Demystifying Artificial Intelligence Sorcery

(Part 1: Fuzzy Logic & Neural Networks)[a]

Abdelbacet Mhamdi
abdelbacet.mhamdi@bizerte.r-iset.tn

Dr.-Ing. in Electrical Engineering
Senior Lecturer at ISET Bizerte

[a]Available @ https://github.com/a-mhamdi/jlai/

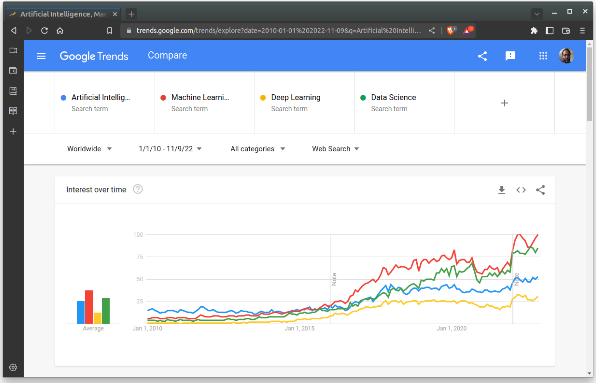
Disclaimer

This document features some materials gathered from multiple online sources. Please note no copyright infringement is intended, and I do not own nor claim to own any of the original materials. They are used for educational purposes only. I have included links solely as a convenience to the reader. Some links within these slides may lead to other websites, including those operated and maintained by third parties. The presence of such a link does not imply a responsibility for the linked site or an endorsement of the linked site, its operator, or its contents.

Roadmap

An overview

Trends



"Numbers represent search interest relative to the highest point on the chart for the given region and time.

- A value of 100 is the peak popularity for the term;
- A value of 50 means that the term is half as popular;
- A score of 0 means there was not enough data for this term."

Top Uses

Artificial intelligence is a branch of computer science which focuses on automation of intelligent behavior.

Some definitions can be categorized into four frames.

Artificial intelligence is a branch of computer science which focuses on automation of intelligent behavior.

Some definitions can be categorized into four frames.

Systems that think like humans

[Bel78]

"[The automation of] activities that we associate with human thinking, activities such as decision-making, problem-solving, learning..."

Bellman, R. E. *An Introduction to Artificial Intelligence: Can Computers Think?* **Boyd & Fraser Publishing Company.**

[Hau89]

"The exciting new effort to make computers think[...] *machines with minds,* in the full and literal sense"

Haugeland, J. (1989). *Artificial Intelligence: The Very Idea.* **A Bradford book. MIT Press.**

Systems that think rationally

[CMM85]

"The study of mental faculties through the use of computational models."

Charniak, E., McDermott, D., and McDermott, D. V. (1985). *Introduction to Artificial Intelligence*. **Addison-Wesley series in computer science and information processing. Addison-Wesley.**

[Win92]

"The study of the computations that make it possible to perceive, reason, and act."

Winston, P. H. (1992). *Artificial Intelligence*. **A-W Series in Computer Science. Addison-Wesley Publishing Company.**

Systems that act like humans

[Kur92]

"The art of creating machines that perform functions that require intelligence when performed by people."

Kurzweil, R. (1992). *The Age of Intelligent Machines*. **Viking.**

[RK91]

"The study of how to make computers do things at which, at the moment, people are better."

Rich, E. and Knight, K. (1991). *Artificial Intelligence*. **Artificial Intelligence Series. McGraw-Hill.**

Systems that act rationally

[Sch90]

"A field of study that seeks to explain and emulate intelligent behavior in terms of computational processes."

Schalkoff, R. J. (1990). *Artificial Intelligence: An Engineering Approach*. **McGraw-Hill Computer science series. McGraw-Hill.**

[LS93]

"The branch of computer science that is concerned with the automation of intelligent behavior."

Luger, G. F. and Stubblefield, W. A. *Artificial Intelligence: Structures and Strategies for Complex Problem Solving*. **Artificial intelligence. Benjamin/Cummings Publishing Company.**

Thought-Provoking Questions

How to achieve intelligence on a computer system

What do we mean by "Intelligence"?

- Single faculty or gathering of abilities
- Learned or existing
- What happens when we learn
- Are creativity and intuition measurable
- Does observable behavior infer to intelligence
- How knowledge is routed in the human brain

Thought-Provoking Questions

How to achieve intelligence on a computer system

What do we mean by "Intelligence"?

- Single faculty or gathering of abilities
- Learned or existing
- What happens when we learn
- Are creativity and intuition measurable
- Does observable behavior infer to intelligence
- How knowledge is routed in the human brain

Turing Test

Alan Turing (1950)

The ability to achieve human level performance in all cognitive tasks, sufficient to fool an interrogator.

- ✓ Natural Language Processing (NLP) *(Communicate in human language)*
- ✓ Knowledge Representation *(Store information)*
- ✓ Automated Reasoning *(Answer questions & draw conclusions)*
- ✓ Machine Learning (ML) *(Adapt to new circumstances, detect & extrapolate patterns)*

Forms of AI

- ☆ Expert Systems *(Based on knowledge or rule settings)*
- ☆ Fuzzy Systems *(Based on fuzzy set theory)*
- ☆ Artificial Neural Networks
- ☆ Genetic Algorithms
- ☆ Belief Networks
- ☆ Hybrid Systems *(Combine two or more approaches)*

Programming Language

julialang.org/

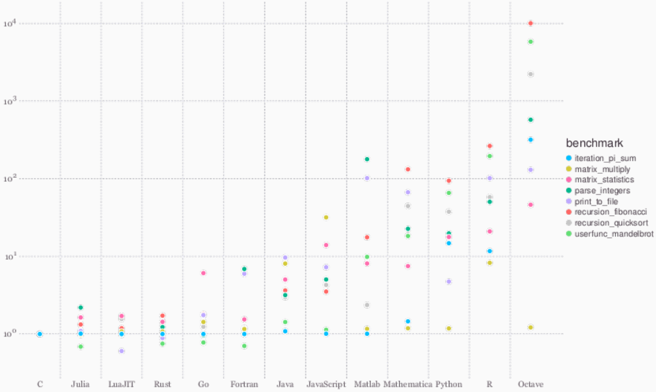
Development Environments

- ▲ $ docker compose up
- ▼ $ docker compose down

Julia in a Nutshell

- Fast
- Dynamic
- Reproducible
- Composable
- General
- Open Source

Julia Micro-Benchmarks (1/2)



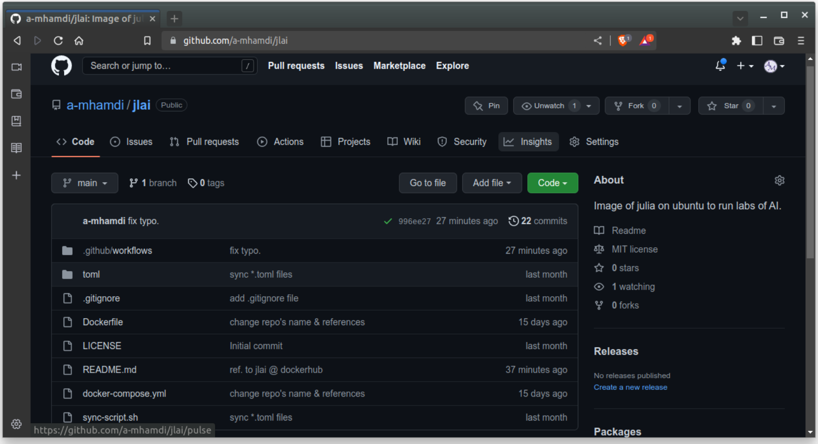
https://julialang.org/benchmarks

Julia Micro-Benchmarks (2/2)

Geometric Means of Micro-Benchmarks by Language

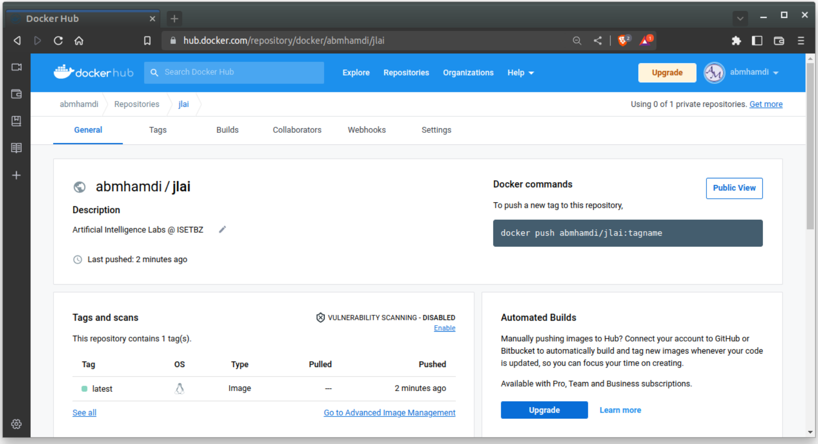
1	C	1.0
2	Julia	1.17006
3	LuaJIT	1.02931
4	Rust	1.0999
5	Go	1.49917
6	Fortran	1.67022
7	Java	3.46773
8	JavaScript	4.79602
9	Matlab	9.57235
10	Mathematica	14.6387
11	Python	16.9262
12	R	48.5796
13	Octave	338.704

Source Control Management (SCM)



https://github.com/a-mhamdi/jlai

Continuous Integration (CI)



https://hub.docker.com/r/abmhamdi/jlai

Fuzzy Logic

What is Fuzzy Logic?

