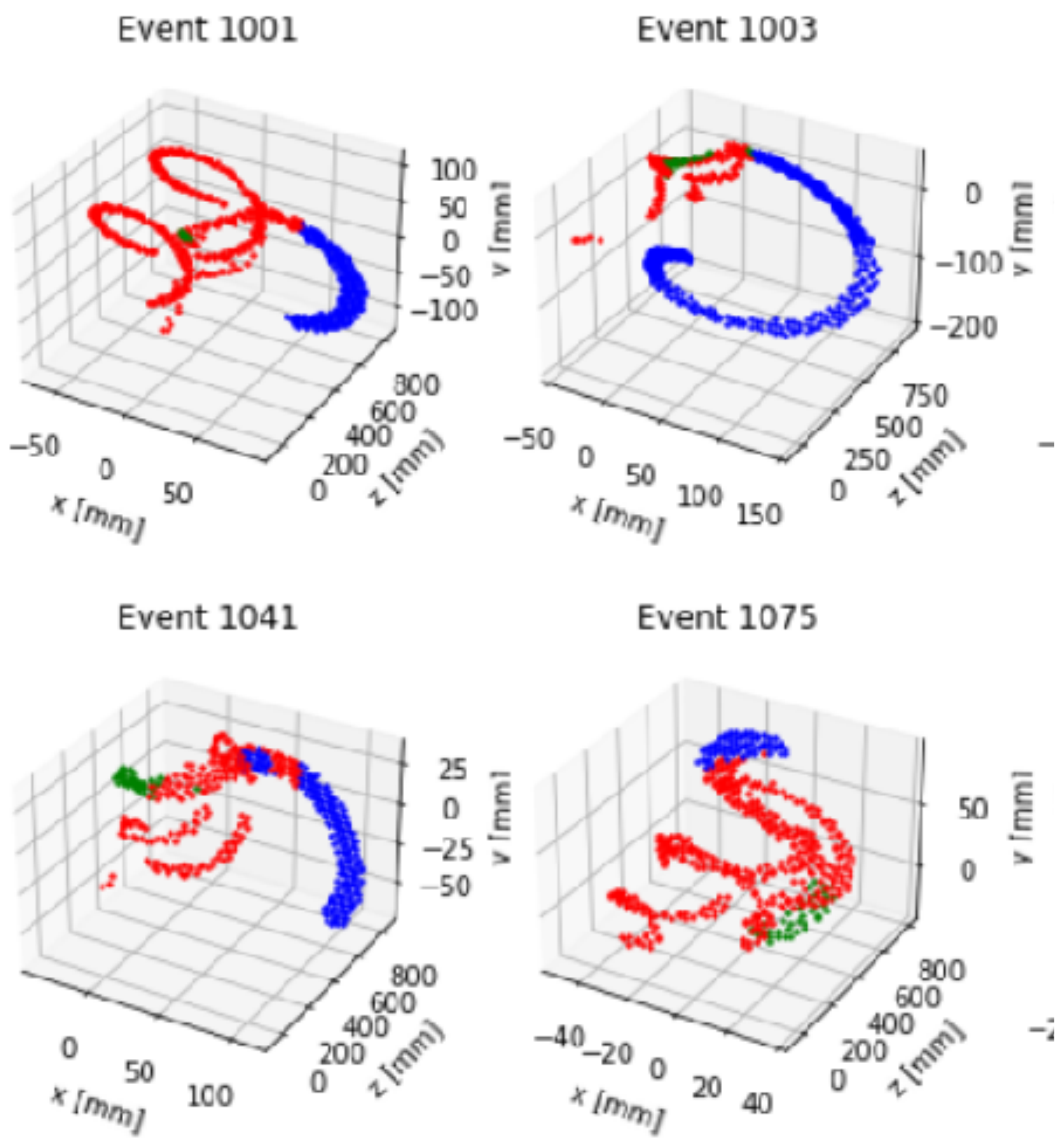
Jefferson Lab

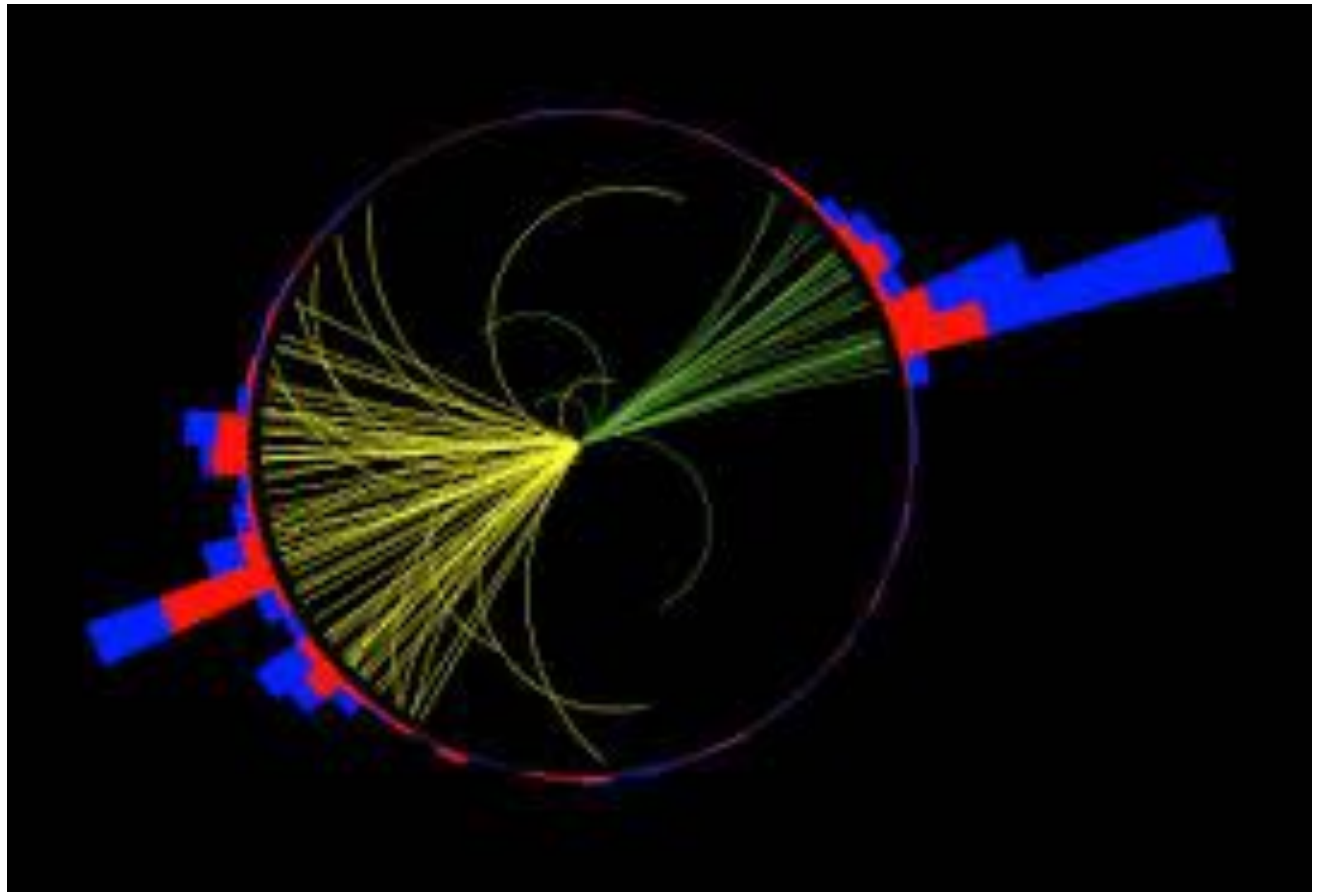
CLAS 12

CERN

CMS

EXPERIMENTAL DATA





FRIB

Jefferson Lab

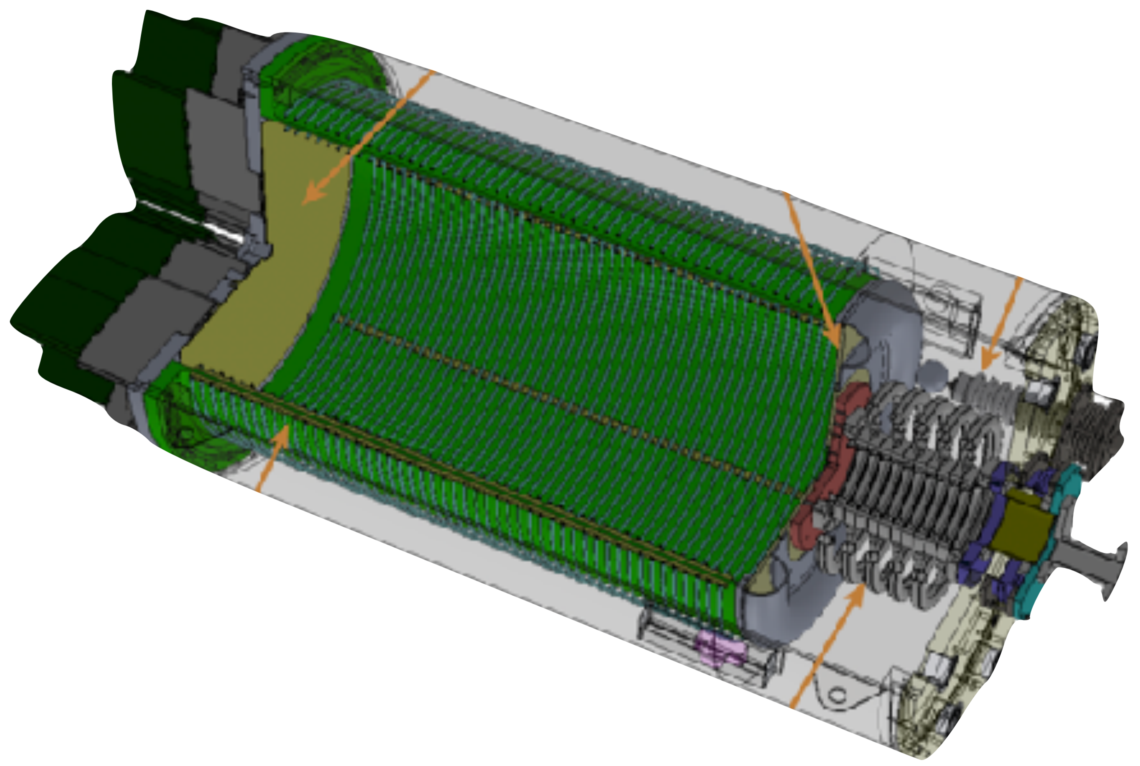
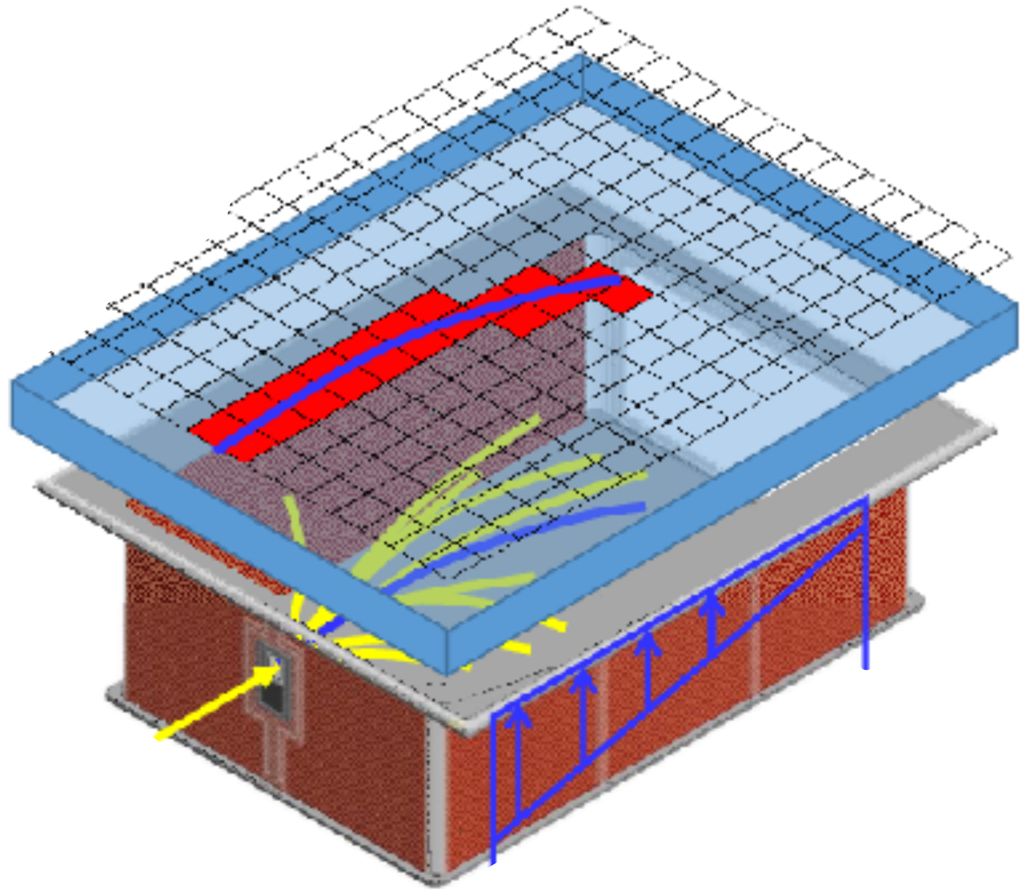
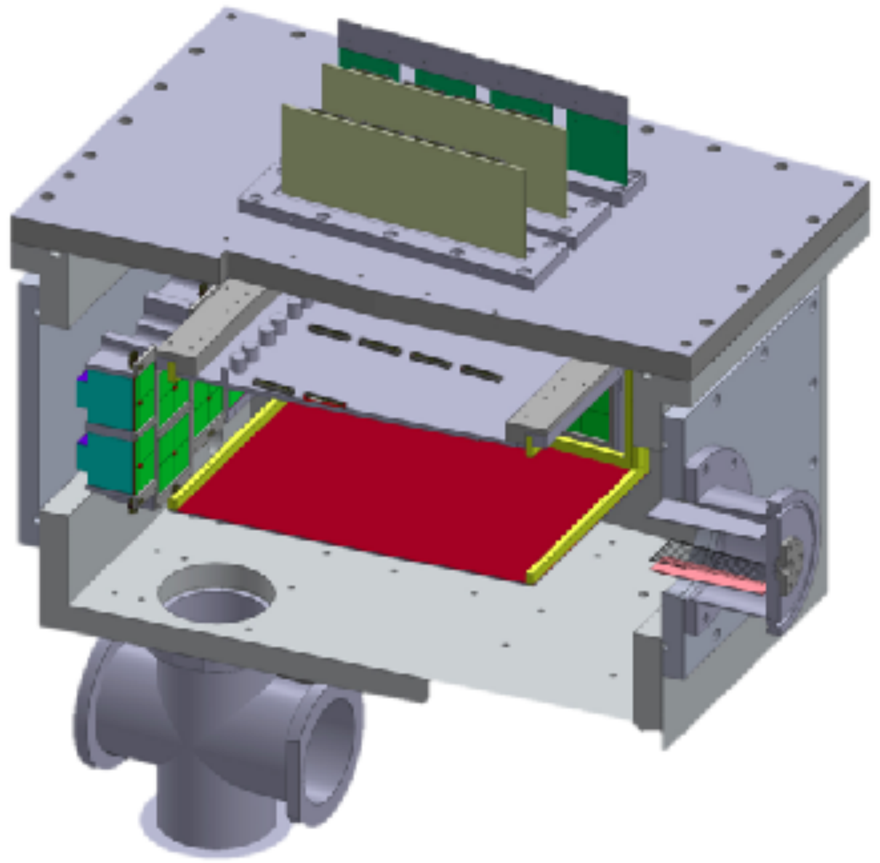
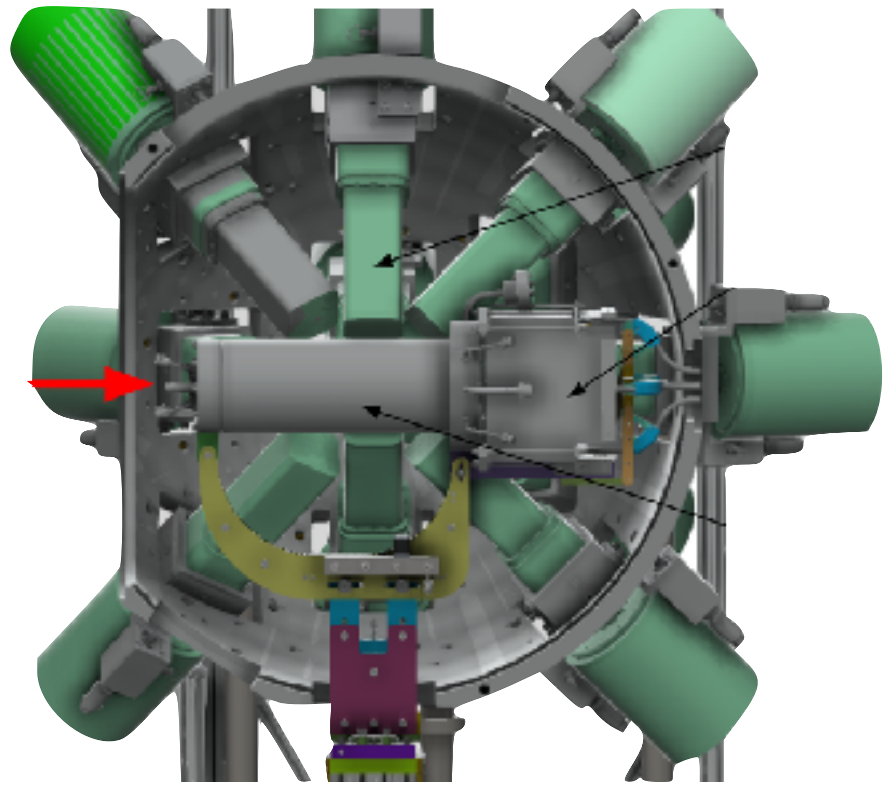
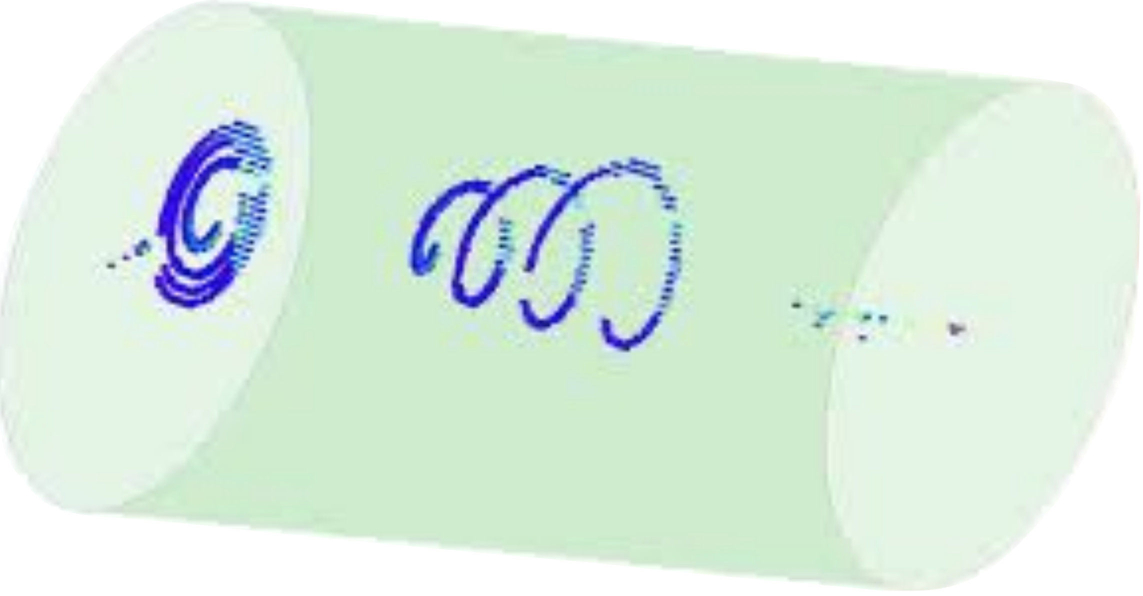
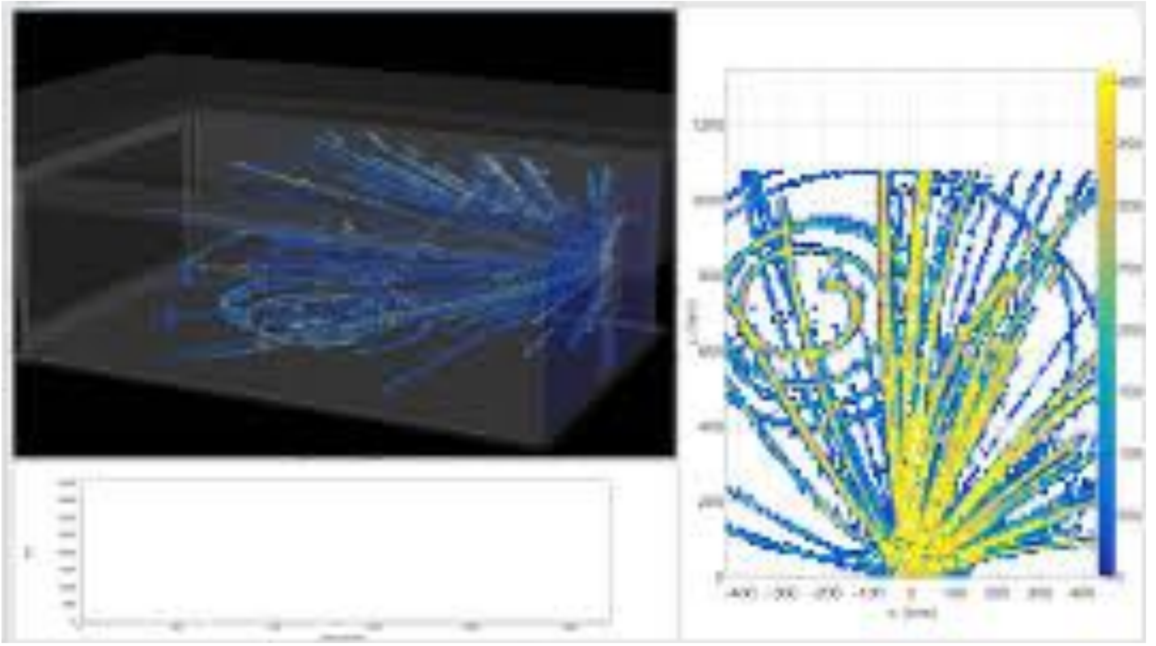
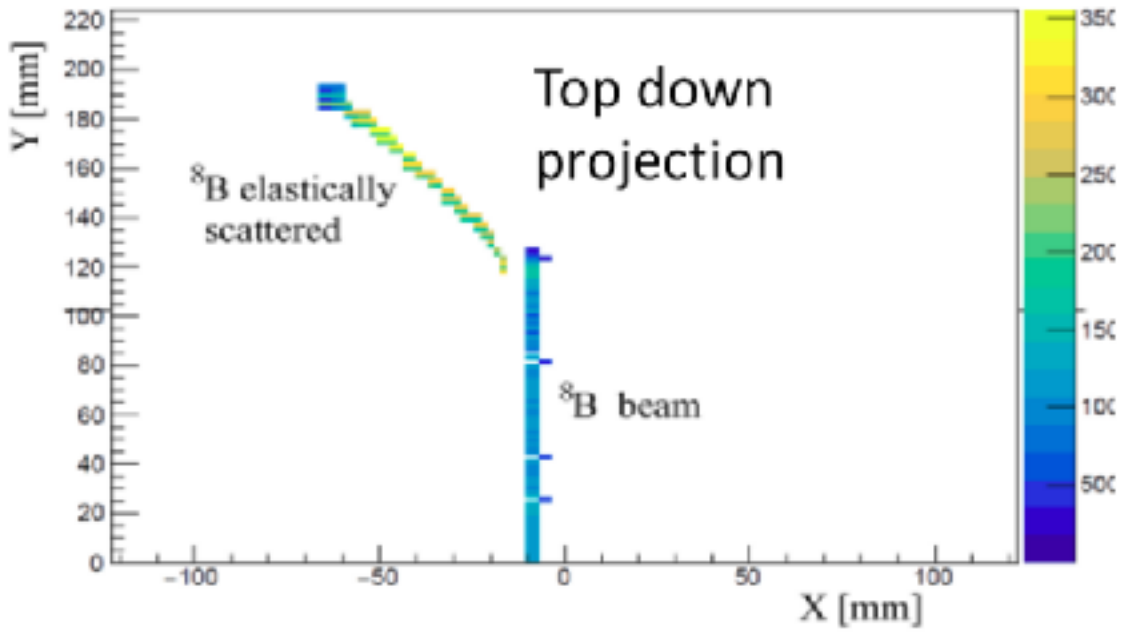
CERN

