Mathematics of Artificial Intelligence

G. Giorgobiani, V. Kvaratskhelia, T. Saghinadze

Muskhelishvili Institute of Computational Mathematics of the Georgian Technical University (MICM)

2nd Int. Conf.: Science, Education, Innovations and Chemical Technologies – From Idea to Implementation. Tbilisi, Georgia, 2023

Supported by the Horizon Europe Grant Project "GAIN"

Introduction

This talk is a short survey of mathematical branches used in AI problems.
At the end we also briefly present research, which is carried out at MICM in the fields of ML, DNN and NLP by the support of currently running HORIZON EUROPE project “GAIN”.
We mostly touch the (DL) - Deep Learning (Deep Neural Networks – DNN) problems, a branch of ML.
We partially follow a monograph

- [1] The Modern Mathematics of Deep Learning (J. Berner, P. Grohs, G. Kutyniok, and P. Petersen)

We cite other relevant literature along the presentation.

Introduction. Applications of DL

- Most famous application of DL is image classification
- HEP (high-energy physics):*DL can improve the power of collider searches for exotic particles.*
  [3] P. Baldi et al, Searching for exotic particles in high-energy physics with deep learning.
- Development of drugs: *DNNs can make better predictions on large data sets.*
  [4] Junshui Ma et al, Deep neural nets as a method for quantitative structure-activity relationships

Introduction. Applications of DL

- Molecular ML models
  [5] Felix A Faber et al, Prediction errors of molecular machine learning models lower than hybrid DFT error
- DL for protein structure prediction
  [6] Andrew W Senior et al, Improved protein structure prediction using potentials from deep learning
- ML for Healthcare, Psychiatry (MEPHESTO project)
  [7] A. Konig, P. Müller et al. Multimodal phenotyping of psychiatric disorders from social interaction: Protocol of a clinical multicenter prospective study.

Neural Network

Definition 1. An artificial neuron with weights $w_1, ..., w_n \in \mathbb{R}$, the bias $b \in \mathbb{R}$ and activation function $\rho : \mathbb{R}^n \to \mathbb{R}$ is the function

$$f(x_1, ..., x_n) = \rho(\sum_{i=1}^{n} x_i w_i - b)$$

Denoting $x = (x_1, ..., x_n), w = (w_1, ..., w_n), x, w \in \mathbb{R}^n$

$$f(x) = \rho(\langle x, w \rangle - b)$$

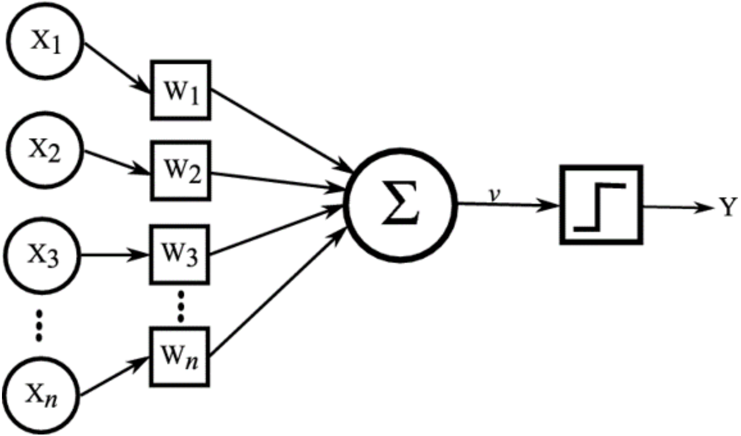
where

$$\langle x, w \rangle = \sum_{i=1}^{n} x_i w_i$$

is the inner (scalar) product on $\mathbb{R}^n$.
A "zoo" of activation functions: Heaviside function, Sigmoid function, Rectifiable Linear Unit (ReLU)