"There are many misconceptions about fuzzy logic. To begin with, fuzzy logic is not fuzzy. Basically, fuzzy logic is a precise logic of imprecision. [...] fuzzy logic is designed to deal with imperfect information. Imperfect information is information which in one or more aspects is imprecise, uncertain, incomplete, unreliable, vague or partially true. In the real world, such information is the norm rather than exception."

Lotfi Zadeh, *WCECS 2014*

" Fuzzy Logic, in computer science, is a form of logic used in some expert systems and other artificial-intelligence applications in which variables can have degrees of truthfulness or falsehood represented by a range of values between 1 (true) and 0 (false). With fuzzy logic, the outcome of an operation can be expressed as a probability rather than as a certainty. For example, in addition to being either true or false, an outcome might have such meanings as probably true, possibly true, possibly false, and probably false."

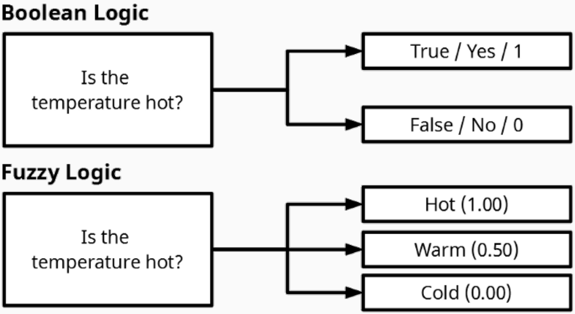
Fuzzy Logic, Microsoft® Encarta® Online Encyclopedia 2009
https://www.refseek.com/data/cache/en/1/Fuzzy_Logic.html

What does Fuzzy Logic have to offer?

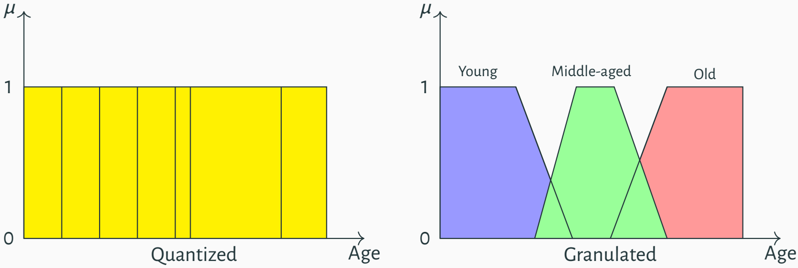
Fuzzy Logic aims at formalizing/mechanizing two noticeable human capabilities:

1. communicating, reasoning and rational decision making
   (in presence of imprecision, uncertainty & partiality of truth)
2. performing a wide variety of tasks
   (w/o measurements or computations)

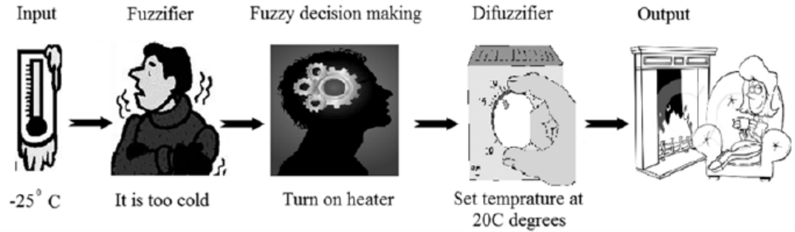
Fuzzy logic as an extension of the Boolean Logic



Continuous → Quantized → Granulated



Example of a Fuzzy Control System



Architecture

Rule Base is provided by experts. It contains the set of rules to govern the decision making.

Fuzzification converts crisp numbers to fuzzy sets.

Inference Engine decides which rules to be fired matching degree of the current fuzzy inputs.

Defuzzification converts the fuzzy sets delivered by the inference engine into some crisp value

Defuzzification

A fuzzy value can be defuzzified through multiple ways.

1. Center of Sums
2. Centroid Method
3. Center of Area
4. Weighted Average Method
5. Max-Membership Principal

Tipping Problem

What should be the TIP at a restaurant, given the quality of FOOD and of SERVICE. These latter are represented by some scores ranging from 0 *(poor)* to 10 *(excellent)*.

Rules Base

1. FOOD is rancid || SERVICE is poor $\Longrightarrow$ TIP is cheap;
2. SERVICE is good $\Longrightarrow$ TIP is average;
3. FOOD is delicious || SERVICE is excellent $\Longrightarrow$ TIP is generous.

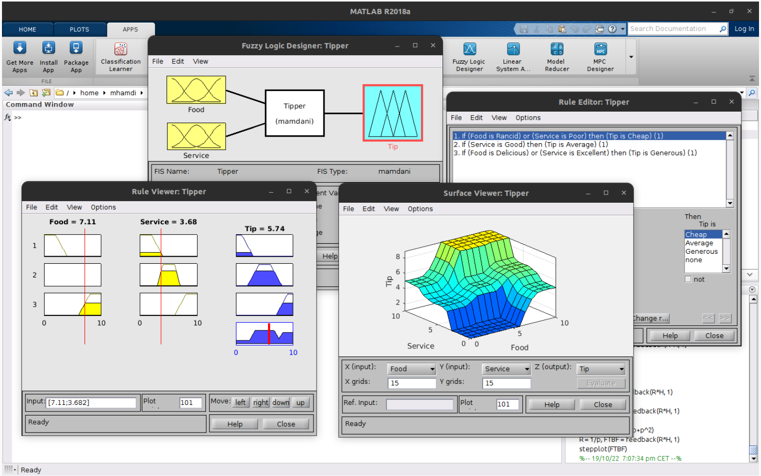
Tipping Problem

What should be the TIP at a restaurant, given the quality of FOOD and of SERVICE. These latter are represented by some scores ranging from 0 *(poor)* to 10 *(excellent)*.

Rules Base

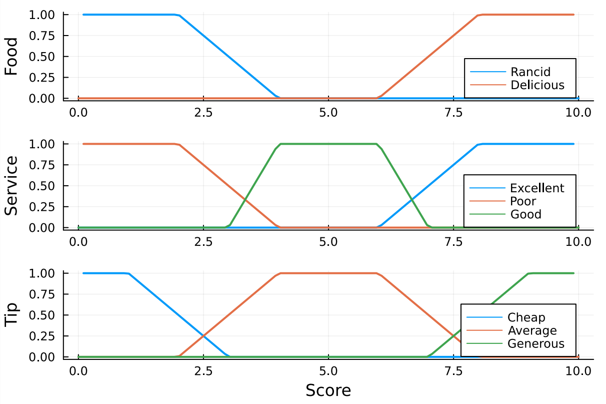
1. FOOD is rancid || SERVICE is poor $\Longrightarrow$ TIP is cheap;
2. SERVICE is good $\Longrightarrow$ TIP is average;
3. FOOD is delicious || SERVICE is excellent $\Longrightarrow$ TIP is generous.