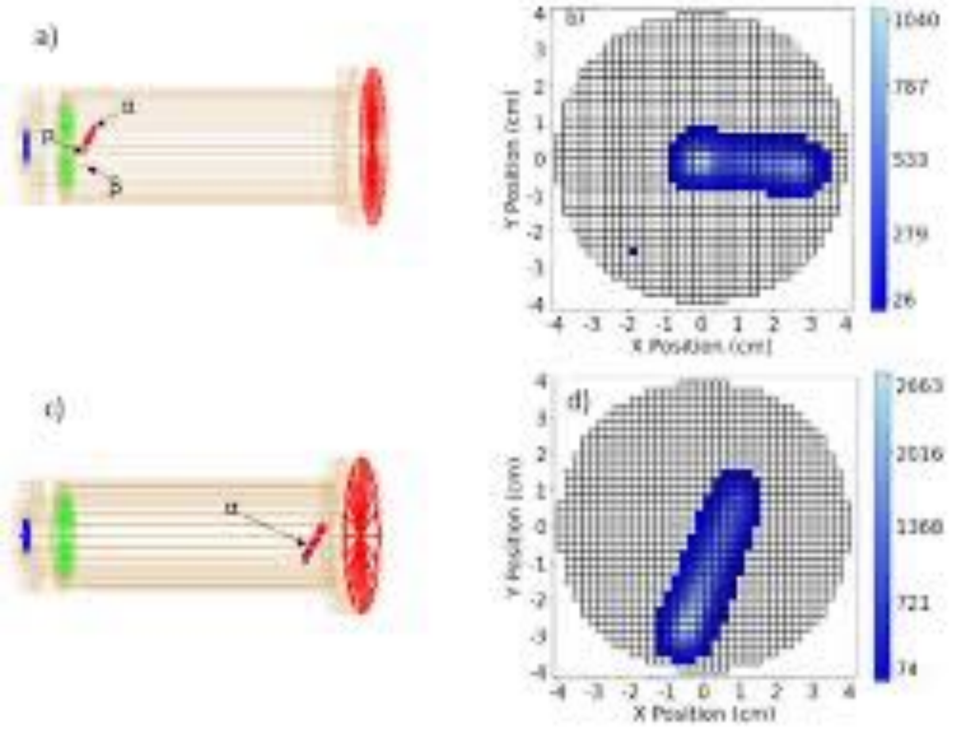
AT-TPC

CLAS 12

CMS

A POINT-CLOUD FOUNDATION MODEL FOR TIME PROJECTION CHAMBERS



ALPhA

A POINT-CLOUD FOUNDATION MODEL FOR TIME PROJECTION CHAMBERS

ELSEVIER
Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment
Volume 1072, March 2025, 170002

Full Length Article

Point-cloud based machine learning for classifying rare events in the Active-Target Time Projection Chamber

Poulomi Dey [a b], Adam K. Anthony [a c d], Curtis Hunt [a], Michelle P. Kuchera [e f], Raghuram Ramanujan [f], William G. Lynch [a c], ManYee Betty Tsang [a c], Joseph M. Wieske [a c], Jessica W. Ajongbah [a b], Saul Beceiro-Novo [a], Kyle W. Brown [a h], Zbigniew Chajecki [g], Kaitlin J. Cook [a h], Skyler Gangestad [d], Tom Ginter [a], Bergen Kendziorski [a], Fanurs Chi Eh Teh [a], HoTing Wong [i g]

ELSEVIER
Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment
Volume 1010, 11 September 2021, 165461

Unsupervised learning for identifying events in active target experiments

R. Solli [a b], D. Bazin [c], M. Hjorth-Jensen [c d], M.P. Kuchera [e], R.R. Strauss [f g]

ELSEVIER
Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment
Volume 940, 1 October 2019, Pages 156-167

Machine learning methods for track classification in the AT-TPC

M.P. Kuchera [a], R. Ramanujan [b], J.Z. Taylor [a], R.R. Strauss [b], D. Bazin [c], J. Bradt [c], Ruiming Chen [a]

Unpaired Translation of Point Clouds for Modeling Detector Response

Mingyang Li[1] Michelle Kuchera[1] Raghuram Ramanujan[1]
Adam Anthony[2] Curtis Hunt[3] Yassid Ayyad[4]
[1] Davidson College [2] High Point University
[3] Michigan State University [4] University of Santiago de Compostela
{mili, mikuchera, raramanujan}@davidson.edu
{aanthon2}@highpoint.edu
{huntc}@frib.msu.edu
{yassid.ayyad}@usc.es

ELSEVIER
Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment
Volume 1080, November 2025, 170659

Full Length Article

Object detection with deep learning for rare event search in the GADGET II TPC

Tyler Wheeler [a b c], S. Ravishankar [c d], C. Wrede [b a], A. Andalib [b a], A. Anthony [b a s], Y. Ayyad [a e], B. Jain [b a], A. Jaros [b a], R. Mahajan [a], L. Schaedig [b a], A. Adams [b a], S. Ahn [f], J.M. Allmond [m], D. Bardayan [g], D. Bazin [a], K. Bosmpotinis [b a], T. Budner [h], S.R. Carmichael [g], S.M. Cha [f], A. Chen [i]...L.E. Weghorn [b a]

LECTURE 1 TOPICS

- Computational graphs
- Gradient-descent optimization
- Logistic regression
- Regression neural networks

MICHELLE KUCHERA
DAVIDSON COLLEGE

CPS-FR
MIT
22 AUGUST 2024

GOALS

- Each of us learns something today
- Stop me with any questions

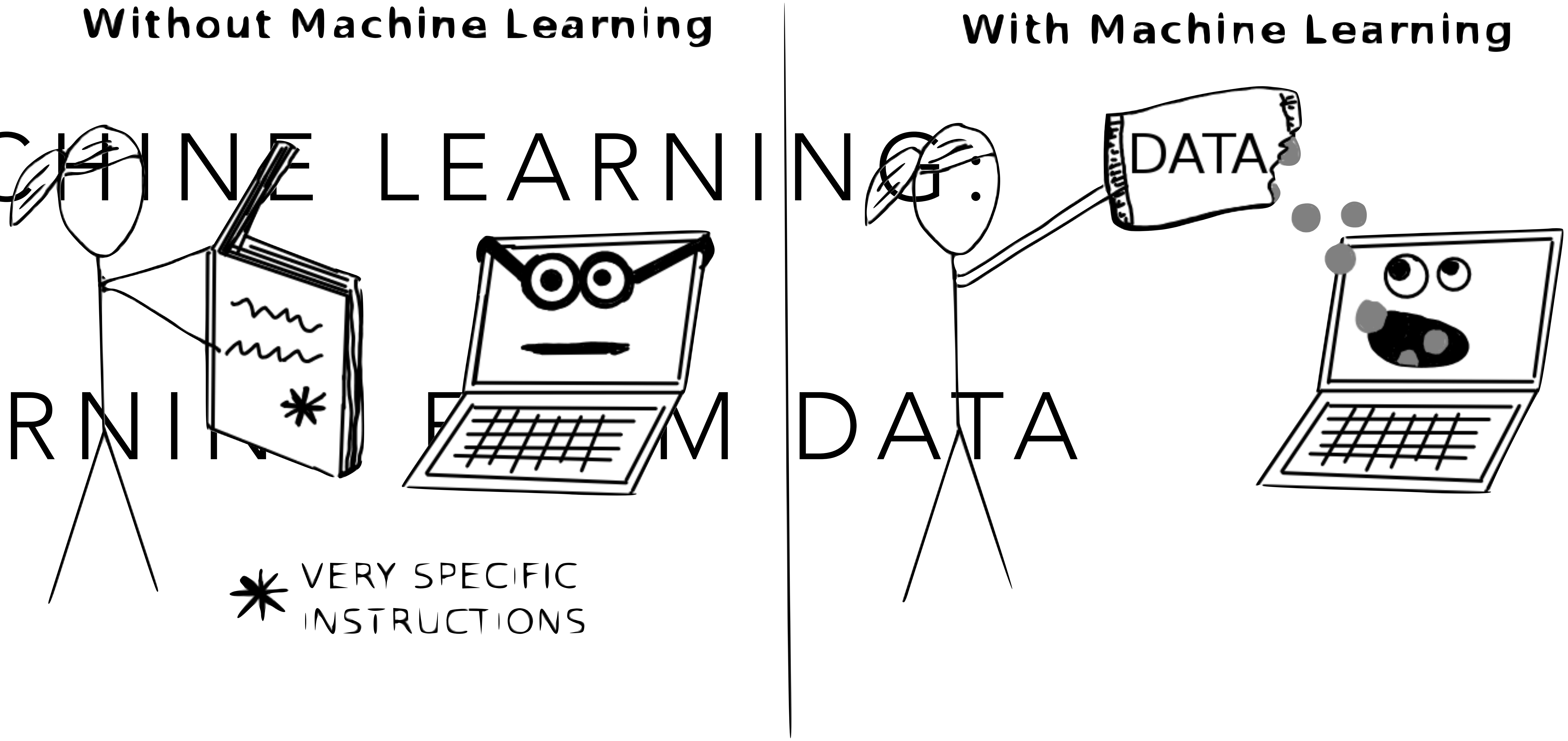
MICHELLE KUCHERA
DAVIDSON COLLEGE

CPS-FR
MIT
22 AUGUST 2024

COMMUNITY

- Each of you arrived here with your own backgrounds, specialty, and path in life
- Your experience and expertise are valuable here, no matter what it is
- If the activity is within your background, help others!
- If you are totally (or a little) lost, ask for help!
- It is our shared goal to have **each** of us leave with some new skill/knowledge/understanding