$$\text{ReLU} \equiv \rho(t) = \max(t, 0)$$

Neural Network/ Perceptron - NN with 1 neuron

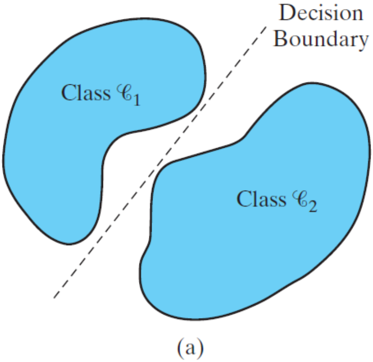


Neural Network/ Perceptron - NN with 1 neuron

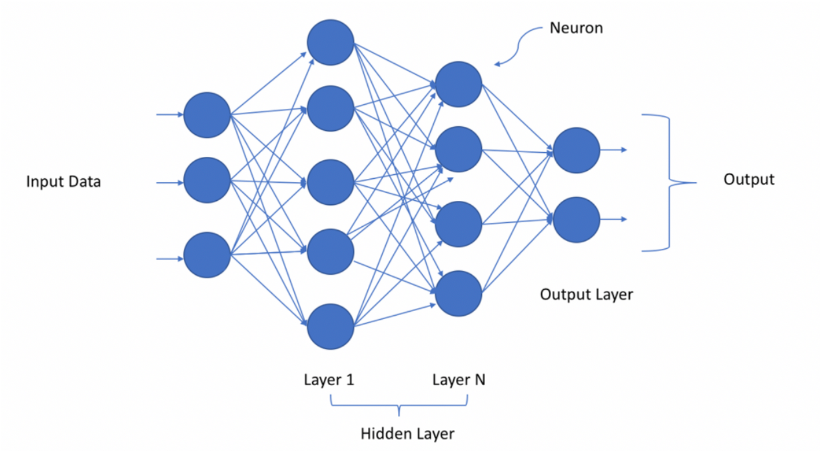
1943 - Warren MuCulloch (neuroscientist), Walter Pitts (logician)
1958 - Frank Rosenblatt American psychologist
- Implemented hardware (bigger than person)
- Weights stored in potentiometers
- Updated using electric motors
- Could recognize alphabet.

Frank Rosenblatt is recognized as the "Father of Deep Learning" together with Hinton, LeCun, Bengio and others.

Perceptron works well when the classes are linearly separated



DNN


Neuron
Input Data
Output
Output Layer
Layer 1
Layer N
Hidden Layer

DNN

For N layers and input $x \in \mathbb{R}^d$, DNN of depth N is

$$\Phi(x) = T_N \rho(T_{N-1} \rho(...\rho(T_1(x)))),$$

where for $n = 1, ..., N$

$$T_n(x) = W^n x + b^n \text{ – Affine (Linear) transforms}$$

$$W^n \text{– Weight Matrices}$$

$$b^n \text{ – Bias Vectors}$$

$$\rho \text{ – Activation function}$$

How DNN works?

Given inputs and outputs (Training Examples) - $(x^i, y^i)_{i=1}^m$

- Step 1 - Split Training Examples into a Training, Validation and Test Sets: D_{tr}, D_v, D_{test}
- Step 2 - Choose Architecture: number of layers N, number of neurons in each layer, activation function
- Step 3 - Training process. Using only Training Set, minimize the *Loss (Cost) Function* with respect to the weights and bias

$$\mathcal{L} = \frac{1}{card(D_{tr})} \sum_{i \in D_{tr}} (\Phi(x^i) - y^i)^2$$

  Typical algorithm - Gradient Descent
- Step 4 -Testing: Performance of the trained DNN is tested using the Test Set D_{test}, i.e. evaluate the closeness of

$$\Phi(x^i) \text{and} y^i, (x_i, y_i) \in D_{test}$$

- Validation Set is used to control the over-fitting

How DNN works? Example - Classification Problem

Classification Problem: Given inputs and outputs (Training Examples)

$$(x^i, y^i)_{i=1}^m, y^i \in \{1, 2, ...d\} - \text{Class Labels}$$

Softmax function as an activation function:

$$\rho : \mathbb{R}^d \to (0,1)^d, \rho(x)_i = \frac{e^{x_i}}{\sum_{j=1}^d e^{xj}}, x = (x_1, ..., x_d)$$

Using Softmax on the last layer returns as output the probabilities $P \equiv (p_i), i = 1, ..., m$ for the classes.
To compare DNN's estimation (output) and the true distribution $Q \equiv (q_i), i = 1, ..., m$ the Cross-entropy is applied as a loss function (Logistic function):

$$H(P, Q) = -\sum_{i=1}^{m} p_i \log(q_i)$$

Basic Math for DL

- PROBABILITY/STATISTICS
- ALGEBRA/Vectors, Matrices etc.
- ANALYSIS/Optimization, Gradient etc.

Advanced Math for DL

- Universality of ReLu. Classical Stone-Weierstrass theorem [8] -Kurt Hornik, Maxwell Stinchcombe, and Halbert White, Multilayer feedforward networks are universal approximators
- Laplacian eigen values,Heat Equation for Shape Recognition [10] - Mohamed Ben Haj Rhouma, Mohamed Ali Khabou, Lotfi Hermi. Shape Recognition Based on Eigenvalues of the Laplacian
- Permutations- our 3 papers are cited [11] - Yucheng Lu, Wentao Guo, Christopher De Sa. GraB: Finding Provably Better Data Permutations than Random Reshuffling.
- Geometric Deep Learning. Finding symmetries. [2] - M. M. Bronstein, J. Bruna, T. Cohen, P. Velickovic. Geometric Deep Learning, Grids, Groups, Graphs...