Using Fuzzy Logic Toolbox



Code is available at https://github.com/a-mhamdi/cosnip/ $\longrightarrow$ *Matlab* $\longrightarrow$ *Fuzzy* $\longrightarrow$ *Tipper.fis*

Using Fuzzy.jl Package



Code is available at https://github.com/a-mhamdi/jlai/
→ *Codes* → *Julia* → *Part-1* → *tipper.jl*

Fuzzy Numbers (1/6)

★ Represent imprecise numbers: number & linguistic modifier *(e.g., nearly, around, etc.)*

- approximately five kilos
- about 12 pm

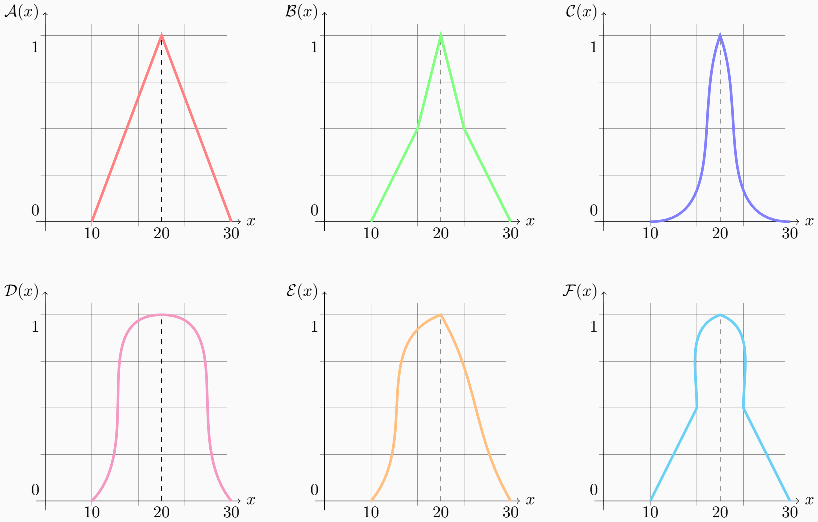
★ Play an important role in decision making, approximate reasoning, statistics with imprecise probabilities and fuzzy control.

We need to perform arithmetic operations on fuzzy numbers *(e.g., calculate a ratio of some fuzzy output over some fuzzy input)*

"around 20"

- includes some number values on either side of the central value of 20
- Central value is fully compatible with concept
- Number around central value are compatible with it to lesser degrees
- Degree of compatibility represented by fuzzy set; Membership value decreases from 1.0 to 0.0 on both sides of central value = fuzzy number.

Fuzzy Numbers (2/6)



Fuzzy Numbers (3/6)

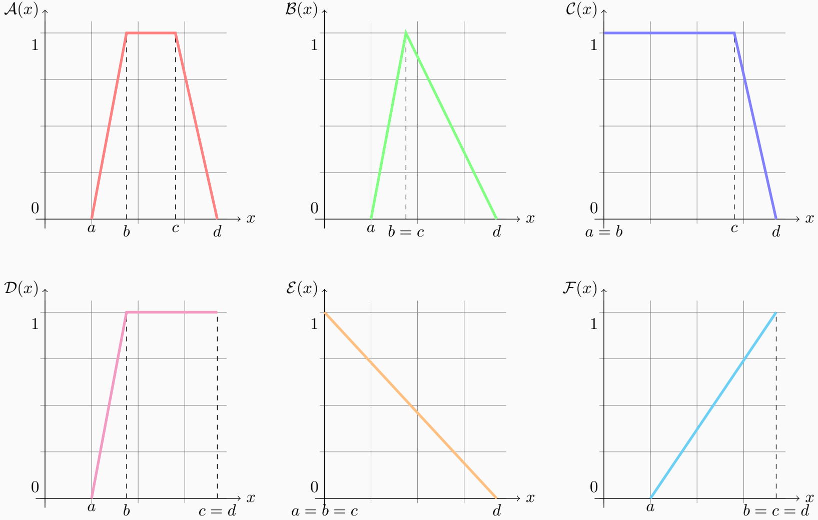
For a fuzzy membership function to qualify as a fuzzy number, it must capture our intuitive concept of a set of numbers around a given real number or interval of real numbers

$$\mathcal{A}(x) = \begin{cases} f(x) & \text{for} \quad x \in [a,\ b] \\ 1 & \text{for} \quad x \in [b,\ c] \\ g(x) & \text{for} \quad x \in [c,\ d] \\ 0 & \text{for} \quad x < a \text{ or } x > d \end{cases} \tag{1}$$

Common shapes of Fuzzy Numbers

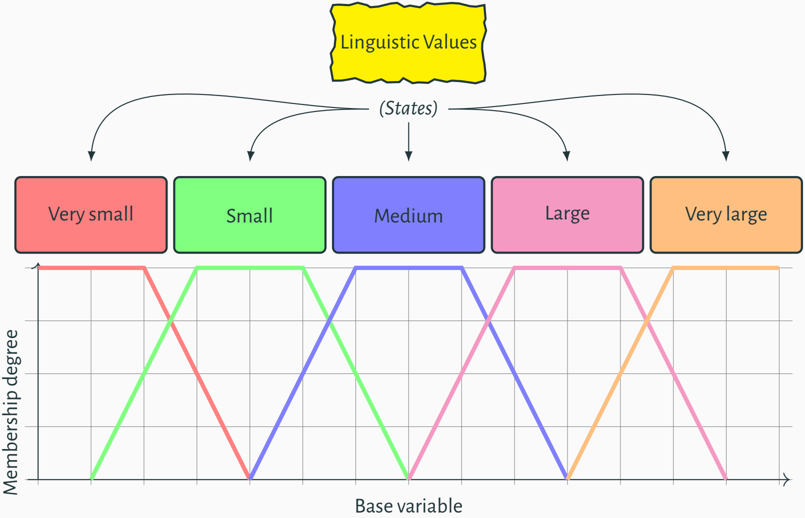
- Most common membership functions are trapezoidal and triangular *(easy to construct and manipulate)*
- Choice of a, b, c and d is important and is highly context-dependant
- Most applications not significantly affected by shapes of functions *(i.e., use linear shapes)*
- When some of real numbers (a, b, c, d) are equal, get degenerated forms of fuzzy numbers

Fuzzy Numbers (4/6)



Fuzzy Numbers (5/6)

States are fuzzy numbers which represent linguistic concepts



Fuzzy Numbers (6/6)

1. Fuzzy numbers are normal fuzzy sets (height=1)
2. Fuzzy numbers are convex fuzzy sets
3. Support of every fuzzy number is open interval (a, d) of real numbers (support must be bounded)
4. Interval analysis can be used to define arithmetic operations on fuzzy numbers

Basic arithmetic operations:

- addition $[a,\ b] + [c,\ d] = [a + c,\ b + d]$
- Subtraction $[a,\ b] - [c,\ d] = [a - d,\ b - c]$
- Multiplication $[a,\ b] \times [c,\ d] = [\min(ac, ad, bc, bd),\ \max(ac, ad, bc, bd)]$
- Division[1] $[a,\ b] \div [c,\ d] = [a,\ b] \times [1/c,\ 1/d]$

[1]Interval division assumes that the number 0 is not one of the elements in the divisor interval [c, d].

Constructing Fuzzy Sets (1/5)

How would you assess today's temperature?

We can describe a parameter describing a phenomena *(e.g., Temperature for environment or Error for distance measurement)* using a finite, small number of descriptors, referred to as linguistic variables of parameter.

Temperature (T) {Cold, Average, Warm}

Error (E) {Small, Medium, Large}

The number of linguistic variables should be kept small (7 ± 2) due to our limited capacity to distinguish more. Commonly 3 to 5 linguistics variables are used in describing parameters.

Constructing Fuzzy Sets (2/5)

Fuzzy Sets ⟶ Membership functions

- Fuzzy sets offer an important and unique approach to describe linguistic variables
- Membership functions

$$\mathcal{A}(x) = \mathcal{X} \rightarrow [0, 1]$$

  are mathematical functions that are used to describe fuzzy sets
- Choosing membership functions require understanding of:
  - nature of the problem and parameter at hand
  - Level of details to be captured
  - Context of application

Prerequistes

- Concepts and linguistic values *(e.g., cold temperature)*
- Numerical measurements and/or linguistic assessments *(e.g., degrees Celsius)*
- Given context
- Data or Expert

Constructing Fuzzy Sets (3/5)

To construct fuzzy sets:

Expert-Driven Using developer, user, decision-maker, etc.