MACHINE LEARNING:
LEARNING FROM DATA

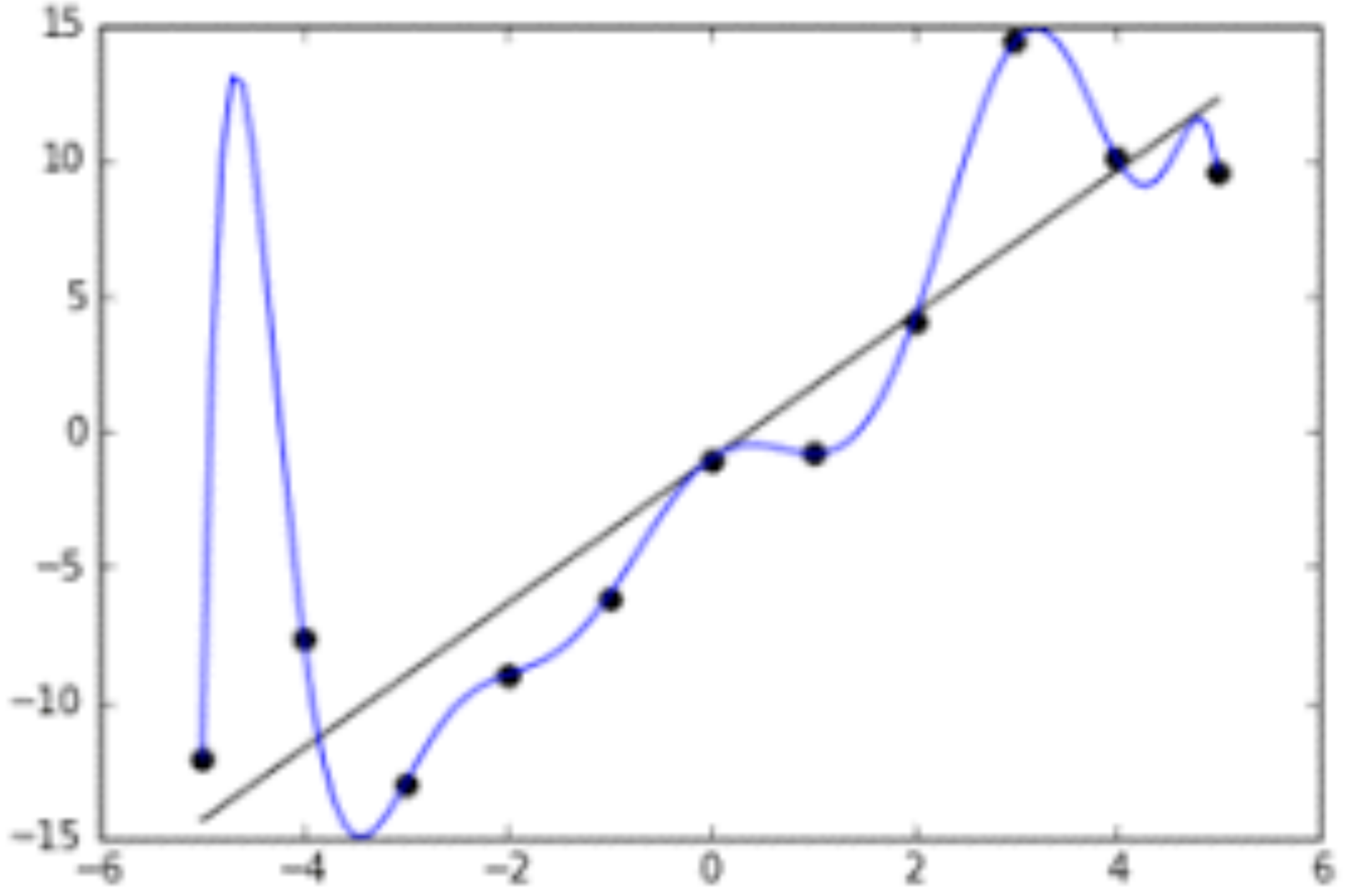


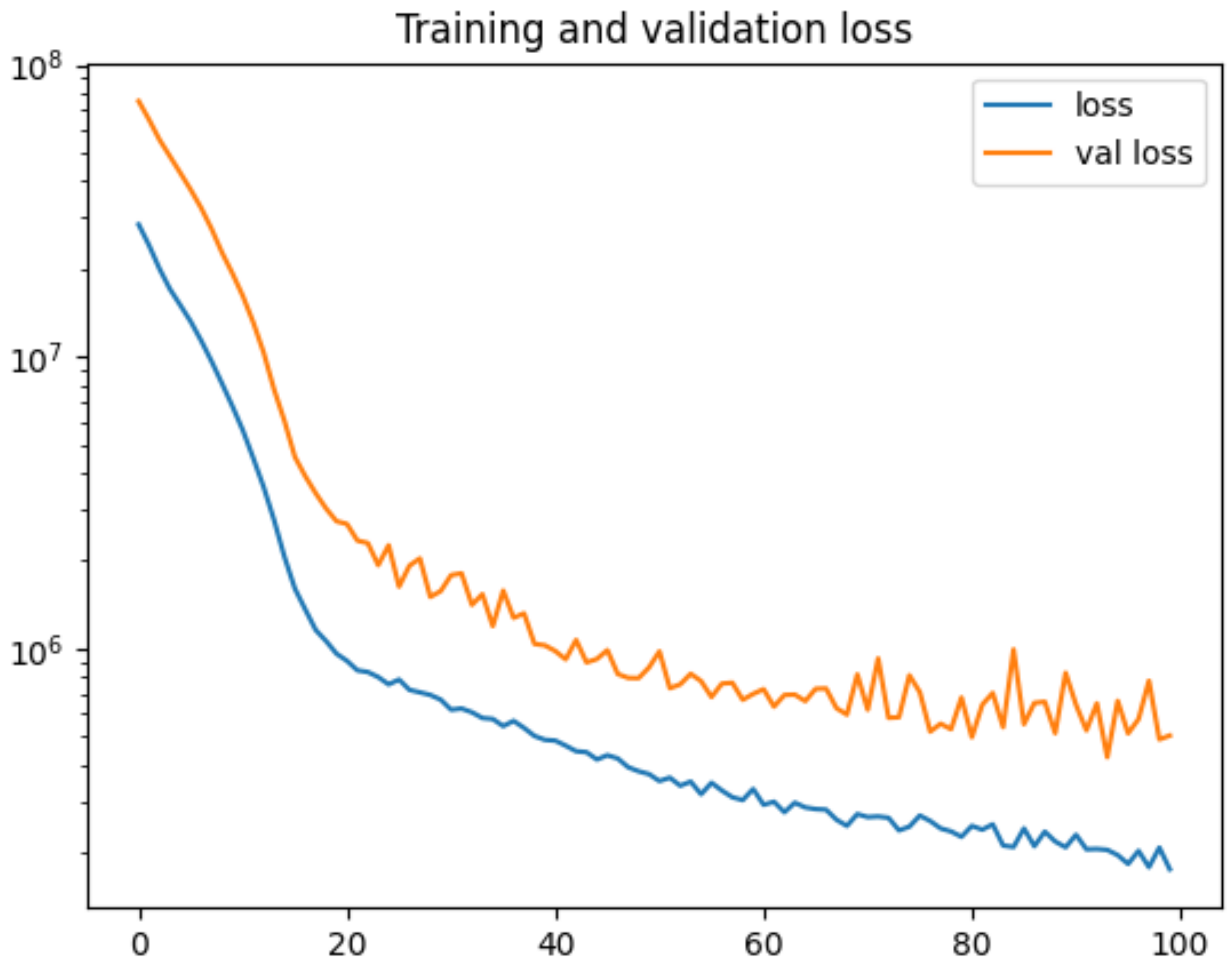
Without Machine Learning

With Machine Learning

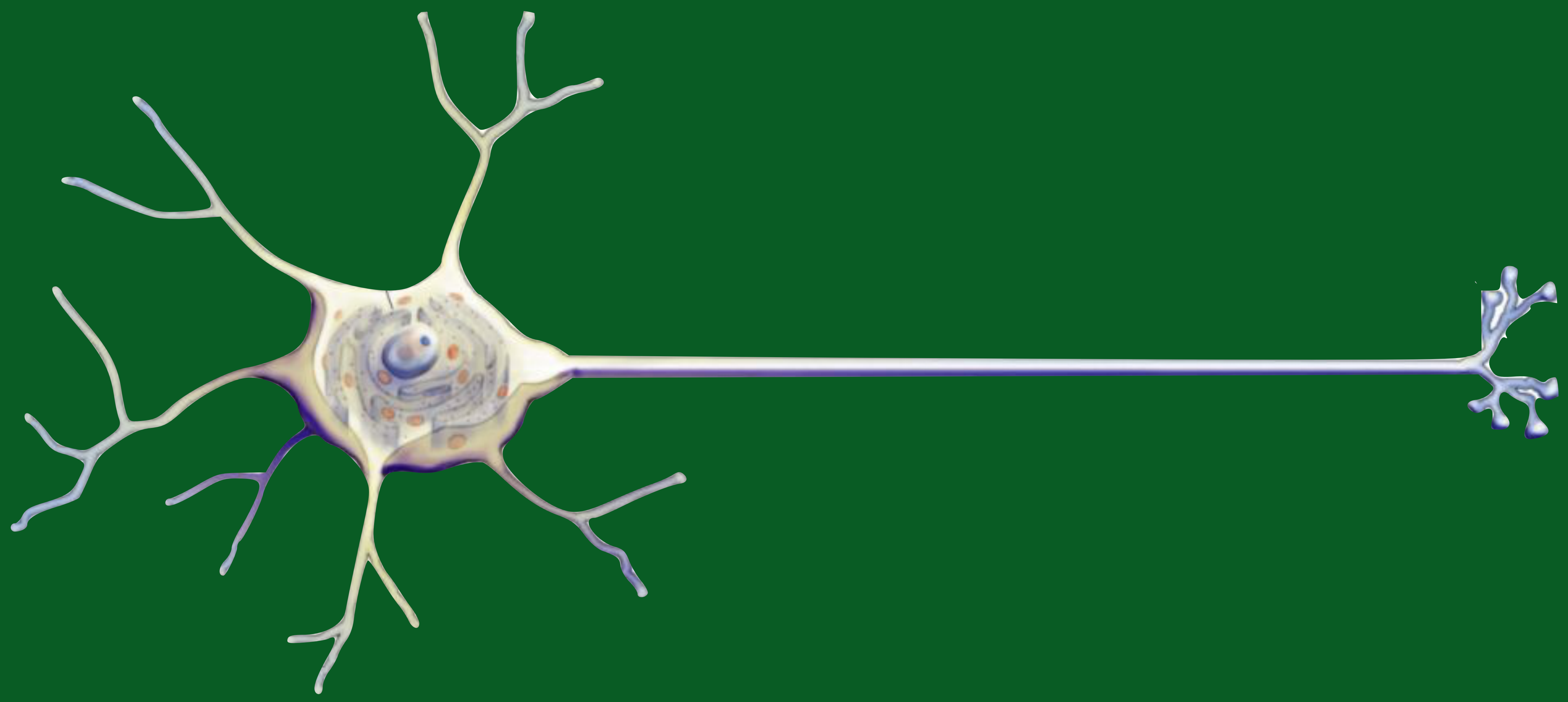
Learning from data was a **paradigm shift** in thinking about predictive models

✱ VERY SPECIFIC INSTRUCTIONS





NEURON



MATHEMATICS

ELSEVIER

Neural Networks
Volume 4, Issue 2, 1991, Pages 251-257

Approximation capabilities of multilayer feedforward networks

Kurt Hornik

Show more

Share Cite

https://doi.org/10.1016/0893-6080(91)90009-T

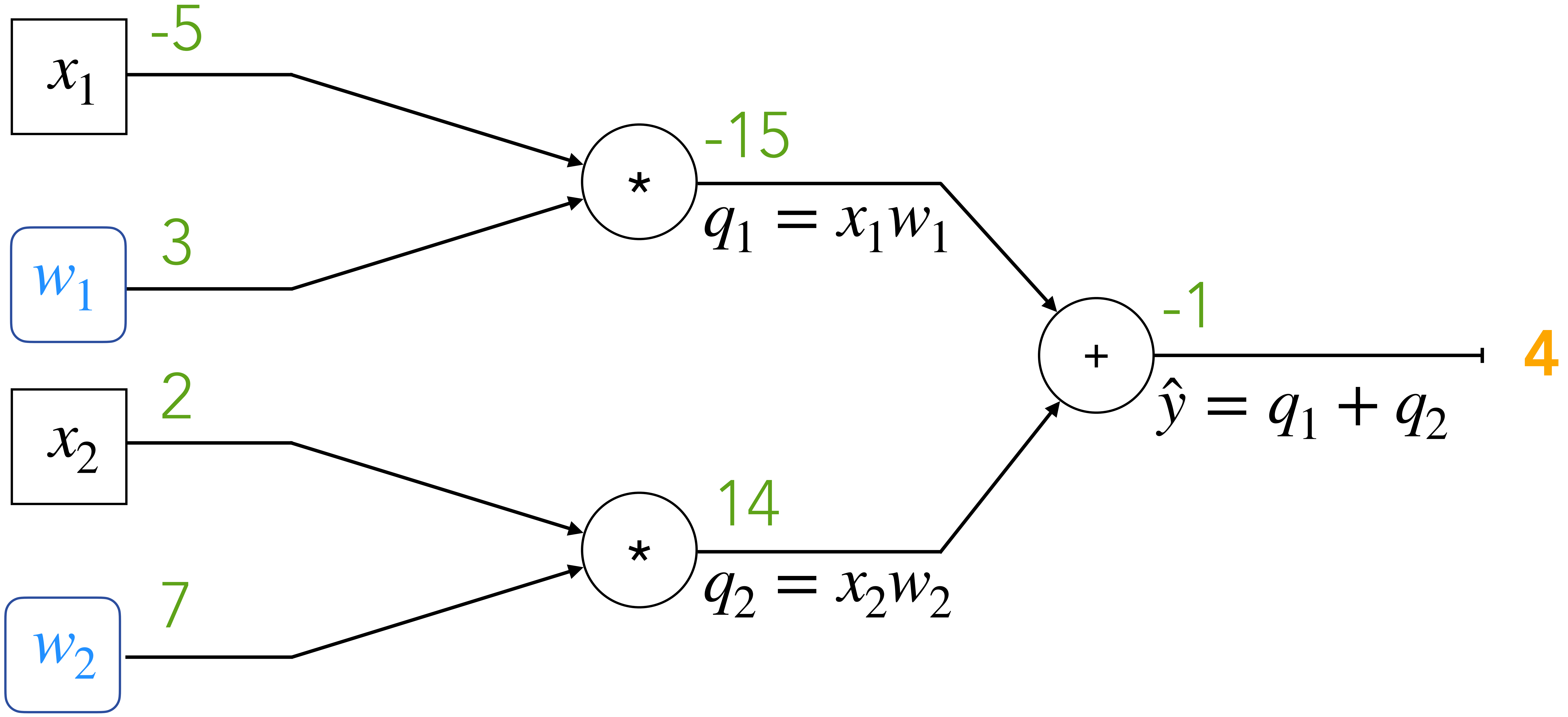
Abstract

We show that standard multilayer feedforward networks with as few as a single hidden layer and arbitrary bounded and nonconstant activation function are universal approximators with respect to $L^p(\mu)$ performance criteria, for arbitrary finite input environment measures μ, provided only that sufficiently many hidden units are available. If the activation function is continuous, bounded and nonconstant, then continuous mappings can be learned uniformly over compact input sets. We also give very general conditions ensuring that networks with sufficiently smooth activation functions are capable of arbitrarily accurate approximation to a function and its derivatives.