AI for Mathematical Problems

- Inverse Problems e.g. MRI
  [12] - Jure Zbontar, Florian Knoll et al. fastMRI: An open dataset and benchmarks for accelerated MRI.
- Partial Differential Equations to avoid the curse of dimensionality

GAIN - Georgian Artificial Intelligence Networking and Twinning Initiative

HORIZON EUROPE - WIDERA-2021-ACCESS-03 (Twinning)
Project Partners - European leaders in AI, Big Data, Robotics, Computer Vision

- DFKI - German Research Centre for Artificial Intelligence.
- INRIA - National Institute for Research in Digital Science and Technology.
- EXOLAUNCH GmbH (EXO) - high-tech spin-off company, TUBerlin

GAIN - Research Topics

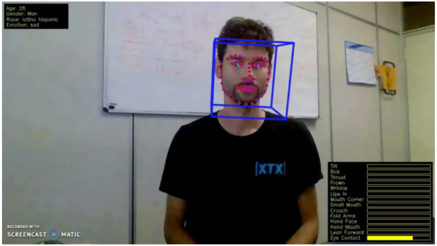
- AI Technologies for Human Behaviour Understanding, Emotions (face and full body video, biosignals)
- AI Methods for Action detection/recognition
- AI Methods for Deep Speech Analysis in Health, NLP and NLP-fMRI
- Georgian Corpus of medical data

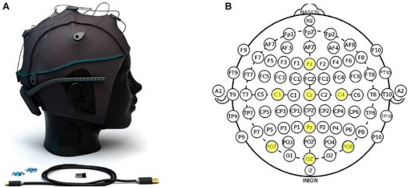
AI in psychiatry - MEPHESTO/GAIN

MEPHESTO - is a large scale DFKI-INRIA joint project.
[7] - Multimodal phenotyping of psychiatric disorders from social interaction: Protocol of a clinical multicenter prospective study.

GAIN: ML, DL applications in psychiatry - MEPHESTO

Video/Audio of Doctor - Patient interviews and biosignals are analysed by ML methods.





Computing Center at MICM - GAIN and SRNSF

Model: HPE Proliant dl385 Gen10 Plus
CPU: 2 x AMD EPYC 7713
RAM: 256 GB DDR4
HD: 7.2 TB
GPU: 2 x Nvidia A100 40GB;

THANK YOU

References

J. Berner, P. Grohs, G. Kutyniok, and P. Petersen. The Modern Mathematics of Deep Learning. In: Mathematical Aspects of Deep Learning, Cambridge University Press.

M. M. Bronstein, J. Bruna, T. Cohen, P. Velickovic. Geometric Deep Learning, Grids, Groups, Graphs, Geodesics, and Gauges. https://arxiv.org/pdf/2104.13478.pdf

P. Baldi, P. Sadowski, D. Whiteson, Searching for exotic particles in high-energy physics with deep learning, Nature Communications 5 (2014), no. 1, 1-9. DOI: 10.1038/ncomms53081

Junshui Ma, Robert P Sheridan, Andy Liaw, George E Dahl, and Vladimir Svetnik, Deep neural nets as a method for quantitative structure-activity relationships, Journal of chemical information and modeling 55 (2015), no. 2, 263-274.

Felix A Faber et al, Prediction errors of molecular machine learning models lower than hybrid DFT error, Journal of

Chemical Theory and Computation 13 (2017), no. 11, 5255-5264. https://doi.org/10.1021/acs.jctc.7b00577

- Andrew W Senior et al, Improved protein structure prediction using potentials from deep learning, Nature 577 (2020), no. 7792, 706-710

- Alexandra Konig, Philipp Müller et al. Multimodal phenotyping of psychiatric disorders from social interaction: Protocol of a clinical multicenter prospective study. Personalized Medicine in Psychiatry Volumes 33–34, 2022, https://doi.org/10.1016/j.pmip.2022.100094

- Kurt Hornik, Maxwell Stinchcombe, and Halbert White, Multilayer feedforward networks are universal approximators, Neural Networks 2 (1989), no. 5, 359-366.

- M. M. Bronstein, J. Bruna, T. Cohen, P. Velickovic. Geometric Deep Learning, Grids, Groups, Graphs, Geodesics, and Gauges. https://arxiv.org/pdf/2104.13478.pdf or https://geometricdeeplearning.com/

Mohamed Ben Haj Rhouma, Mohamed Ali Khabou, Lotfi Hermi. Shape Recognition Based on Eigenvalues of the Laplacian. Advances in Imaging and Electron Physics, V. 167, 2011, pp. 185–254.

Yucheng Lu, Wentao Guo, Christopher De Sa. GraB: Finding Provably Better Data Permutations than Random Reshuffling. 36th Conference on Neural Information Processing Systems (NeurIPS 2022).

Jure Zbontar, Florian Knoll et al. fastMRI: An open dataset and benchmarks for accelerated MRI, 2018, arXiv preprint arXiv:1811.08839.