1. Direct methods
   - Answers to questions that explicitly pertain to the constructed membership function
   - Single or multiple experts
2. indirect methods
   - Simpler questions, easier to answer, less sensitive to subjective biases, pertain to membership function only implicitly
   - Single or multiple experts

Data-Driven Form data to fuzzy sets

Constructing Fuzzy Sets (4/5)

Direct Methods with Multiple Experts

Example

n experts were asked to validate the proposition **"x belongs to A"** as either true or false

True $a_i(x) = 1$

False $a_i(x) = 0$

where $i \in \{1 \cdots n\}$ denotes the i^{th} expert.

$$A(x) = \frac{1}{n} \sum_{i=1}^{n} a_i(x)$$

Can also distinguish degrees of competence c_i of individual experts:

$$A(x) = \sum_{i=1}^{n} c_i a_i(x), \qquad \text{where} \quad \sum_{i=1}^{n} c_i = 1$$

CONSTRUCTING FUZZY SETS (5/5)

- Given 5 labourers $\{\mathfrak{L}1, \mathfrak{L}2, \mathfrak{L}3, \mathfrak{L}4, \mathfrak{L}5\}$
- Need to determine membership function **"A"** that captures linguistic term **"Excellent Labourer"**
- Ask 10 superintendents if particular person is excellent labourer *(answer either yes* (1) *or no* (0)*)*
- For each labourer, calculate membership grade of belonging to fuzzy set **"A"** by taking ratio of total number of yes (1) to total number of responses.

	$\mathfrak{L}1$	$\mathfrak{L}2$	$\mathfrak{L}3$	$\mathfrak{L}4$	$\mathfrak{L}5$
$\mathfrak{E}\#1$	1	1	1	1	1
$\mathfrak{E}\#2$	0	0	1	1	1
$\mathfrak{E}\#3$	0	1	0	1	0
$\mathfrak{E}\#4$	1	0	1	1	1
$\mathfrak{E}\#5$	0	0	1	1	1
$\mathfrak{E}\#6$	0	1	1	1	1
$\mathfrak{E}\#7$	0	0	0	0	0
$\mathfrak{E}\#8$	1	1	1	1	1
$\mathfrak{E}\#9$	0	0	0	1	0
$\mathfrak{E}\#10$	0	0	0	1	0

$\Longrightarrow$ Opinions of individual experts must be aggregated

The resulting set would be: $\boxed{A = 0.3/\mathfrak{L}1 + 0.4/\mathfrak{L}2 + 0.6/\mathfrak{L}3 + 0.9/\mathfrak{L}4 + 0.6/\mathfrak{L}5}$

Fuzzy Inference Systems (FIS)

Mamdani Algorithm

- was introduced by Ebrahim (Abe) H. Mamdani in 1975
- works using rules of linguistics, style like human concepts
  (more intuitive and easier to understand)
- creates a methodology to design control system
- well suited to applications where rules are inspired from human expert knowledge.

Task #1

Consider a fuzzy logic system with two inputs u, v and an output w. We suppose that each variable ranges from $0 \longrightarrow 10$. w changes by a unit step. The membership functions of the fuzzy variables are described below.

- u can be:

  Negative (N) $\mathcal{L}(2, 4)$
  Zero (Z) $\Delta(3, 6, 9)$
  Positive (PS) $\Gamma(6, 8)$.

- v can be:

  Negative (N) $\mathcal{L}(2, 5)$
  Zero (Z) $\Pi(2, 4, 6, 8)$
  Positive (PS) $\Gamma(0, 8)$.

- w can be:

  Small (S) $\mathcal{L}(2, 4)$
  Medium (M) $\Delta(3, 5, 7)$
  High (H) $\Gamma(6, 8)$.

Rule Base - case of $\wedge$

v \ u	**N**	**Z**	**P**
N	**S**	**S**	**M**
Z	**S**	**M**	**H**
P	**M**	**H**	**H**

Evaluate
w if $u = 4$ & $v = 6$.

Task #2[2]

Design a fuzzy lighting controller system, in which the control system dims the bulb light automatically according to the environmental light. Assume that the inputs to the system are the environmental light x_1 and the changing rate of the environmental light x_2. The output y represents the control value of the dimmer.

- x_1 ranges between 120 and 220 lumens. x_1 can be:

  Dark (D) $\mathcal{L}(130,\ 150)$
  Ambient (A) $\Pi(130,\ 150,\ 190,\ 210)$
  Light (L) $\Gamma(190,\ 210)$.

- x_2 ranges between -10 and $+10$. x_2 can be:

  Negative-Small (NS) $\mathcal{L}(-10,\ 0)$
  Zero (Z) $\Delta(-10,\ 0,\ 10)$
  Positive-Small (PS) $\Gamma(0,\ 10)$.

- y ranges between 0 and +10. *dm* can be:

  Very-Small (VS) $\mathcal{L}(2,\ 4)$
  Small (S) $\Delta(2,\ 4,\ 6)$
  Big (B) $\Delta(4,\ 6,\ 8)$
  Very-Big (VB) $\Gamma(6,\ 8)$.

Rule Base - case of $\wedge$

x_2 \ x_1	D	A	L
NS	VB	B	B
Z	B	B	S
PS	B	S	VS

Evaluate
y if $x_1 = 125$ & $x_2 = -6$.

[2] Credit: Dr. Mohammed A. T.

Fuzzy Inference Systems (FIS)

Sugeno Algorithm

Task #3

Suppose we have three fuzzy predicates: $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ described by these trapezoidal fuzzy sets:

$$\mathcal{A} \quad \Pi\,(0,\ 2,\ 5,\ 9)$$
$$\mathcal{B} \quad \Pi\,(2,\ 8,\ 13,\ 16)$$
$$\mathcal{C} \quad \Pi\,(11,\ 16,\ 19,\ 19)$$

x and y are fuzzy variables, each one ranges between 0 and 19. Given the following three rules:

$$\Re_1 \quad (x \text{ is } \mathcal{A}) \wedge (y \text{ is } \mathcal{C}) \longrightarrow u = 10$$
$$\Re_2 \quad !(x \text{ is } \mathcal{A}) \vee (y \text{ is } \mathcal{B}) \longrightarrow u = 2$$
$$\Re_3 \quad (x \text{ is } \mathcal{B}) \wedge !(y \text{ is } \mathcal{C}) \longrightarrow u = 5$$

Compute the degree of satisfaction for each case:

① $x_1 = 5 \,\&\, y_1 = 12$ ② $x_2 = 0 \,\&\, y_2 = 15$ ③ $x_3 = 7 \,\&\, y_3 = 13$

$\boxed{x_1 = 5 \,\&\, y_1 = 12}$

$\Re_1$

$\mu_{\mathcal{A}}(x_1) \min \mu_{\mathcal{C}}(y_1)$

$1 \min 1/5 = 1/5$

$\Re_2$

$\mu_{\mathcal{A}}(x_1) \max \mu_{\mathcal{B}}(y_1)$

$0 \max 1 = 1$

$\Re_3$

$\mu_{\mathcal{B}}(x_1) \min \mu_{\mathcal{C}}(y_1)$

$1/2 \min 4/5 = 1/2$

$\boxed{u_1 = 3.82}$

$\boxed{x_2 = 0 \,\&\, y_2 = 15}$

$\Re_1$

$\mu_{\mathcal{A}}(x_2) \min \mu_{\mathcal{C}}(y_2)$

$0 \min \_ = 0$

$\Re_2$

$\mu_{\mathcal{A}}(x_2) \max \mu_{\mathcal{B}}(y_2)$

$1 \max 1/3 = 1$

$\Re_3$

$\mu_{\mathcal{B}}(x_2) \min \mu_{\mathcal{C}}(y_2)$

$0 \min \_ = 0$

$\boxed{u_2 = 2}$

$\boxed{x_3 = 7 \,\&\, y_3 = 13}$

$\Re_1$

$\mu_{\mathcal{A}}(x_3) \min \mu_{\mathcal{C}}(y_3)$

$1/2 \min 2/5 = 2/5$

$\Re_2$

$\mu_{\mathcal{A}}(x_3) \max \mu_{\mathcal{B}}(y_3)$

$1/2 \max 1 = 1$

$\Re_3$

$\mu_{\mathcal{B}}(x_3) \min \mu_{\mathcal{C}}(y_3)$

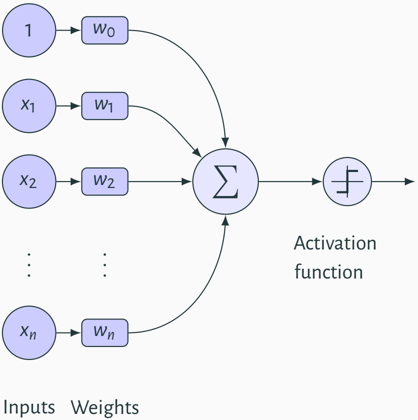
$5/6 \min 3/5 = 3/5$

$\boxed{u_3 = 4.5}$

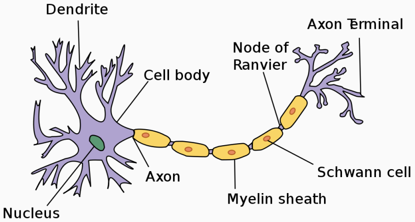
Neural Networks

Fundamental unit of a neural network (1/3)