MATHEMATICS

COMPUTATIONAL GRAPH

$$\hat{y} = x_1 w_1 + x_2 w_2$$

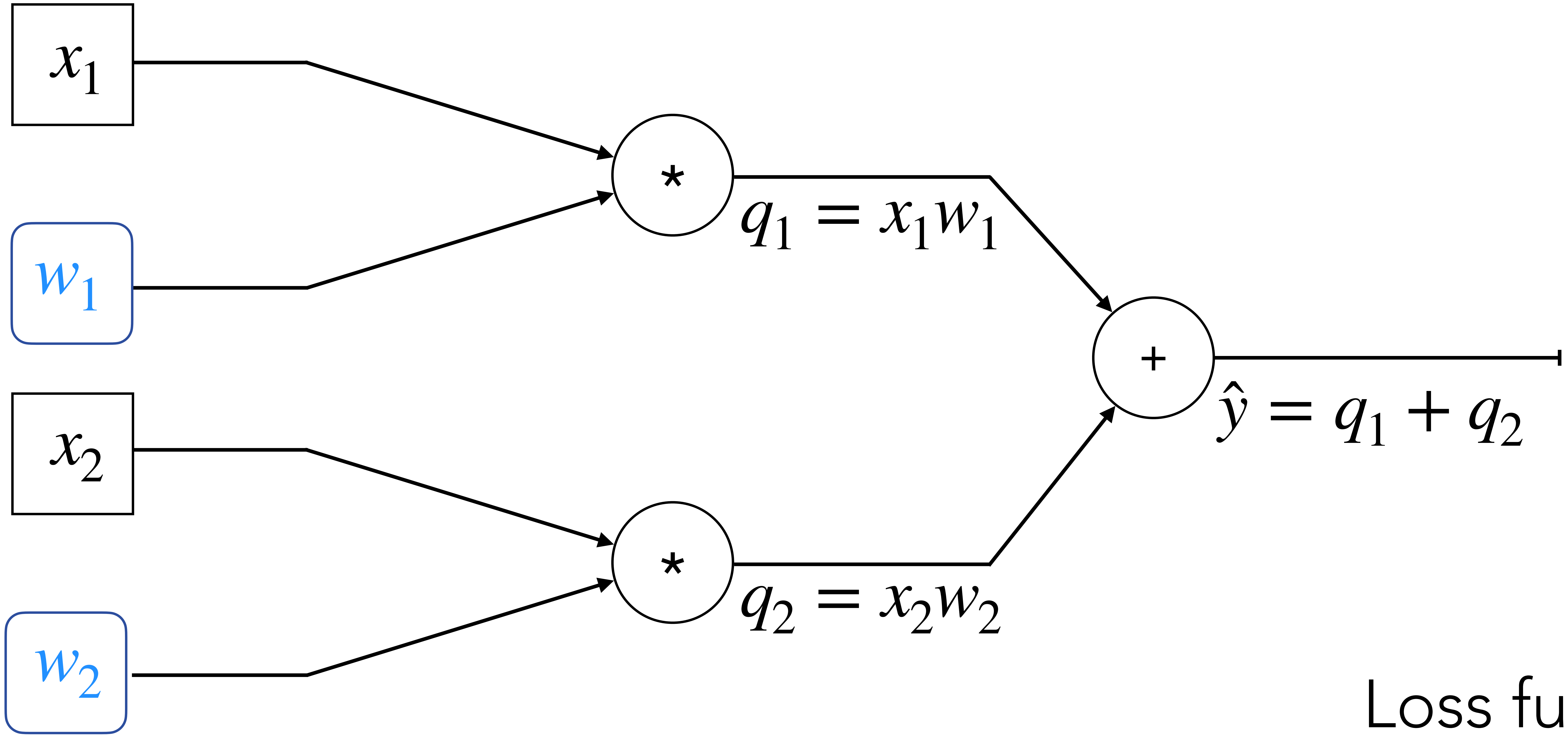


MACHINE LEARNING

SUPERVISED LEARNING

REGRESSION

$$\hat{y} = x_1 w_1 + x_2 w_2$$

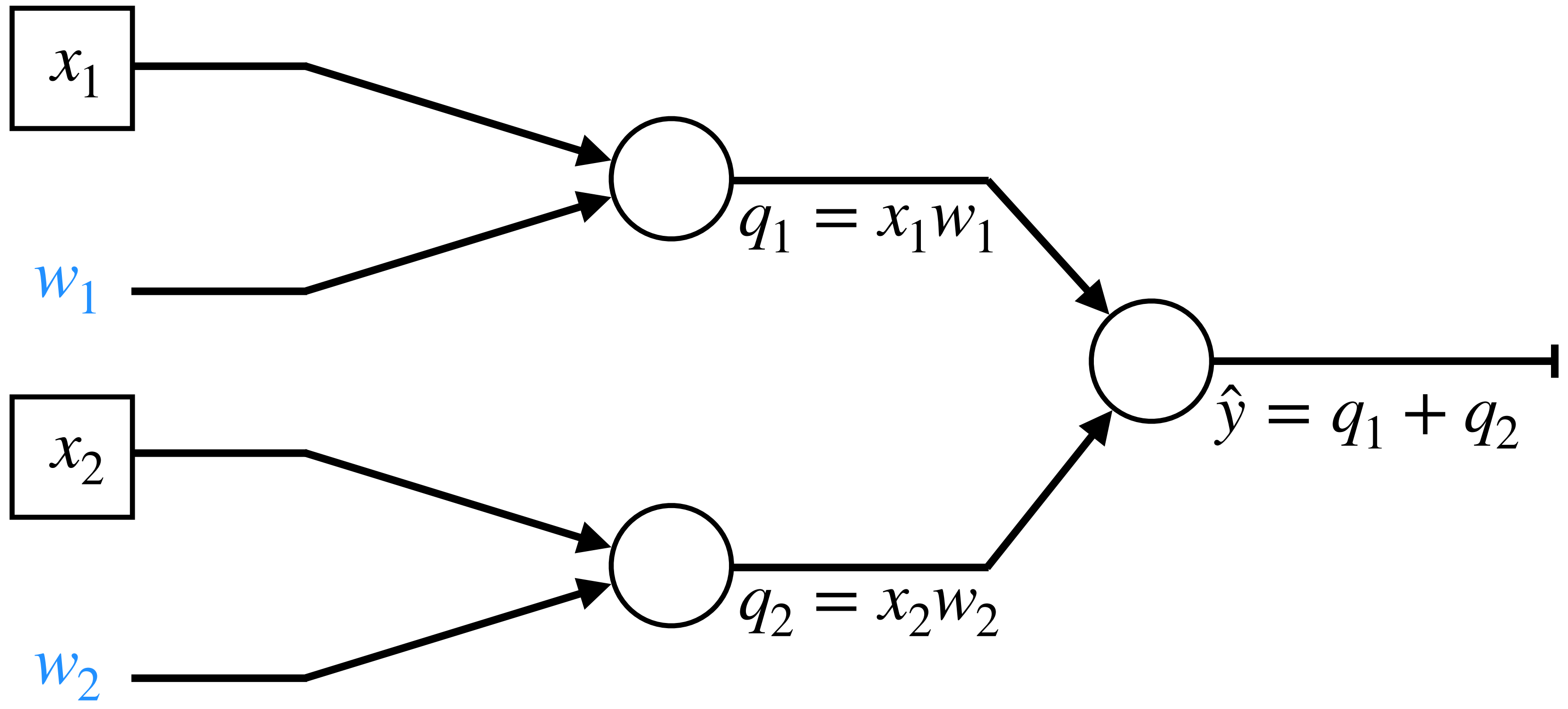


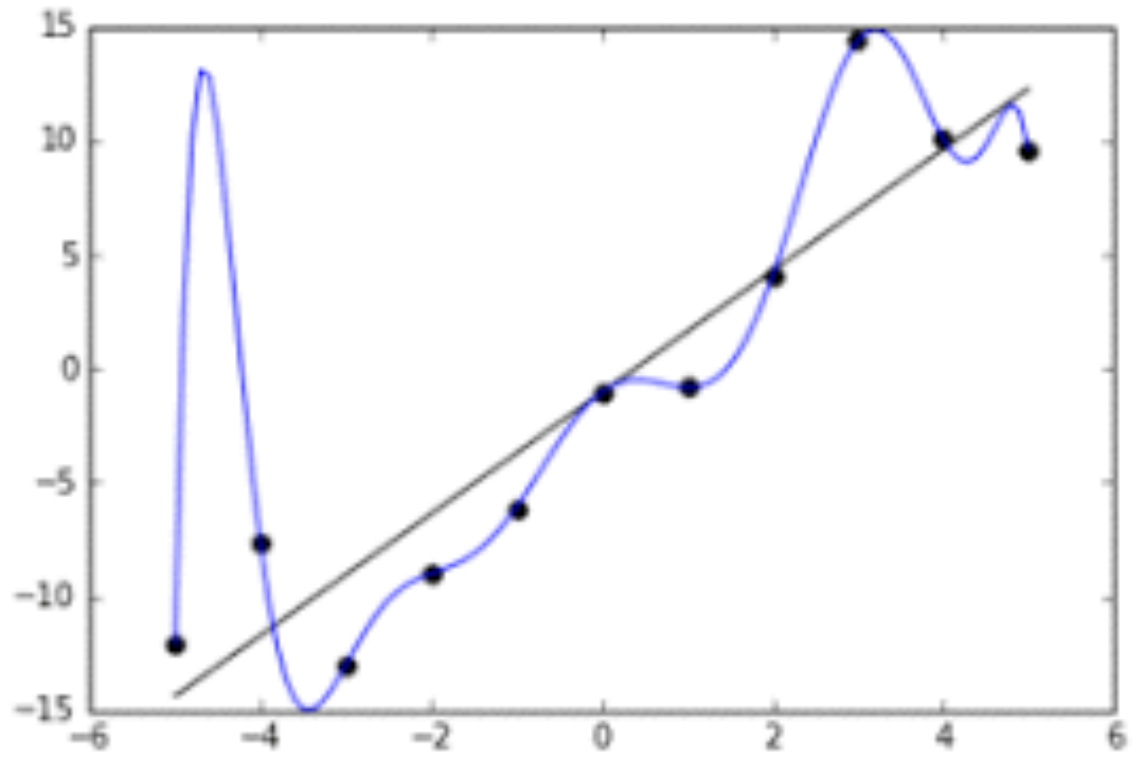
Loss function

$$J(w) = \hat{y} - y$$

SUPERVISED LEARNING

$$\hat{y} = x_1 w_1 + x_2 w_2$$





	Feature 1	Feature 2	Feature 3	Target
Example 1				
Example 2				
Example 3				
Example 4				

Loss function

MSE across N examples

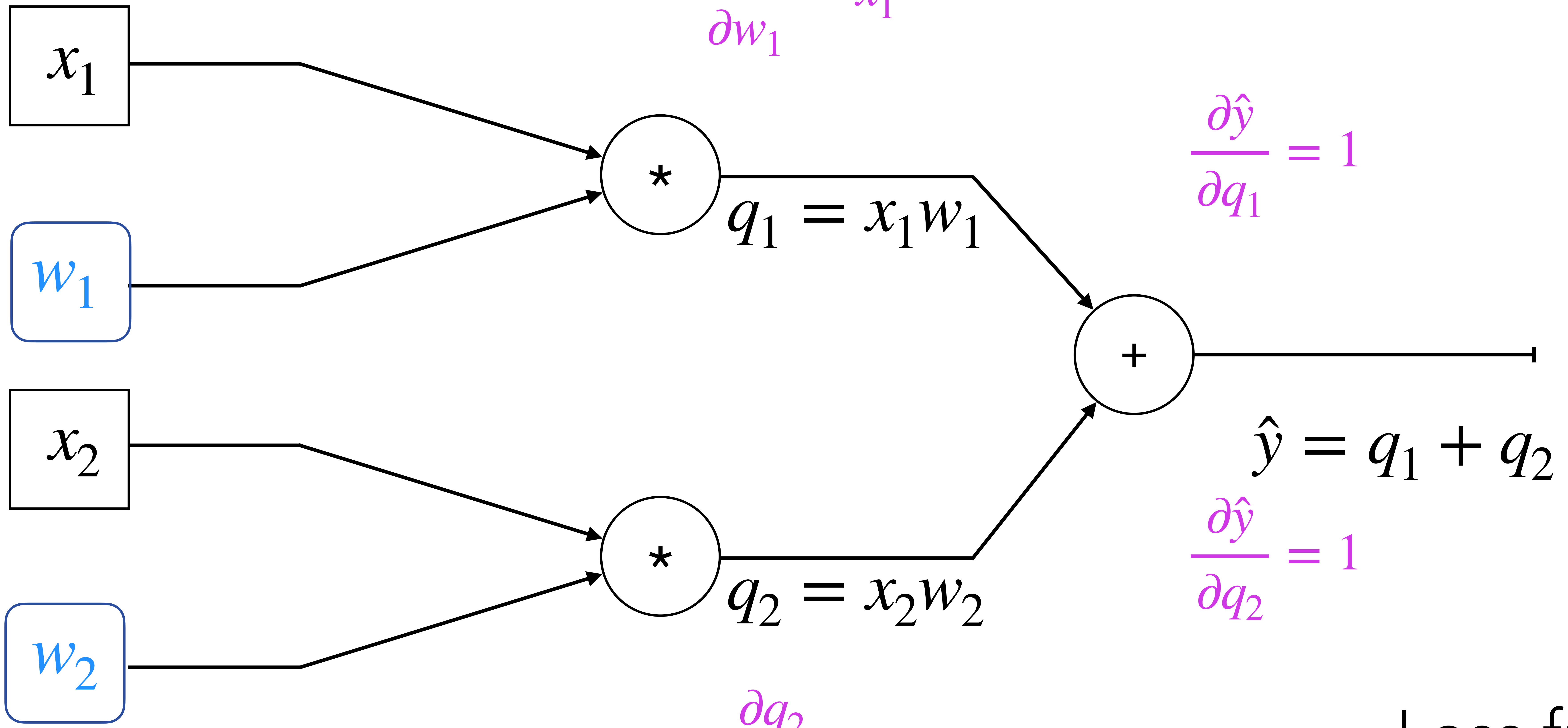
$$J(w) = \frac{1}{N} \sum_{i=0}^{N} (\hat{y}_i - y_i)^2$$

BACKPROPAGATION

$$w_1 = w_1 - \eta * \frac{\partial J}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial q_1} \frac{\partial q_1}{\partial w_1}$$

$$\frac{\partial q_1}{\partial w_1} = x_1$$

$$\frac{\partial \hat{y}}{\partial q_1} = 1$$



$$\frac{\partial J}{\partial \hat{y}} = 1$$

$$\frac{\partial \hat{y}}{\partial q_2} = 1$$

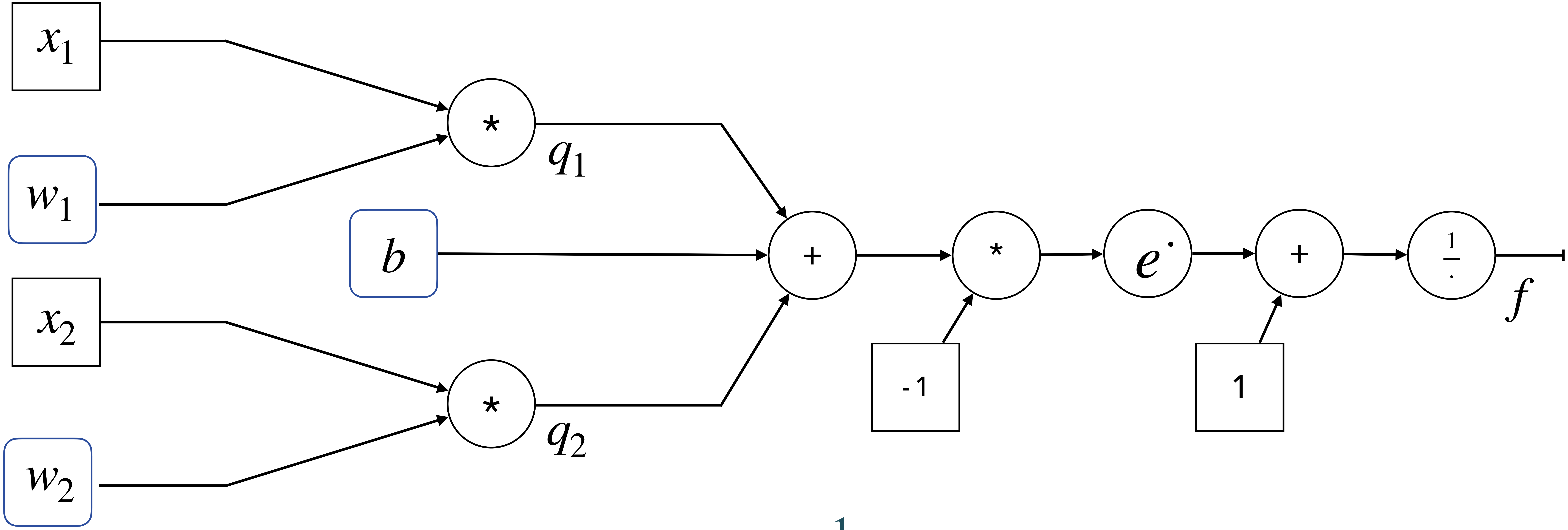
$$\frac{\partial q_2}{\partial w_2} = x_2$$

$$w_2 = w_2 - \eta * \frac{\partial J}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial q_2} \frac{\partial q_2}{\partial w_2}$$