Artificial neuron



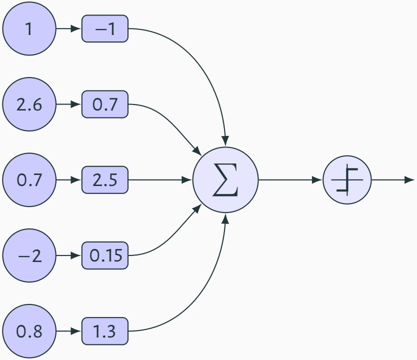
Biological neuron

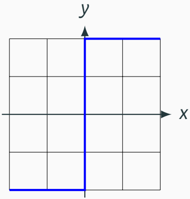


https://id.wikipedia.org/wiki/Sel_saraf

FUNDAMENTAL UNIT OF A NEURAL NETWORK (2/3)

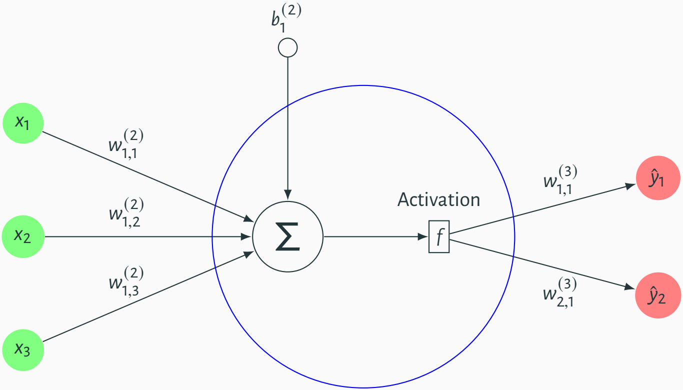
Task #4
Compute the output of the following neuron.



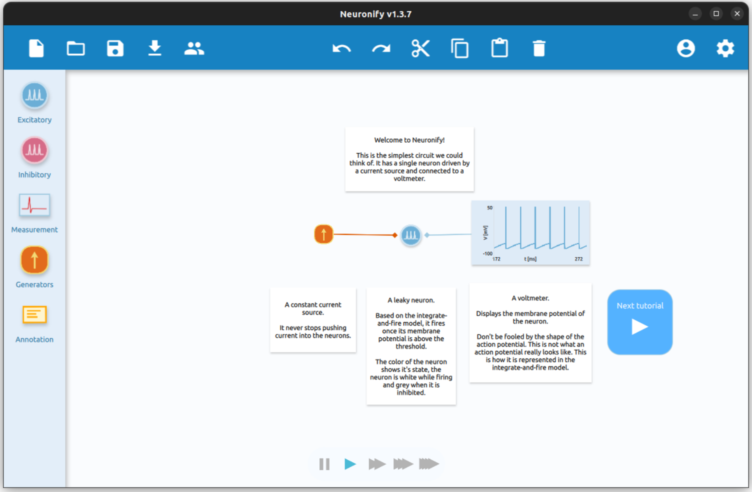


$$y \;=\; \mathrm{sign}(1 \times -1 + 2.6 \times 0.7 + 0.7 \times 2.5 - 2 \times 0.15 + 0.8 \times 1.3) \;=\; 1$$

Fundamental unit of a neural network (3/3)

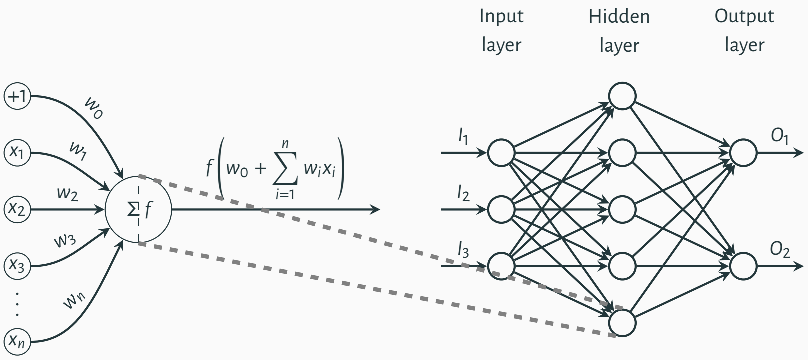


Neural simulation



http://ovilab.net/neuronify/

Multilayer Perceptron (MLP)



Task #5
For the above structure, determine how many parameters are to be adjusted.

$$\text{Params \#} \;=\; 5 \times 3 + 5 + 2 \times 5 + 2 \;=\; 32$$

✓ Design a structure

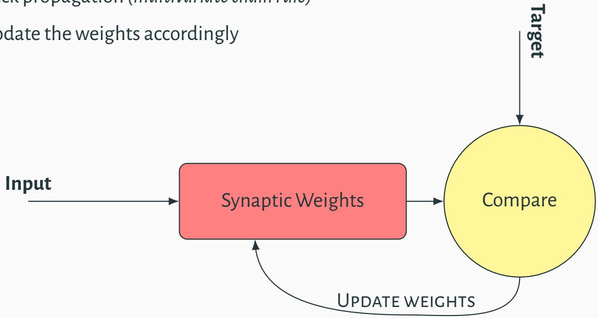
✓ Specify a loss function to minimize

✓ Optimize using gradient descent

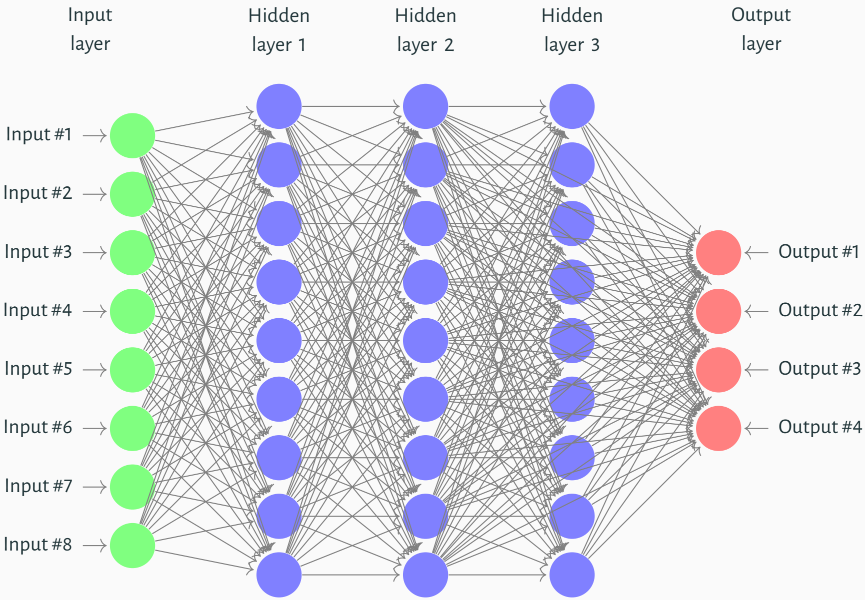
① Feedforward propagation *(matrix multiplication and point-wise activation)*

② Back propagation *(multivariate chain rule)*

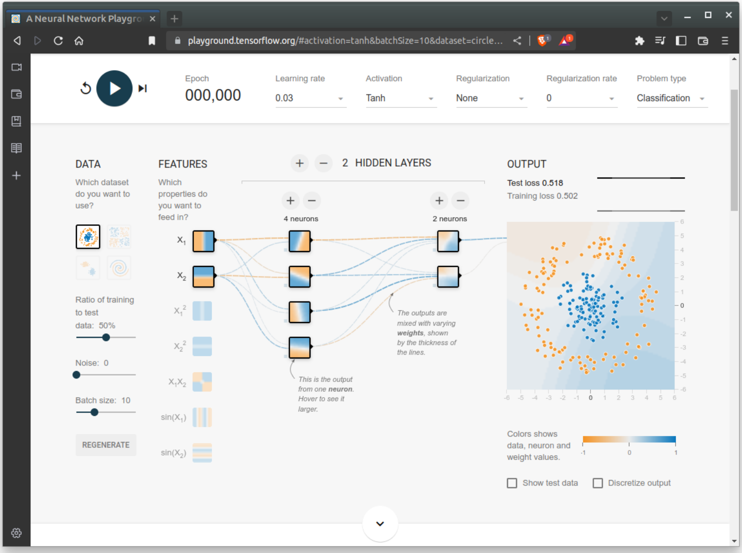
③ Update the weights accordingly



Multilayer Perceptron (MLP)