Loss function

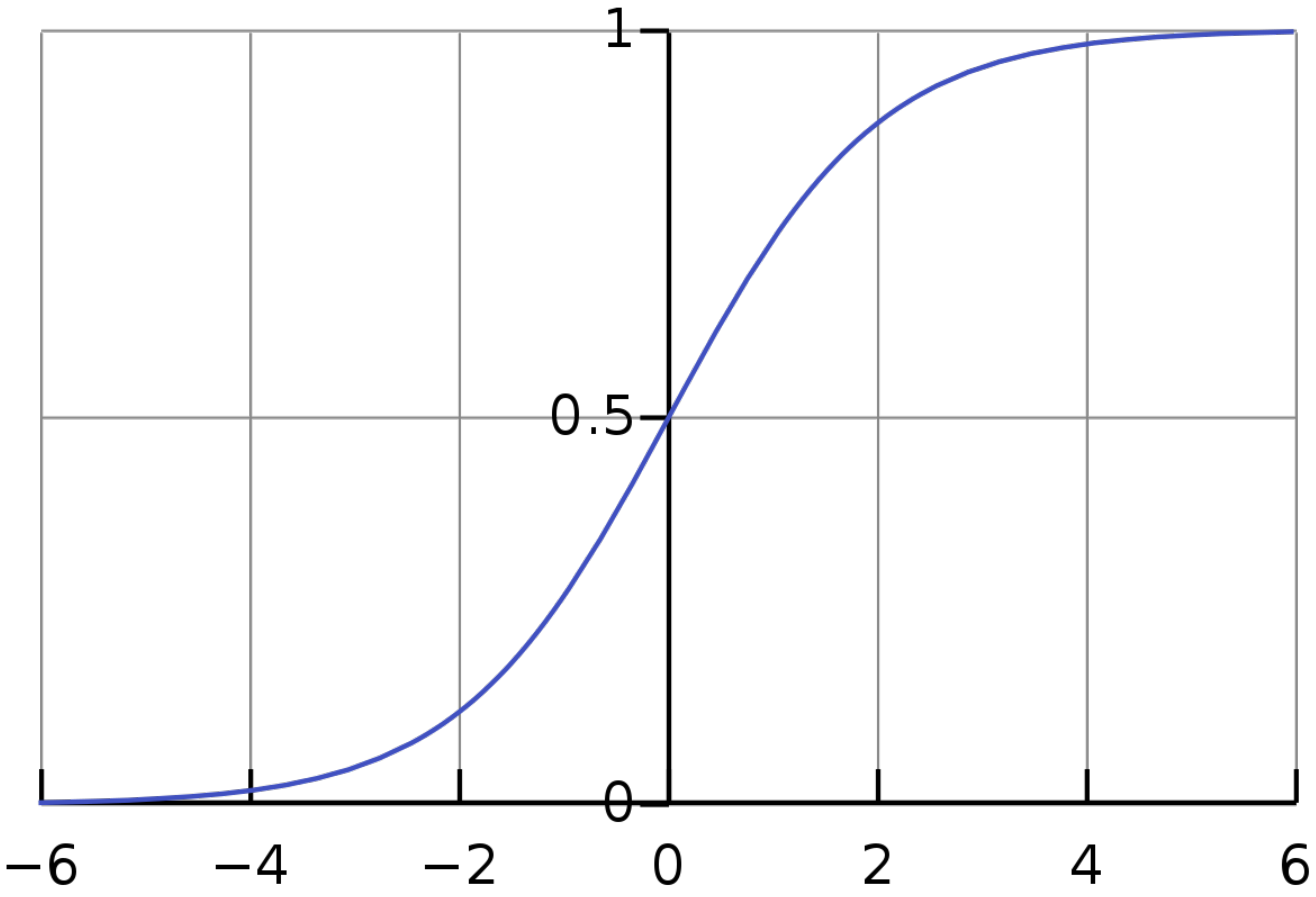
$$J(w) = \hat{y} - y$$

LOGISTIC REGRESSION

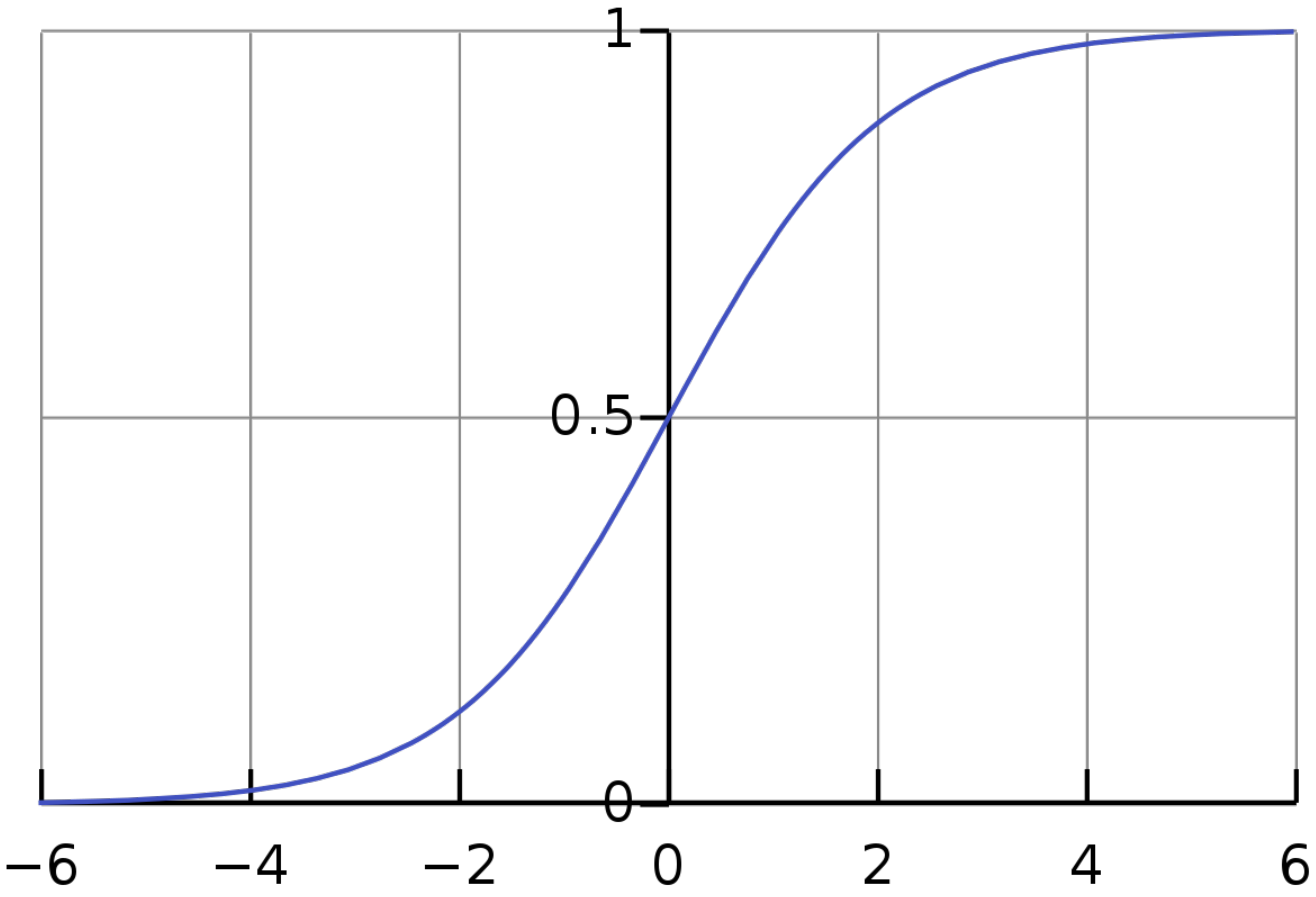


$$f = \frac{1}{1 + e^{-(x_1 w_1 + x_2 w_2 + b)}}$$

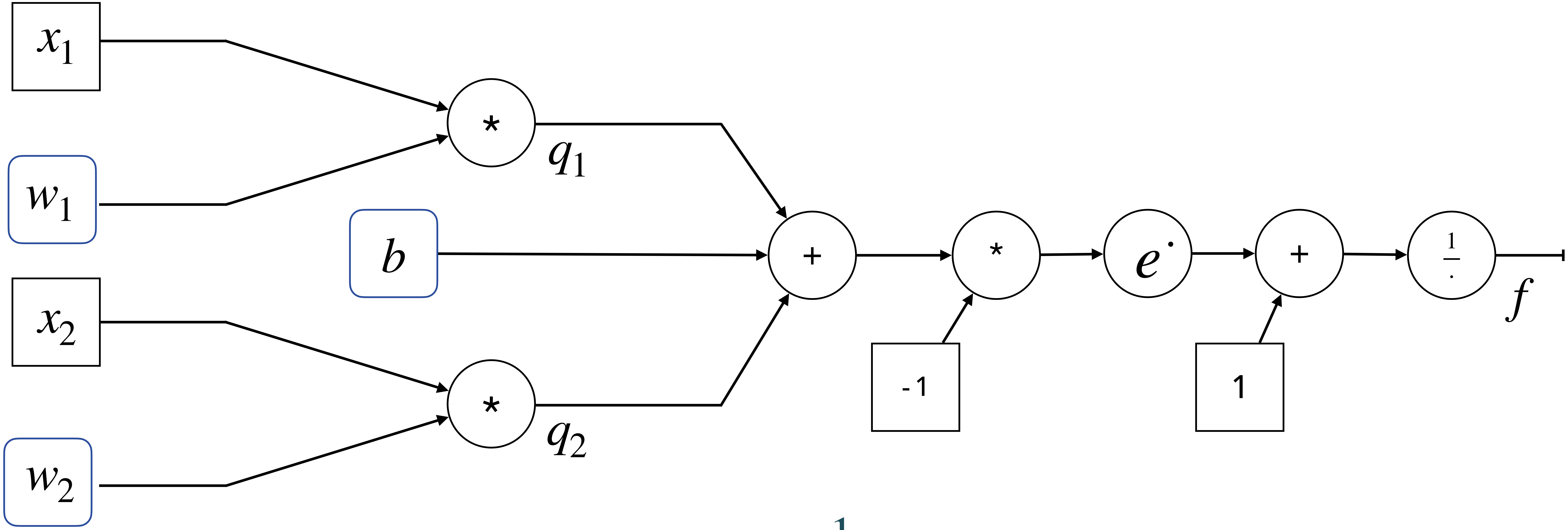
LOGISTIC REGRESSION



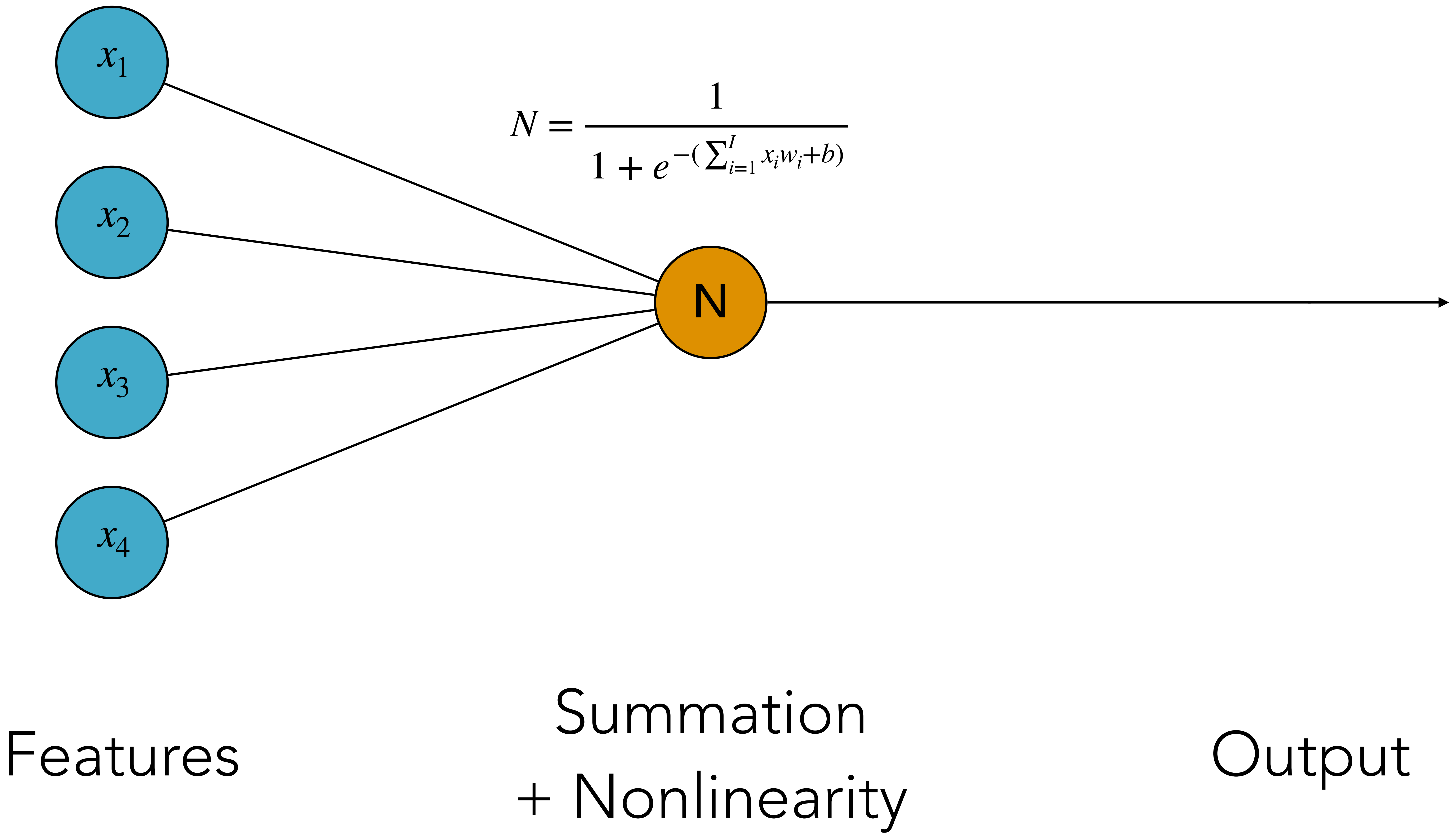
CLASSIFICATION



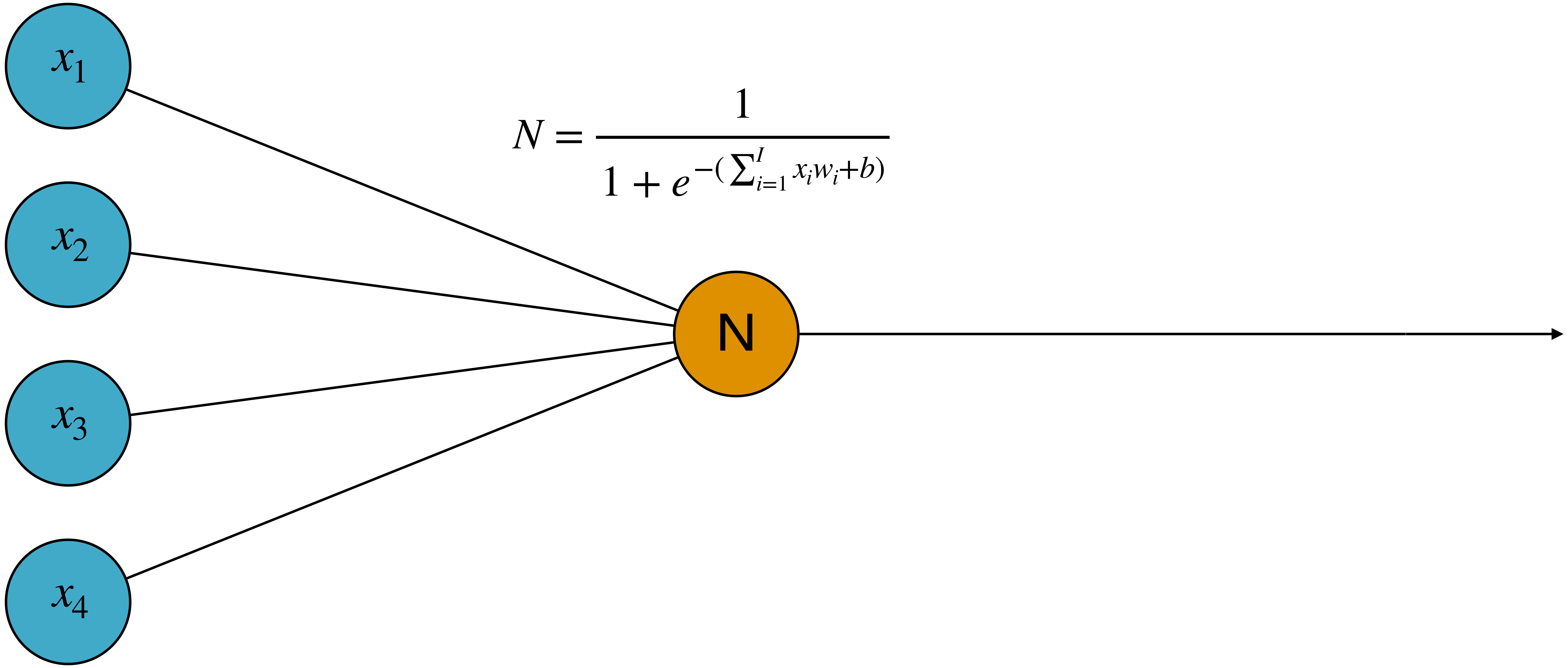
LOGISTIC REGRESSION



$$f = \frac{1}{1 + e^{-(x_1 w_1 + x_2 w_2 + b)}}$$


x_1
x_2
x_3
x_4
$N = \frac{1}{1 + e^{-(\sum_{i=1}^{I} x_i w_i + b)}}$
N
Features
Summation
+ Nonlinearity
Output

CHECK: HOW MANY TRAINABLE PARAMETERS?