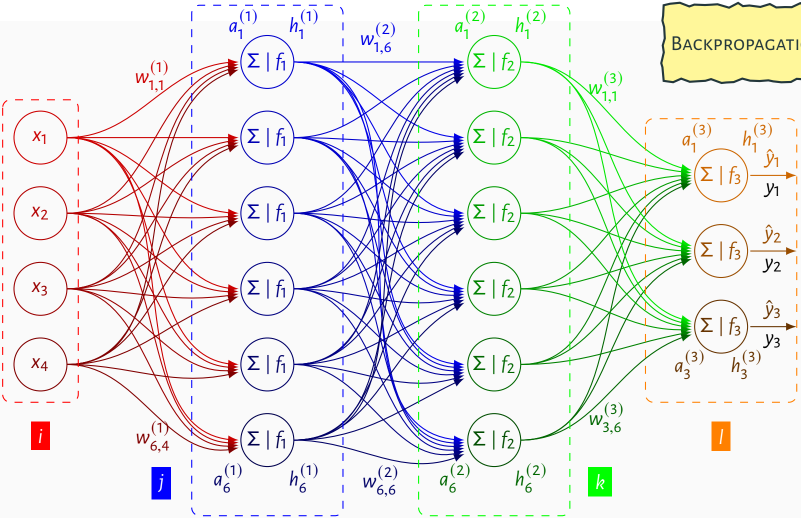


Tinker with a neural network



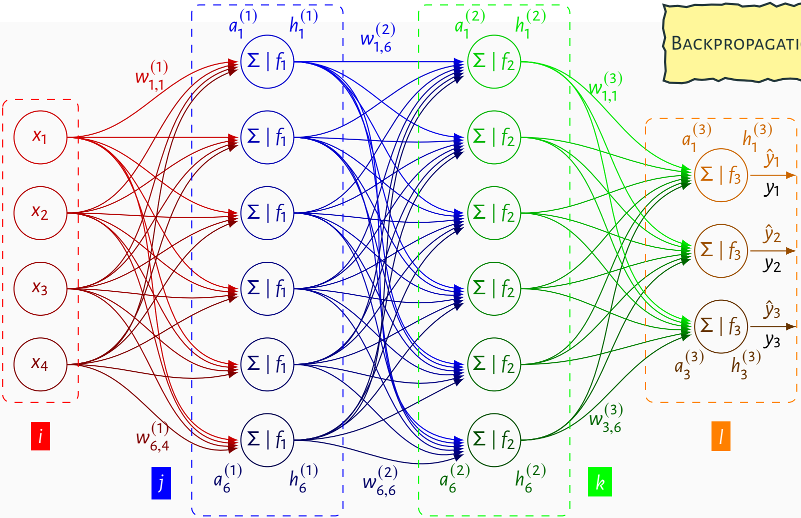
https://playground.tensorflow.org/

Backpropagation



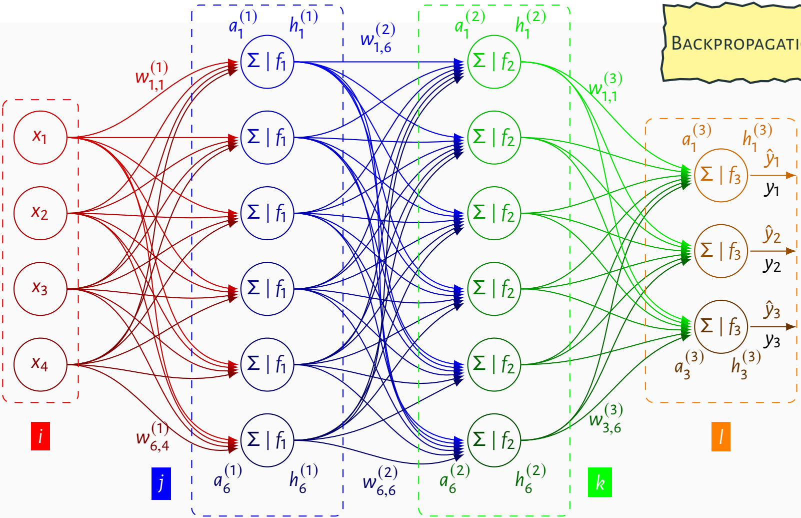
$$\mathcal{J}(\mathcal{W}) = \frac{1}{2}\sum_{l}(y_l - \hat{y}_l)^2 \implies \frac{\partial \mathcal{J}}{\partial w_{l,k}^{(3)}} = -\underbrace{(y_l - \hat{y}_l)\,\dot{f}_3\left(a_l^{(3)}\right)}_{\delta_l^{(3)}} h_k^{(2)}$$

Backpropagation

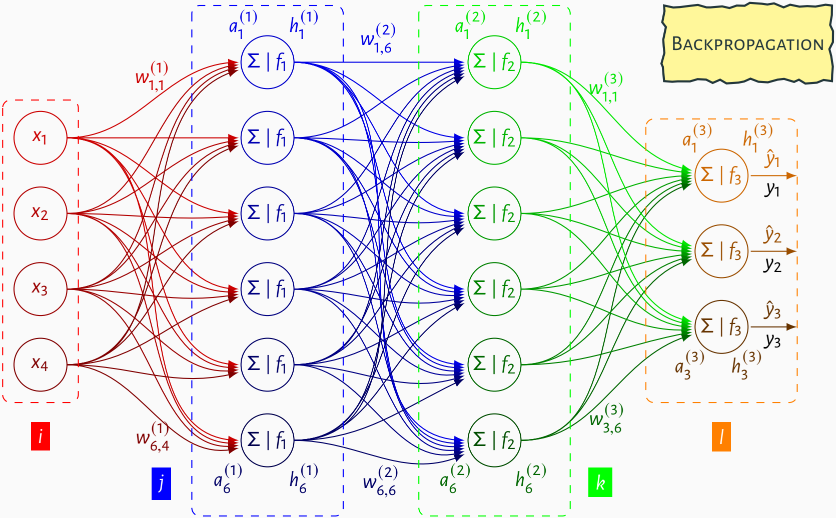


$$\frac{\partial \mathcal{J}}{\partial w^{(2)}_{k,j}} = -\sum_l (y_l-\hat{y}_l)\,\dot{f}_3\left(a^{(3)}_l\right) w^{(3)}_{l,k}\,\dot{f}_2\left(a^{(2)}_k\right) h^{(1)}_j = -\underbrace{\sum_l \delta^{(3)}_l\, w^{(3)}_{l,k}\,\dot{f}_2\left(a^{(2)}_k\right)}_{\delta^{(2)}_k} h^{(1)}_j$$

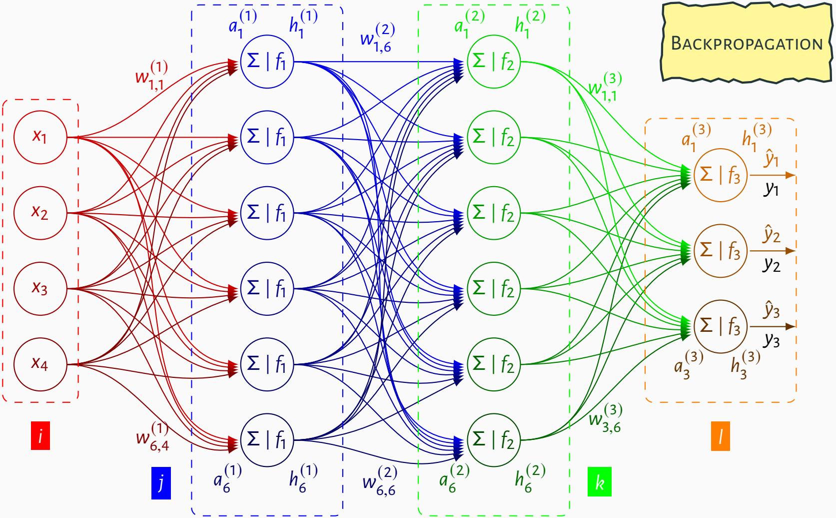
Backpropagation



$$\frac{\partial \mathcal{J}}{\partial w^{(1)}_{j,i}} = -\sum_l (y_l - \hat{y}_l)\, \dot{f}_3\left(a^{(3)}_l\right) \sum_k w^{(3)}_{l,k}\, \dot{f}_2\left(a^{(2)}_k\right) w^{(2)}_{k,j}\, \dot{f}_1\left(a^{(1)}_j\right) x_i = -\underbrace{\sum_k \delta^{(2)}_k w^{(2)}_{k,j}\, \dot{f}_1\left(a^{(1)}_j\right)}_{\delta^{(1)}_j} x_i$$