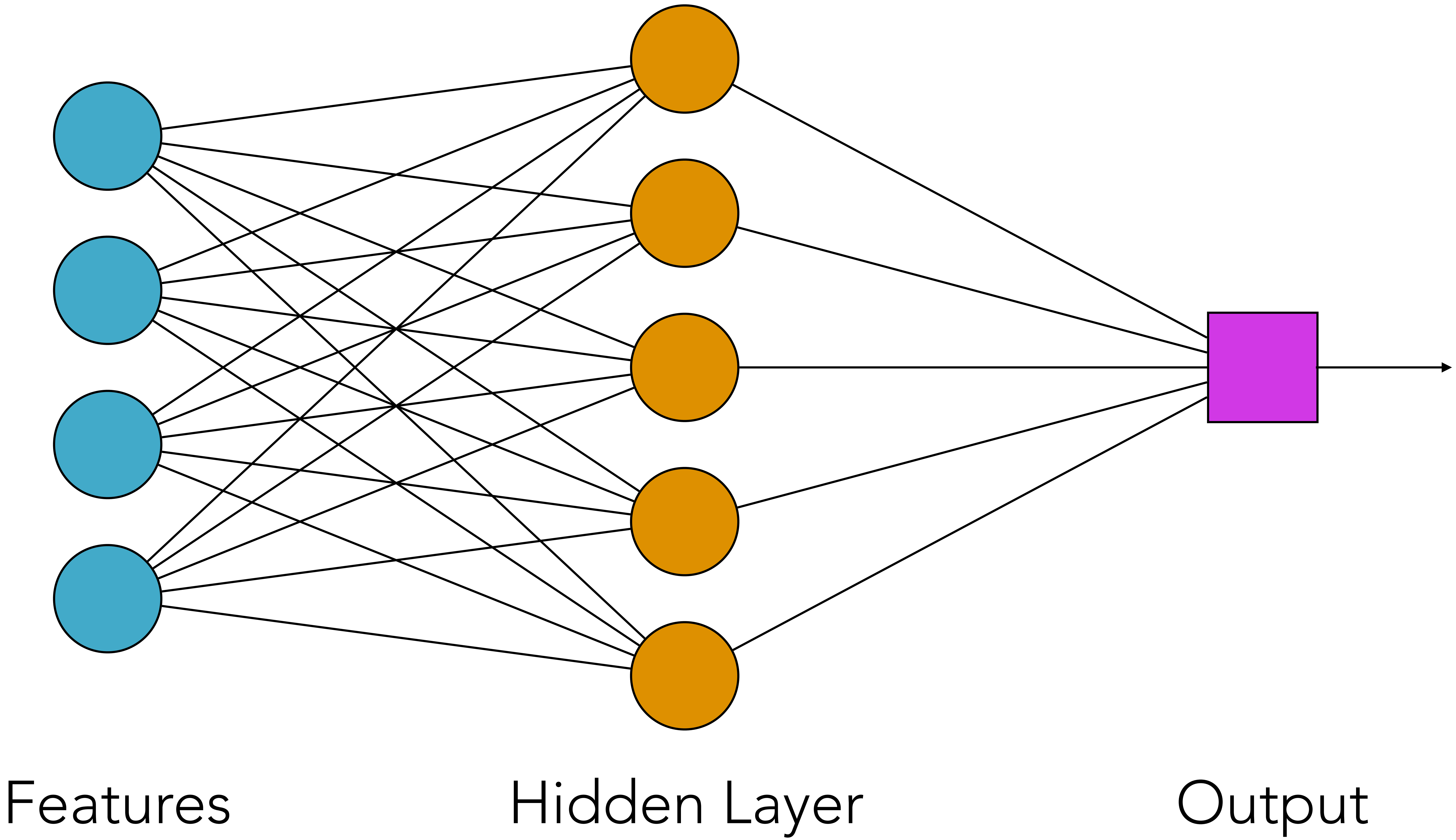


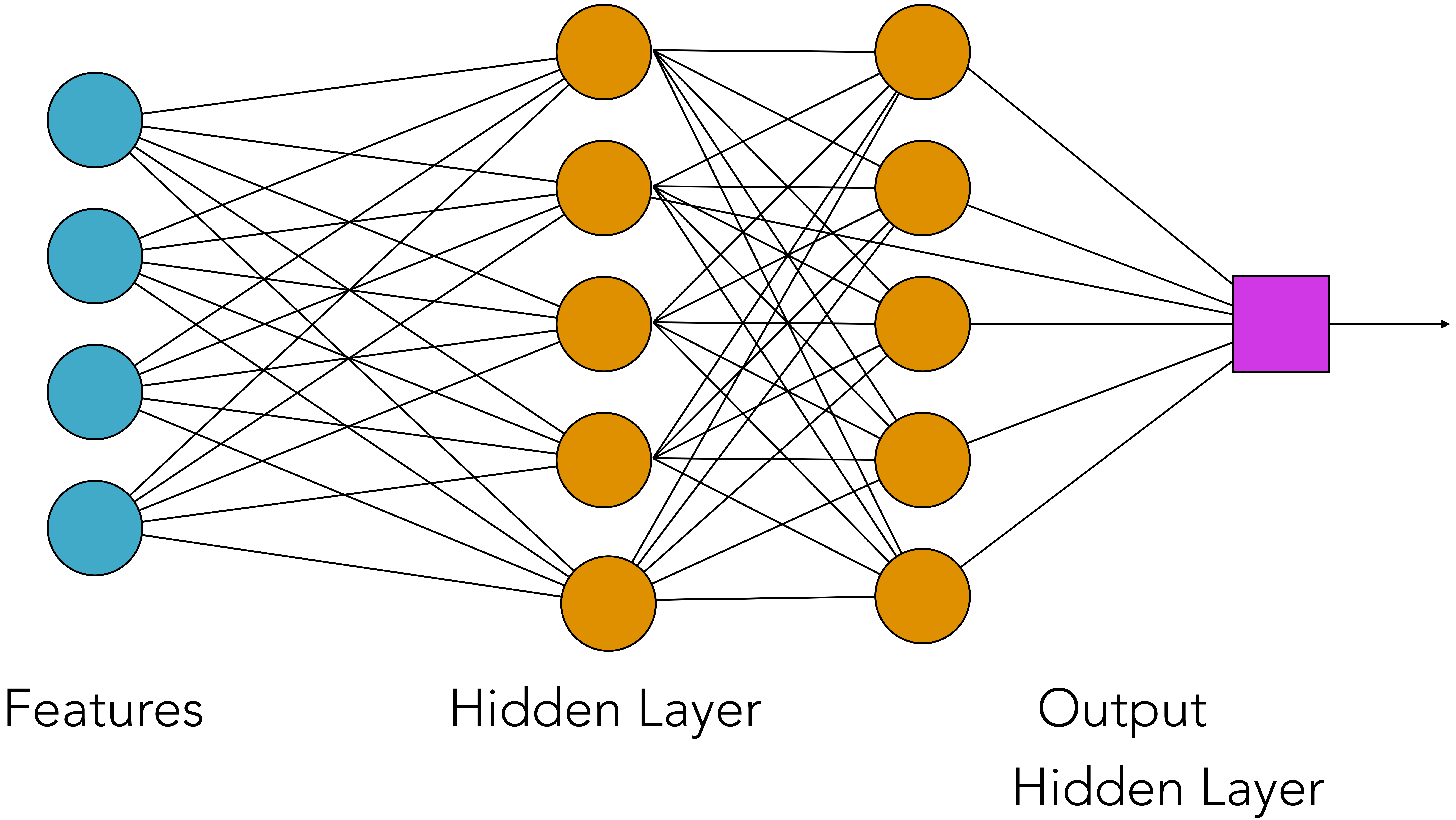
$$N = \frac{1}{1 + e^{-(\sum_{i=1}^{I} x_i w_i + b)}}$$

Features

Summation
+ Nonlinearity

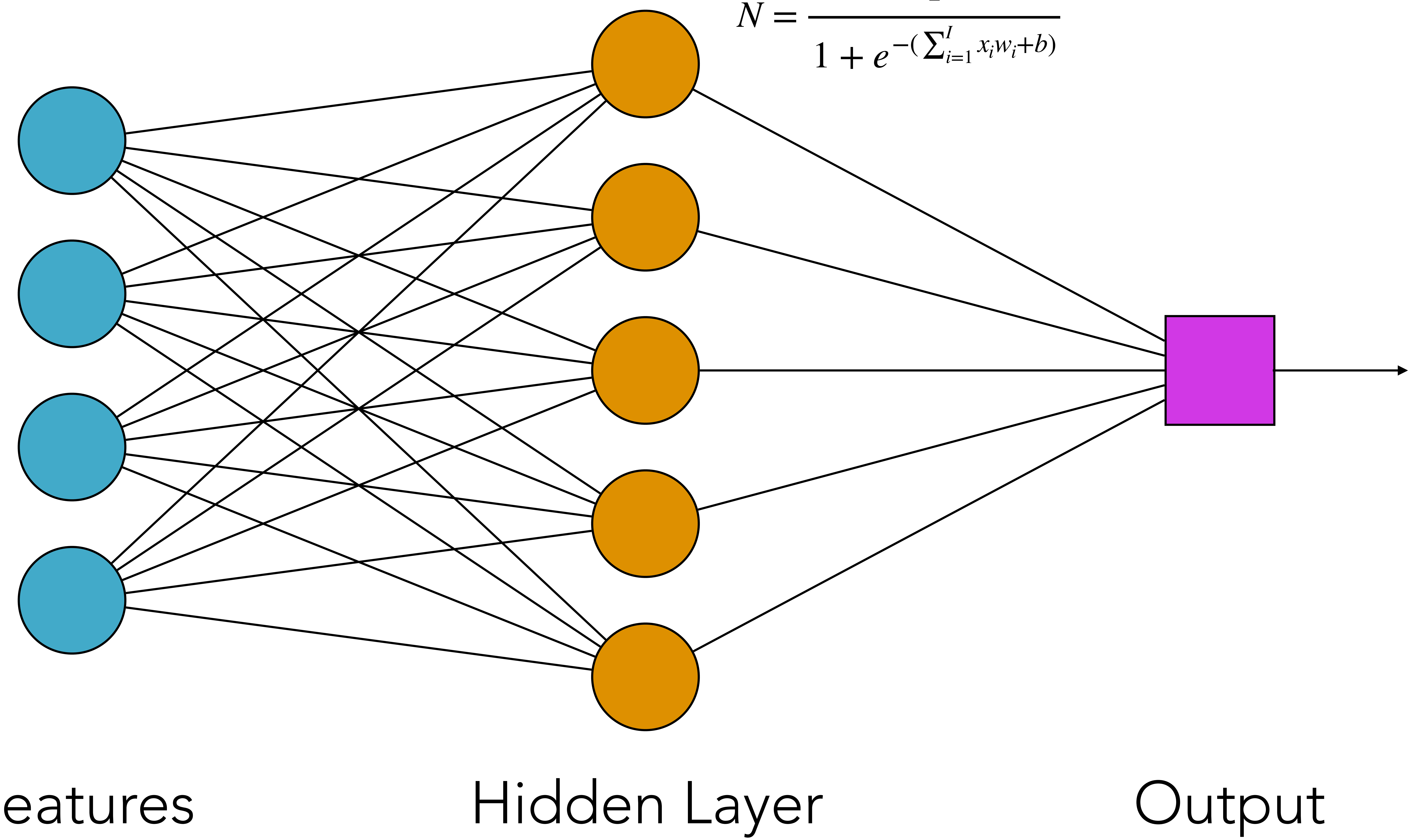
Output


Features
Hidden Layer
Output


Features
Hidden Layer
Output
Hidden Layer

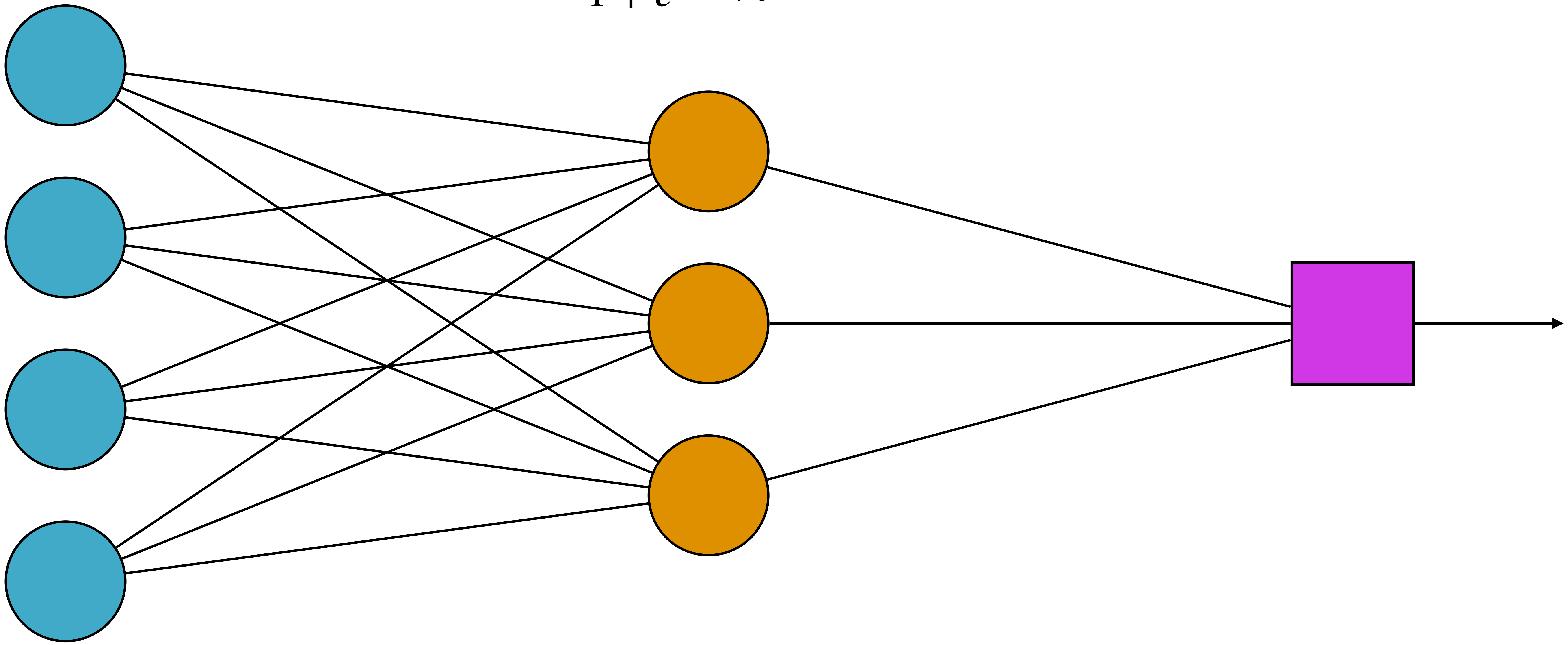
CHECK: HOW MANY TRAINABLE PARAMETERS?

$$N = \frac{1}{1 + e^{-(\sum_{i=1}^{I} x_i w_i + b)}}$$



CHECK: HOW MANY TRAINABLE PARAMETERS?