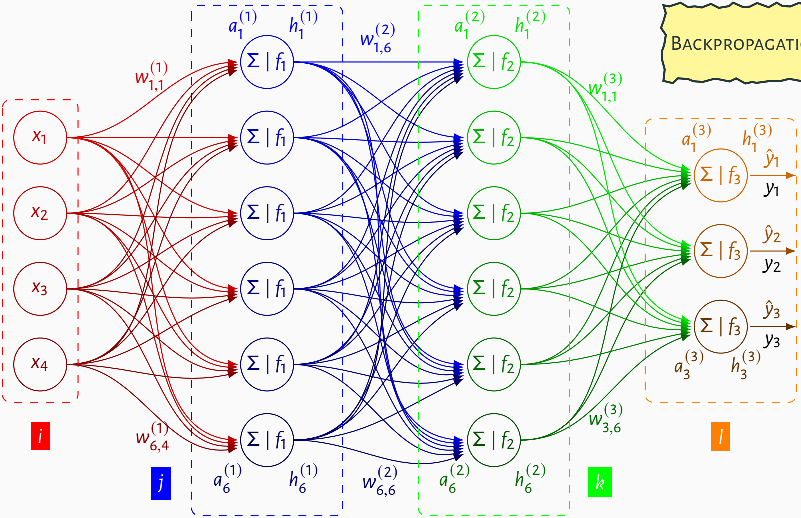


$$\boxed{\delta_l^{(3)} = (y_l - \hat{y}_l) \times \dot{f}_3\left(a_l^{(3)}\right) \implies \Delta\omega_{l,k}^{(3)} = -\eta\delta_l^{(3)} \times h_k^{(2)}}$$



$$\boxed{\delta^{(2)}_k = \left(\delta^{(3)}_1 w^{(2)}_{1,k} + \delta^{(3)}_2 \omega^{(2)}_{2,k} + \delta^{(3)}_3 \omega^{(2)}_{3,k}\right) \times \dot{f}_2\left(a^{(2)}_k\right) \implies \Delta\omega^{(2)}_{k,j} = -\eta\delta^{(2)}_k \times h^{(1)}_j}$$

Backpropagation



$$\delta_j^{(1)} = \left(\delta_1^{(2)} w_{1,j}^{(1)} + \cdots + \delta_6^{(2)} \omega_{6,j}^{(1)}\right) \times \dot{f}_2\left(a_j^{(2)}\right) \implies \Delta\omega_{j,i}^{(1)} = -\eta\delta_j^{(2)} \times x_i^{(1)}$$

Multivariate chain rule

Output layer ⟶ hidden layer #2

$$\frac{\partial \hat{y}_l}{\partial w_{l,k}^{(3)}} = \underbrace{\frac{\partial \hat{y}_l}{\partial a_l^{(3)}}}_{\dot{f}_3\left(a_l^{(3)}\right)} \underbrace{\frac{\partial a_l^{(3)}}{\partial w_{l,k}^{(3)}}}_{h_k^{(2)}}$$

Output layer ⟶ hidden layer #1

$$\frac{\partial \hat{y}_l}{\partial w_{k,j}^{(2)}} = \underbrace{\frac{\partial \hat{y}_l}{\partial a_l^{(3)}}}_{\dot{f}_3\left(a_l^{(3)}\right)} \underbrace{\frac{\partial a_l^{(3)}}{\partial h_k^{(2)}}}_{w_{l,k}^{(3)}} \underbrace{\frac{\partial h_k^{(2)}}{\partial a_k^{(2)}}}_{\dot{f}_2\left(a_k^{(2)}\right)} \underbrace{\frac{\partial a_k^{(2)}}{\partial w_{k,j}^{(2)}}}_{h_j^{(1)}}$$

Output layer ⟶ input layer

$$\frac{\partial \hat{y}_l}{\partial w_{j,i}^{(1)}} = \underbrace{\frac{\partial \hat{y}_l}{\partial a_l^{(3)}}}_{\dot{f}_3\left(a_l^{(3)}\right)} \underbrace{\frac{\partial a_l^{(3)}}{\partial h_k^{(2)}}}_{w_{l,k}^{(3)}} \underbrace{\frac{\partial h_k^{(2)}}{\partial a_k^{(2)}}}_{\dot{f}_2\left(a_k^{(2)}\right)} \underbrace{\frac{\partial a_k^{(2)}}{\partial h_j^{(1)}}}_{w_{k,j}^{(2)}} \underbrace{\frac{\partial h_j^{(1)}}{\partial a_j^{(1)}}}_{\dot{f}_1\left(a_j^{(1)}\right)} \underbrace{\frac{\partial a_j^{(1)}}{\partial w_{j,i}^{(1)}}}_{x_i}$$

List of available optimizers (1/2)

Here is a list of some common optimizers for artificial neural networks:

$$\boxed{\Delta\hat{\mathcal{W}} \triangleq \mathcal{F}\left(\nabla \underbrace{\mathcal{J}\left(\hat{\mathcal{W}}\right)}_{\text{Loss Function}}\right)} \equiv \boxed{\hat{\mathcal{W}} \triangleq \hat{\mathcal{W}} + \mathcal{F}\left(\nabla\mathcal{J}\left(\hat{\mathcal{W}}\right)\right)} \quad \nabla\mathcal{J}\left(\hat{\mathcal{W}}\right) = \begin{bmatrix} \frac{\partial\mathcal{J}}{\partial\hat{w}_0} \\ \vdots \\ \frac{\partial\mathcal{J}}{\partial\hat{w}_n} \end{bmatrix}$$

Stochastic Gradient Descent (SGD)

$$\hat{\mathcal{W}} \triangleq \hat{\mathcal{W}} - \eta\nabla\mathcal{J}\left(\hat{\mathcal{W}}\right)$$

Mini-batch Gradient Descent

$$\hat{\mathcal{W}} \triangleq \hat{\mathcal{W}} - \frac{\eta}{m}\nabla\sum_{i=1}^{m}\mathcal{J}\left(\hat{\mathcal{W}}\right)$$

⟵ m denotes the size of the mini-batch

Momentum

$$\hat{\mathcal{W}} \triangleq \hat{\mathcal{W}} - \mathcal{V}, \quad \text{where} \quad \mathcal{V} \triangleq \alpha\mathcal{V} + \eta\nabla\mathcal{J}\left(\hat{\mathcal{W}}\right)$$

Nesterov Accelerated Gradient (NAG)

$$\hat{\mathcal{W}} \triangleq \hat{\mathcal{W}} - \mathcal{V} \quad \text{where} \quad \mathcal{V} \triangleq \alpha\mathcal{V} + \eta\nabla\mathcal{J}\left(\hat{\mathcal{W}} - \alpha\mathcal{V}\right)$$

List of available optimizers (2/2)

AdaGrad

$$\hat{\mathcal{W}} \triangleq \hat{\mathcal{W}} - \frac{\eta}{\sqrt{\mathcal{G}+\epsilon}} \nabla \mathcal{J}\left(\hat{\mathcal{W}}\right) \quad \text{where} \quad \mathcal{G} \triangleq \mathcal{G} + \left(\nabla \mathcal{J}\left(\hat{\mathcal{W}}\right)\right)^2$$