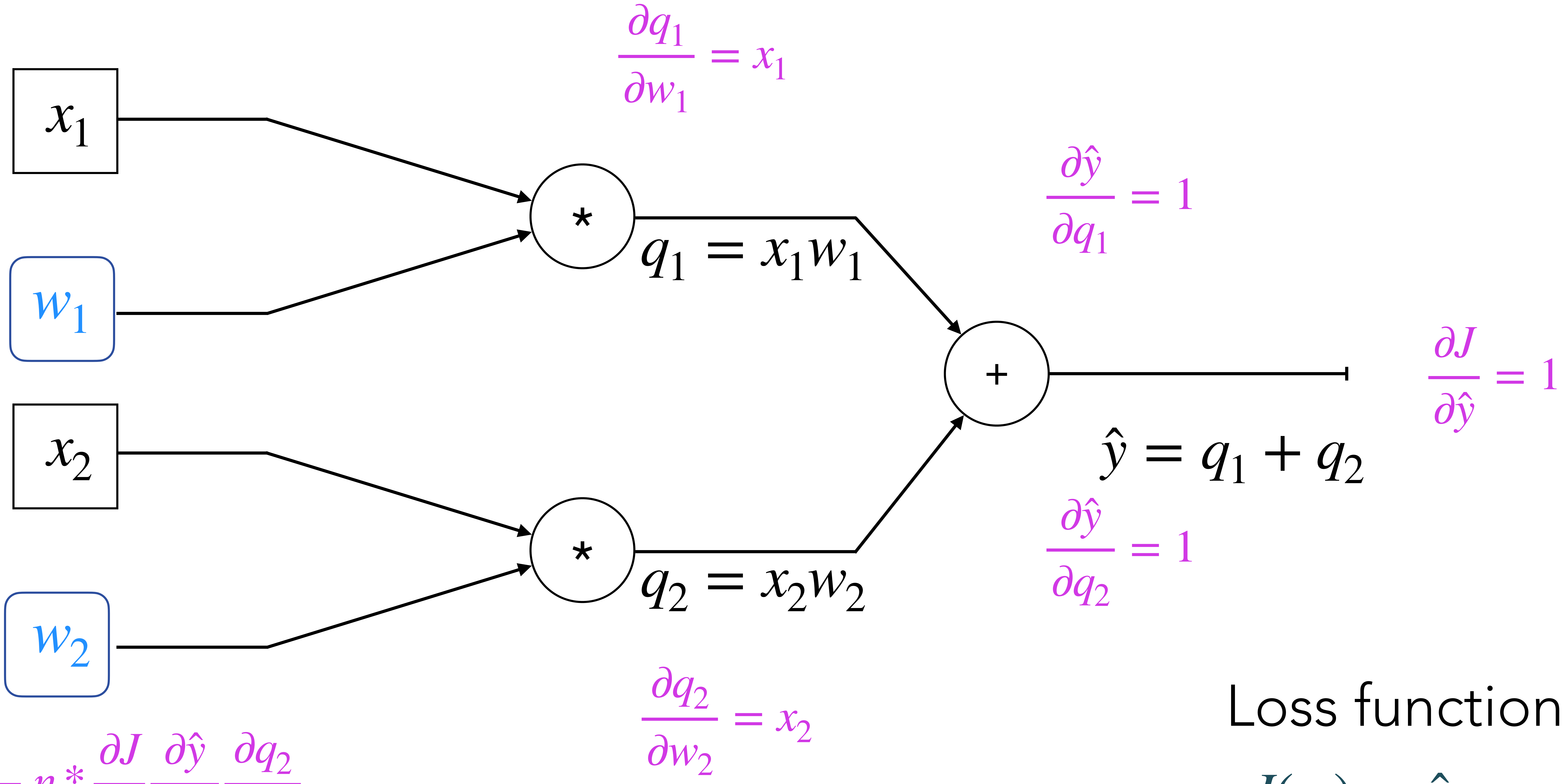
$$N = \frac{1}{1 + e^{-(\sum_{i=1}^{I} x_i w_i + b)}}$$



Features Hidden Layer Output

BACKPROPAGATION

$$w_1 = w_1 - \eta * \frac{\partial J}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial q_1} \frac{\partial q_1}{\partial w_1}$$

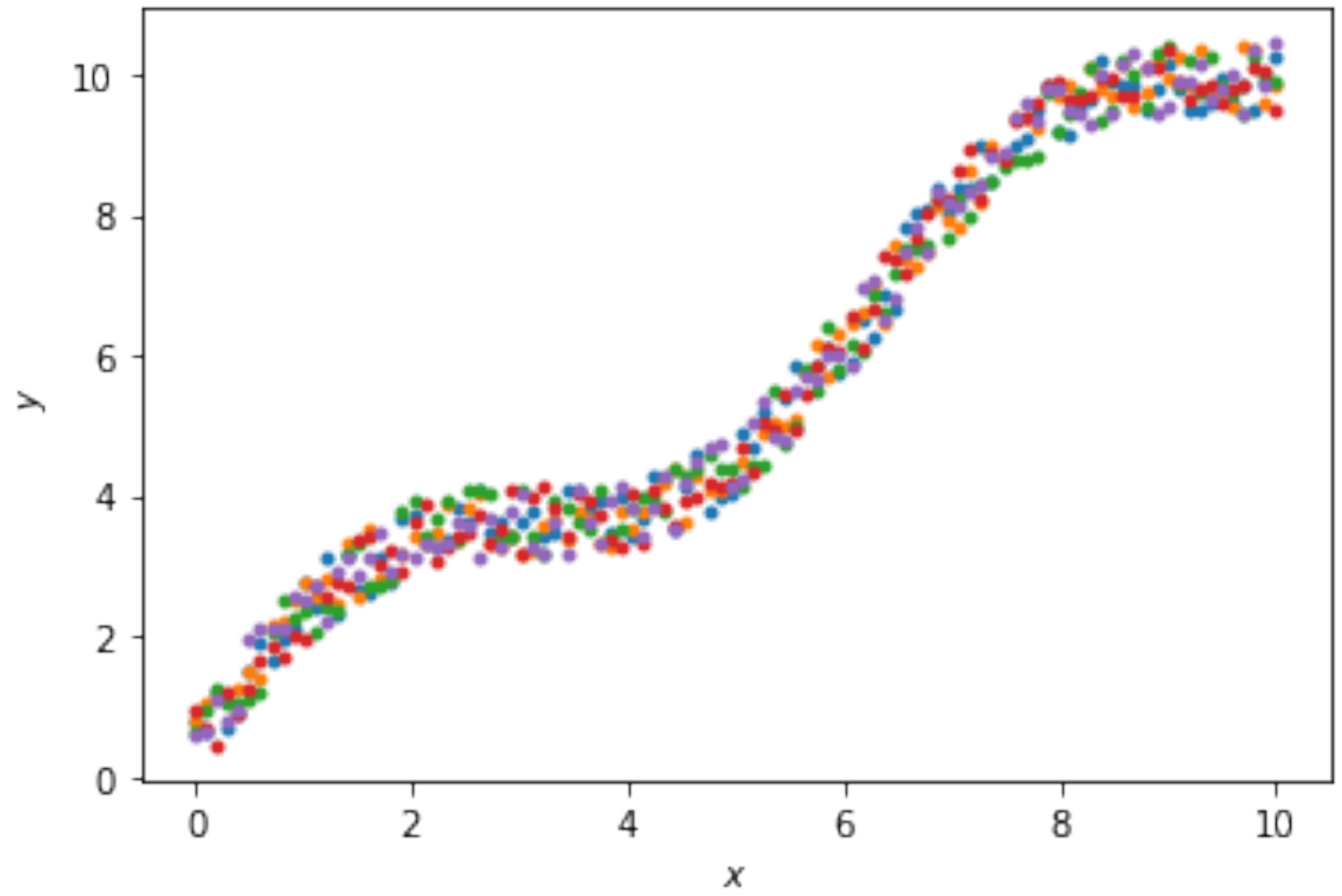


$$w_2 = w_2 - \eta * \frac{\partial J}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial q_2} \frac{\partial q_2}{\partial w_2}$$

Loss function

$$J(w) = \hat{y} - y$$

LOSS FUNCTIONS



Loss function

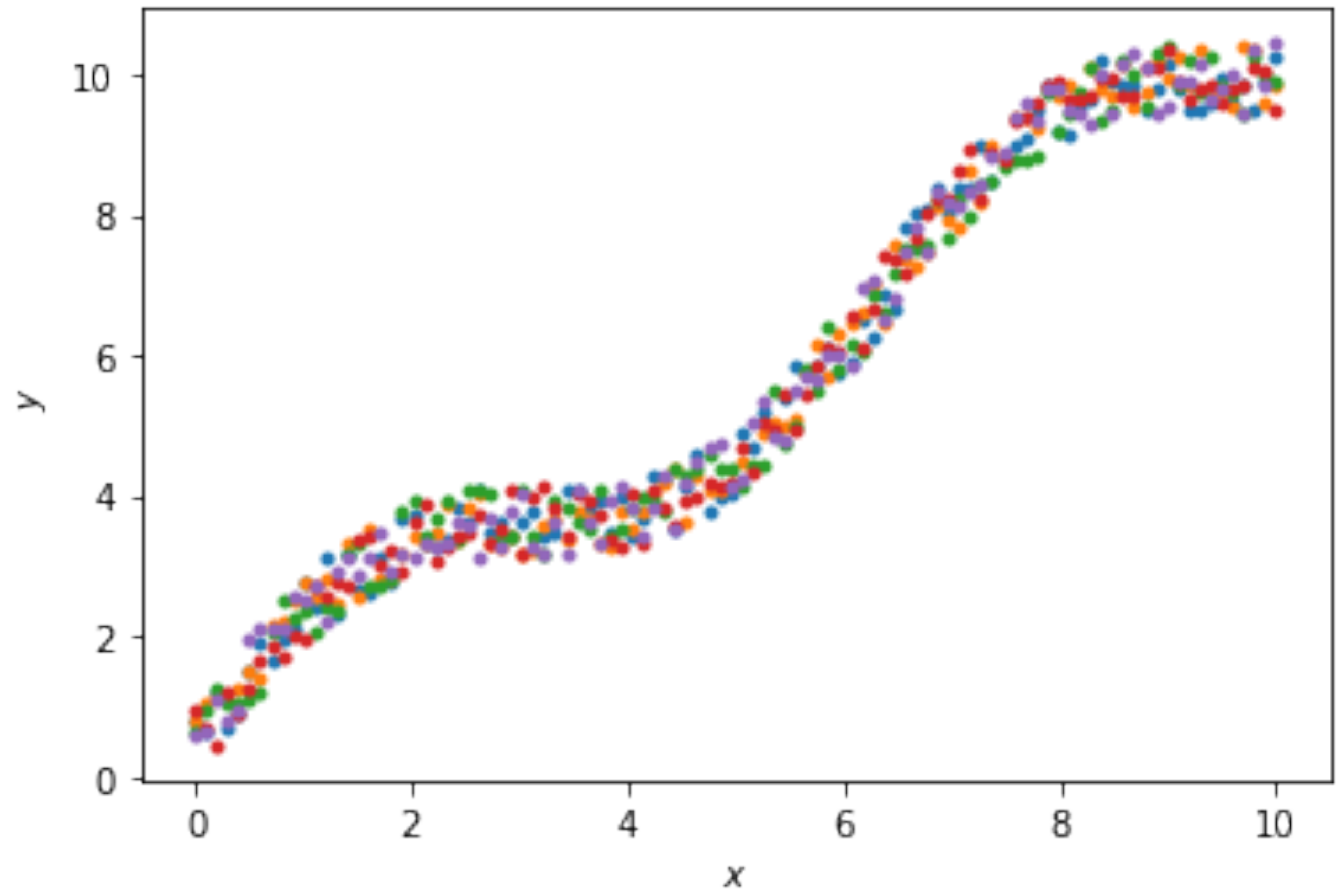
Mean squared error (MSE)

$$J(w) = \frac{1}{N} \sum_{i=0}^{N} (\hat{y}_i - y_i)^2$$

Mean absolute error (MAE)

$$J(w) = \frac{1}{N} \sum_{i=0}^{N} |\hat{y}_i - y_i|$$

LOSS FUNCTIONS



Loss function

Mean squared error

mean

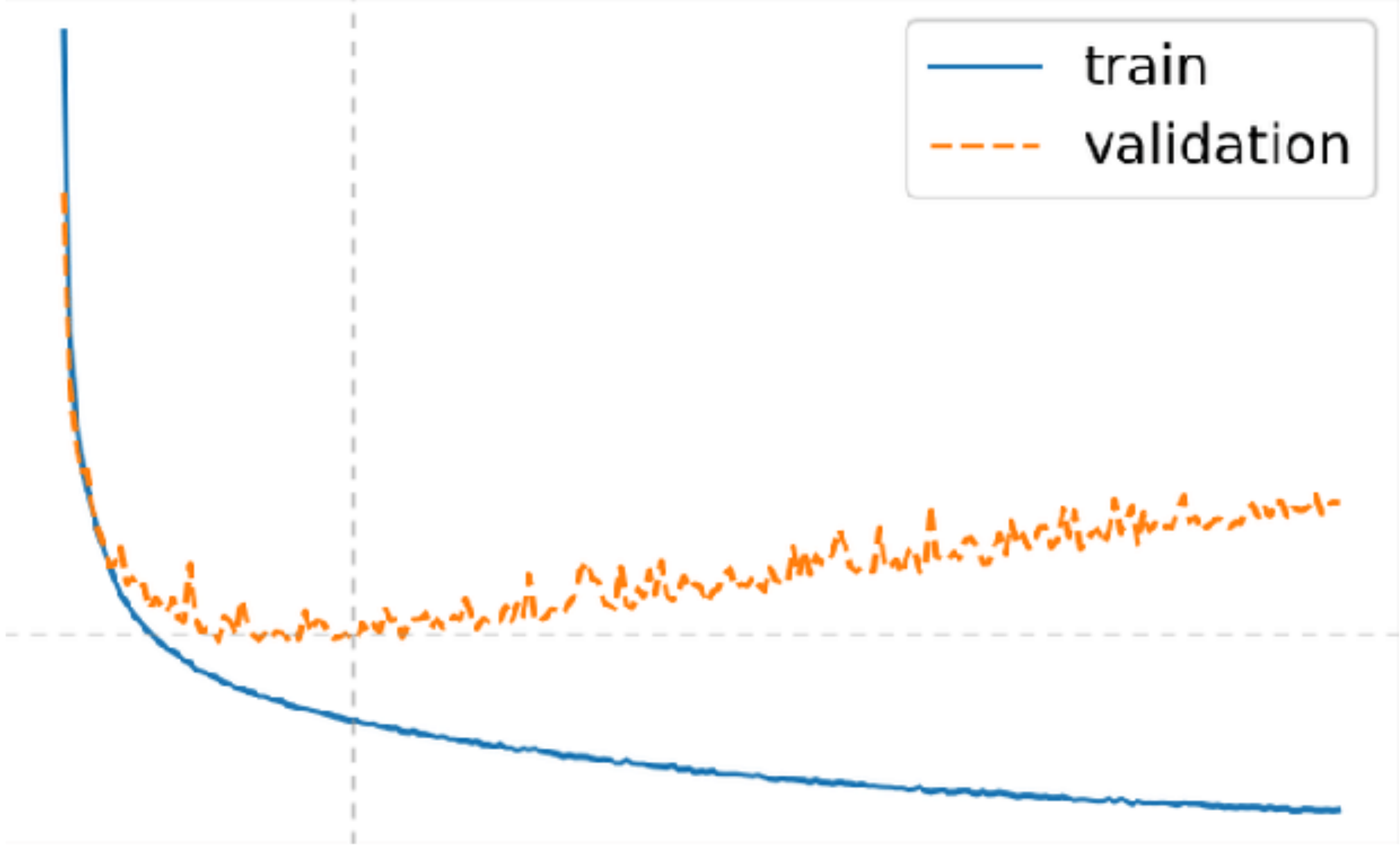
$$J(w) = \frac{1}{N} \sum_{i=0}^{N} (\hat{y}_i - y_i)^2$$

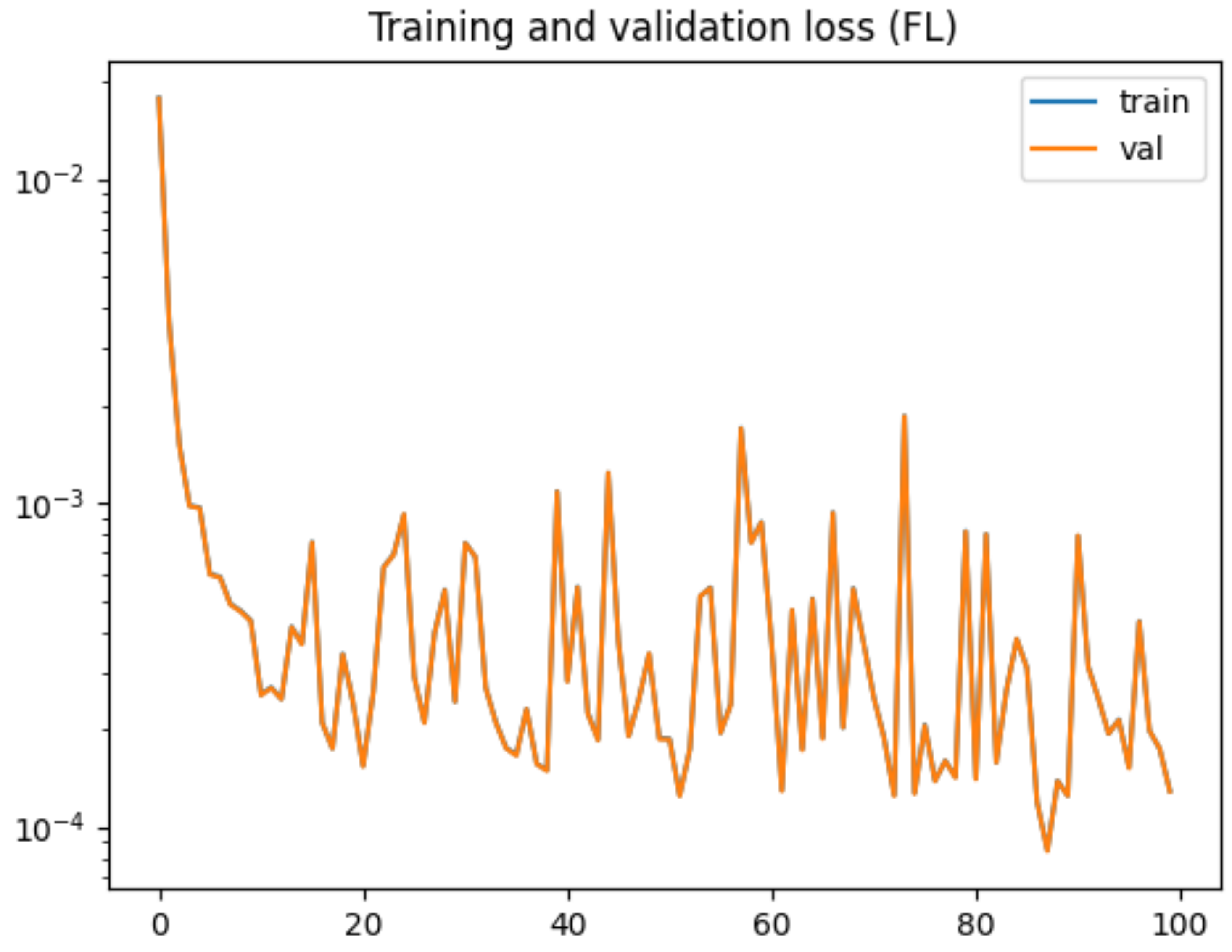
Mean absolute error

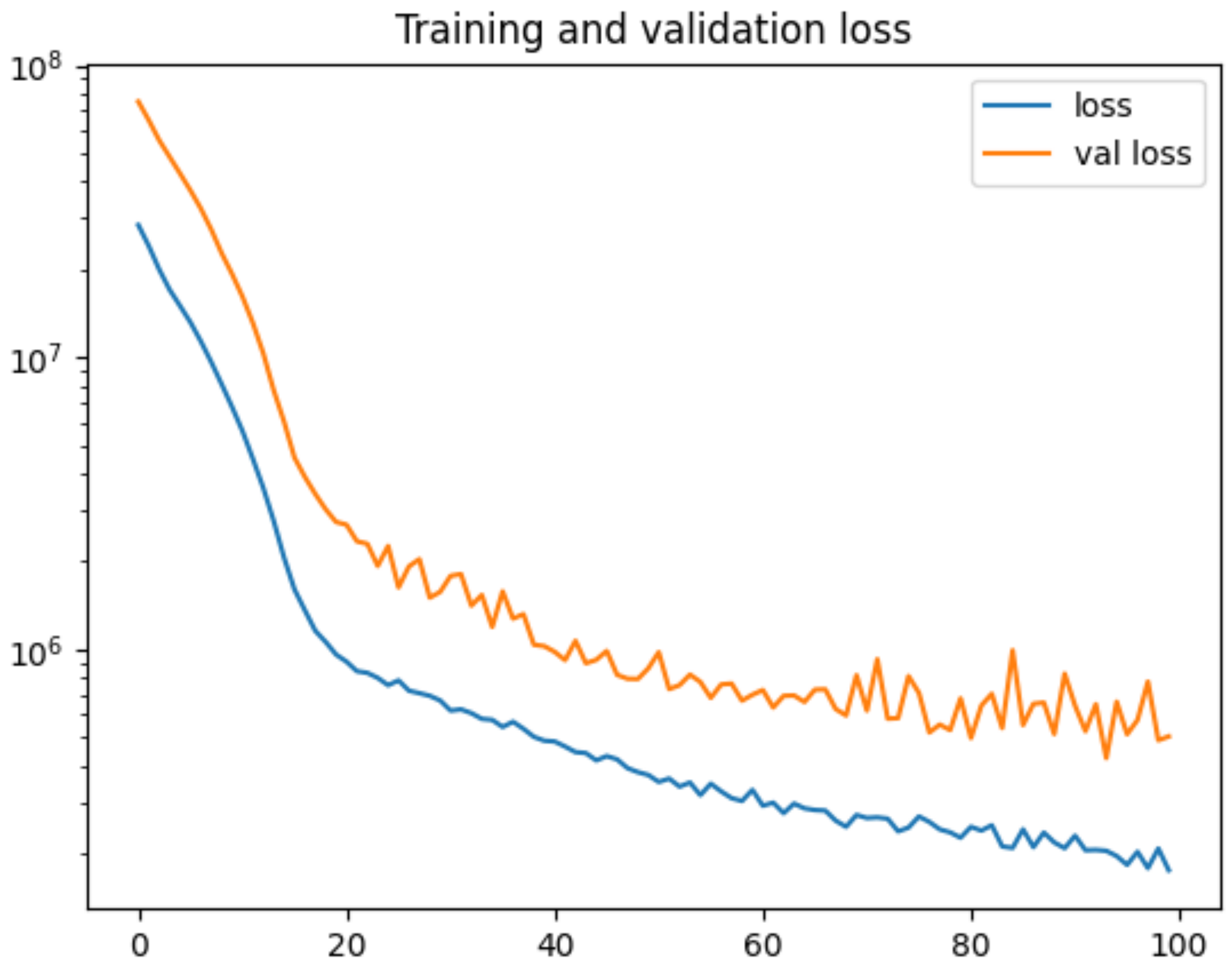
median

$$J(w) = \frac{1}{N} \sum_{i=0}^{N} |\hat{y}_i - y_i|$$

Learning (loss) curves



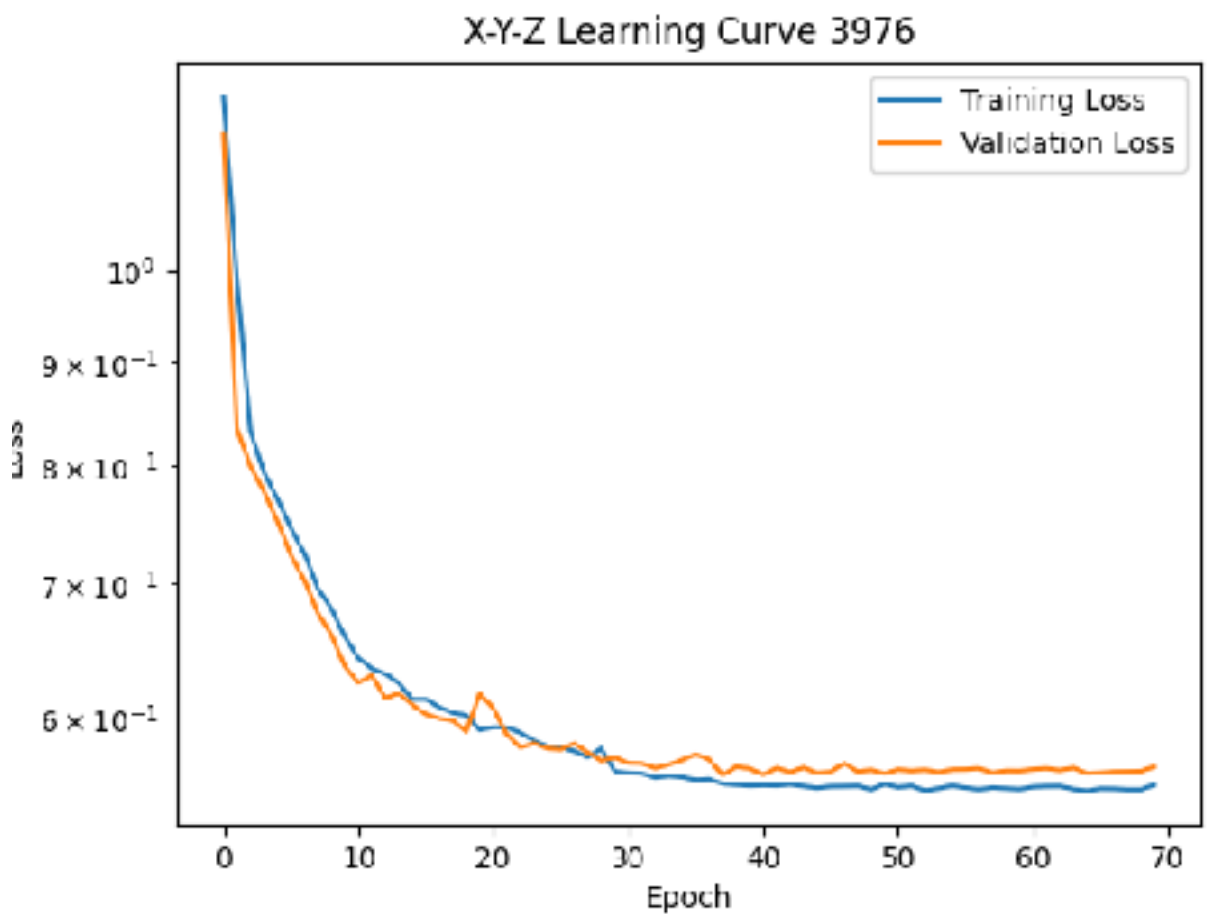


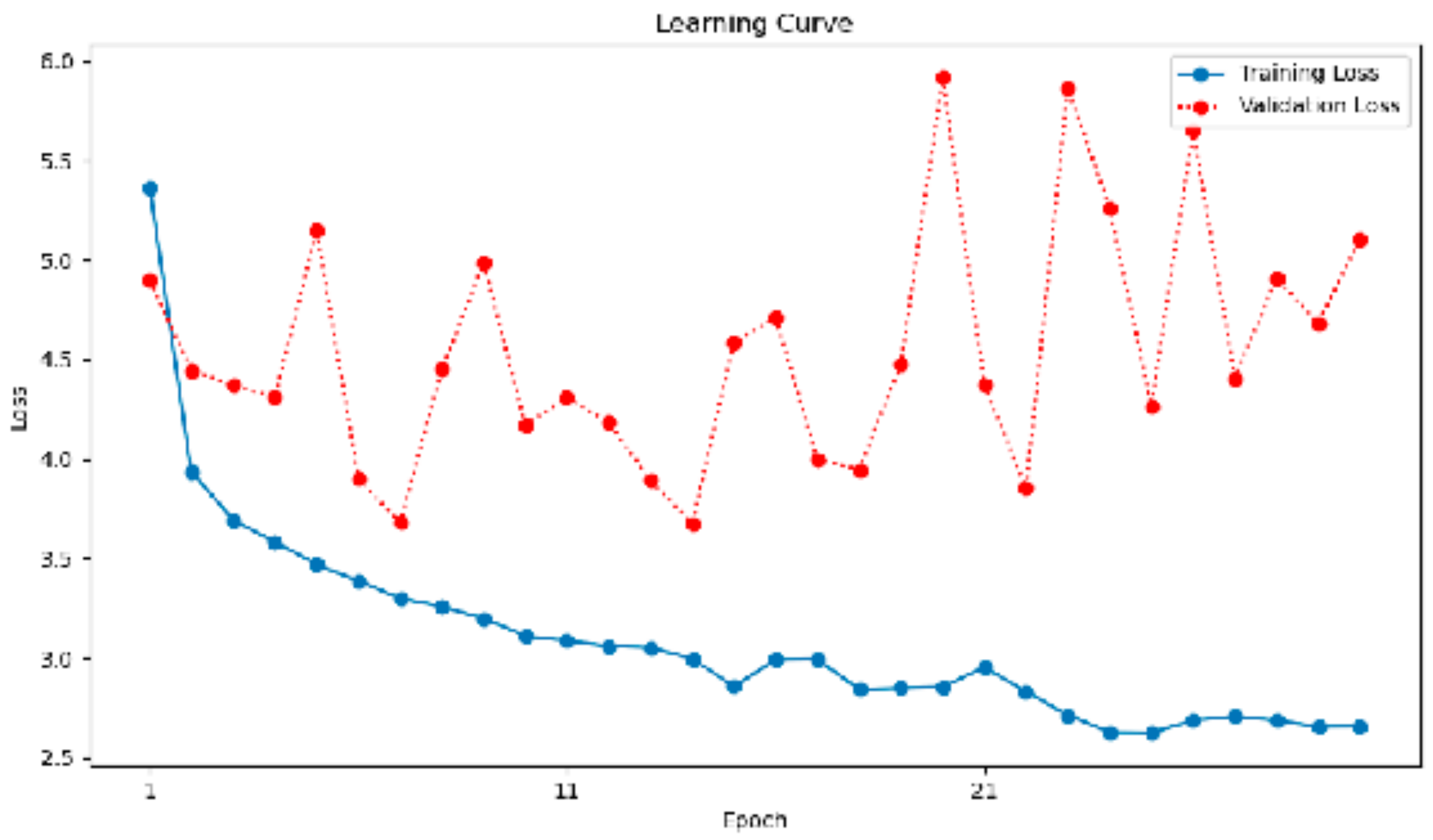


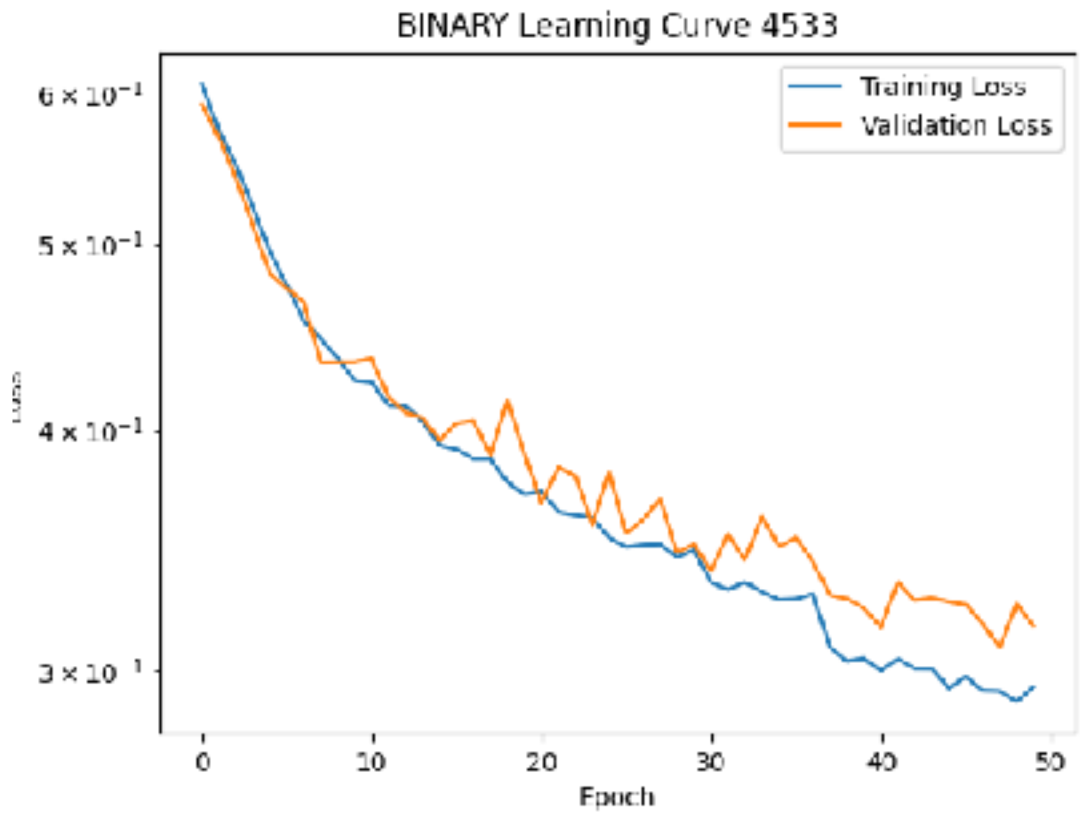
TRAINING

Remember that our goal is NOT to minimize loss on training data!

Learning curves







AUTOMATIC DIFFERENTIATION

TensorFlow PyTorch

Keras

MODERN NEURAL NETWORK ARCHITECTURES

CHRISTIAN SZEGEDY ET. AL. GOING DEEPER WITH CONVOLUTIONS.

“GoogLeNet network with all the bells and whistles”

MODERN NEURAL NETWORK ARCHITECTURES

CHRISTIAN SZEGEDY ET. AL. GOING DEEPER WITH CONVOLUTIONS.

"GoogLeNet network with all the bells and whistles"

PRACTICAL TIPS FOR TRAINING MODELS

DATA

	Feature 1	Feature 2	Feature 3	Target
Example 1				
Example 2				
Example 3				
Example 4				

NORMALIZATION

- Puts each feature on same scale

WHY?

DATA

	Feature 1	Feature 2	Feature 3	Target
Example 1				
Example 2				
Example 3				
Example 4				

NORMALIZATION

- Puts each feature on same scale
- Allows default hyperparameters to be a good starting point
  - learning rate, initialization of weights, etc.
- Options depend on data distribution
  - Standardization: mean: 0 stdev: 1
  - Min-max: [0,1]

DATA

	Feature 1	Feature 2	Feature 3	Target
Example 1				
Example 2				
Example 3				
Example 4				

ENCODING

- Non-numeric data
- Class-based features:
  - One-hot encoding: 3 → [0 0 1] [0 1 0] [1 0 0]

WHY??

DATA

	Feature 1	Feature 2	Feature 3	Target
Example 1				
Example 2				
Example 3				
Example 4				

ENCODING

- Non-numeric data
- Class-based features:
  - One-hot encoding: 3 → [0 0 1] [0 1 0] [1 0 0]
  - When classes do not have sequential meaning: ✅ cars vs dogs vs plants ❌ months

BUILDING AND TRAINING MODELS

TRAINING

- A challenging part of machine learning is gaining the experience for tuning models well.
- We will work on this skill!

COMMUNITY

- Each of you arrived here with your own backgrounds, specialty, and path in life
- Your experience and expertise are valuable here, no matter what it is
- If the activity is within your background, help others!
- If you are totally (or a little) lost, ask for help!
- It is our shared goal to have **each** of us leave with some new skill/knowledge/understanding