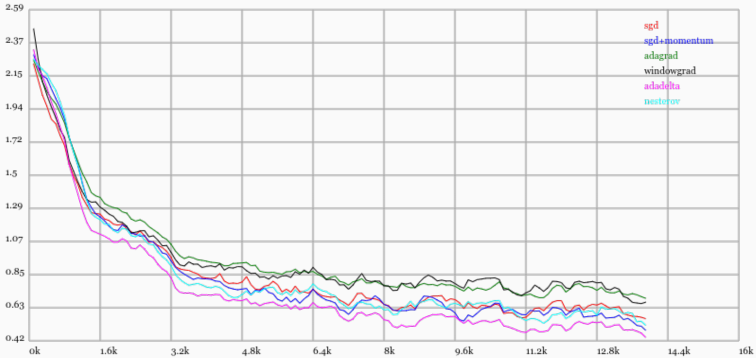
RMSProp

$$\hat{\mathcal{W}} \triangleq \hat{\mathcal{W}} - \frac{\eta}{\sqrt{\mathcal{G}+\epsilon}} \nabla \mathcal{J}\left(\hat{\mathcal{W}}\right) \quad \text{where} \quad \mathcal{G} \triangleq \mathcal{G} + (1-\beta)\left(\nabla \mathcal{J}\left(\hat{\mathcal{W}}\right)\right)^2$$

Adam

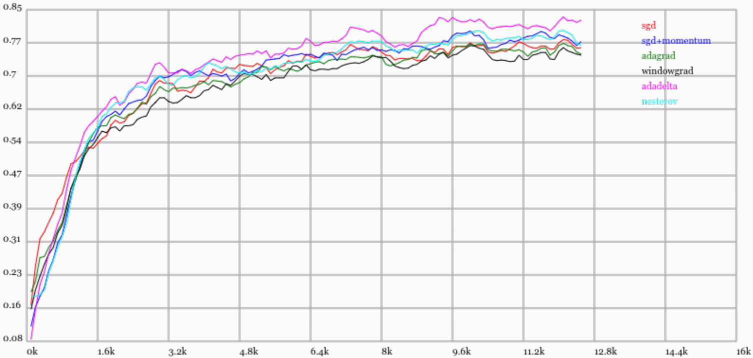
$$\begin{aligned}
\mathcal{M} &\triangleq \beta_1 \mathcal{M} + (1-\beta_1) \nabla \mathcal{J}\left(\hat{\mathcal{W}}\right) && \longleftarrow \text{Estimate of first moment} \\
\mathcal{V} &\triangleq \beta_2 \mathcal{V} + (1-\beta_2) \left(\nabla \mathcal{J}\left(\hat{\mathcal{W}}\right)\right)^2 && \longleftarrow \text{Estimate of second moment} \\
\hat{\mathcal{M}} &= \frac{\mathcal{M}}{1-\beta_1^k} && \longleftarrow \text{@ every } k^{\text{th}} \text{ iteration} \\
\hat{\mathcal{V}} &= \frac{\mathcal{V}}{1-\beta_2^k} && \longleftarrow \text{@ every } k^{\text{th}} \text{ iteration} \\
\hat{\mathcal{W}} &\triangleq \hat{\mathcal{W}} - \frac{\eta}{\sqrt{\hat{\mathcal{V}}+\epsilon}} \hat{\mathcal{M}}
\end{aligned}$$

Effect of optimizer on loss values



https://cs.stanford.edu/people/karpathy/convnetjs/demo/trainers.html

Effect of optimizer on testing accuracy values



https://cs.stanford.edu/people/karpathy/convnetjs/demo/trainers.html

Frameworks to be used

https://juliapackages.com/p/mlj

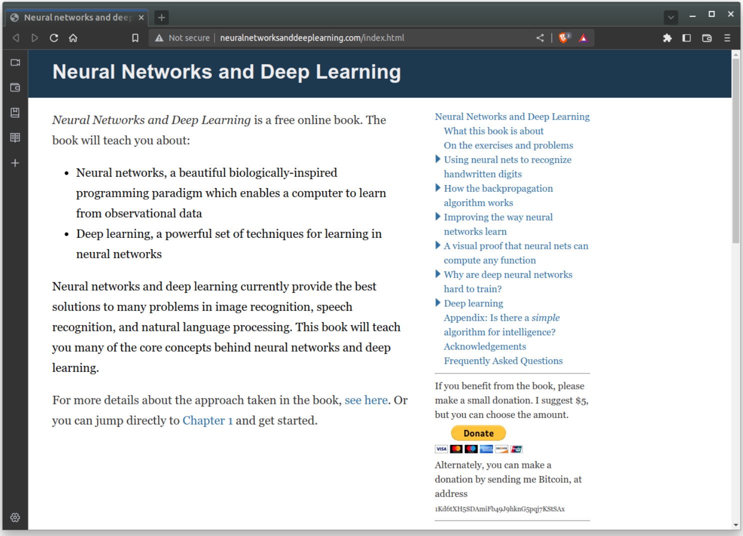
https://juliapackages.com/p/flux

ANN

CODE SNIPPET

Code is available at https://github.com/a-mhamdi/jlai/
⟶ *Codes* ⟶ *Julia* ⟶ *Part-1* ⟶ *xor-gate.jl*

To be continued in the following e-book



Neural Networks and Deep Learning

Neural Networks and Deep Learning is a free online book. The book will teach you about:

- Neural networks, a beautiful biologically-inspired programming paradigm which enables a computer to learn from observational data
- Deep learning, a powerful set of techniques for learning in neural networks

Neural networks and deep learning currently provide the best solutions to many problems in image recognition, speech recognition, and natural language processing. This book will teach you many of the core concepts behind neural networks and deep learning.

For more details about the approach taken in the book, see here. Or you can jump directly to Chapter 1 and get started.

If you benefit from the book, please make a small donation. I suggest $5, but you can choose the amount.

Donate

Alternately, you can make a donation by sending me Bitcoin, at address 1Kd6tXH5SDAmiFb49J9hknG5pqj7KStSAx

http://neuralnetworksanddeeplearning.com/

Quizzes

MCQ (1/4)

1. ... is a part of machine learning that works with neural networks.
   - × Artificial intelligence
   - ✓ Deep learning
   - × All of the above
   - × None of the above
2. In training a neural network, we notice that the loss does not increase in the first few starting epochs: What is the reason for this?
   - × The learning rate is low
   - × Regularization parameter is high
   - × Stuck at the local minima
   - ✓ All of the above
3. What is the use of validation dataset in Machine Learning?
   - × To train the machine learning model.
   - ✓ To tune the hyperparameters of the machine learning model
   - × To evaluate the performance of the machine learning model
   - × None of the above

MCQ (2/4)

4. Which of the following is true about model capacity
   (where model capacity means the ability of neural network to approximate complex functions)?
   - ✓ As number of hidden layers increases, model capacity increases
   - × As dropout ratio increases, model capacity increases
   - × As learning rate increases, model capacity increases
   - × None of these
5. Overfitting is a type of modelling error which results in the failure to predict future observations effectively or fit additional data in the existing model. Yes/No?
   - ✓ Yes
   - × No
   - × Can not say
   - × Probably
6. ... is the scenario when the model fails to decipher the underlying trend in the input data.
   - ✓ Underfitting
   - × Overfitting
   - × All of the above
   - × None of the above

MCQ (3/4)

7. The average positive difference between computed and desired outcome values
   - ✗ Root Mean Squared Error
   - ✗ Mean Squared Error
   - ✗ Mean Absolute Error
   - ✓ Mean Positive Error
8. ... is used as an input to the machine learning model for training and prediction purposes.
   - ✗ Target variable
   - ✓ Feature vector
   - ✗ All of the above
   - ✗ None of the above
9. The correlation between the number of years an emplyee has worked for a company and the salary of the employee is 0.75. What can be said about emplyee salary and years worked?
   - ✗ There is no relationship between salary and years worked.
   - ✓ Individuals that have worked for the company the longest have higher salaries.
   - ✗ Individuals that have worked for the company the longest have lower salaries.
   - ✗ The majority of employees have been with the company a long time.

MCQ (4/4)

10. As the amount of training data increases
    - × Training error usually increases and generalization error usually increases
    - ✓ Training error usually increases and generalization error usually decreases
    - × Training error usually decreases and generalization error usually decreases
    - × Training error usually decreases and generalization error usually increases

Some Useful Links

1. https://setosa.io/ev/
2. https://karpathy.ai/
3. http://yann.lecun.com/
4. https://www.hackingnote.com/
5. https://machinelearningmastery.com/
6. https://stanford.edu/ shervine/teaching/
7. https://www.ibm.com/downloads/cas/GB8ZMQZ3
8. https://colah.github.io/posts/2014-03-NN-Manifolds-Topology/

References

[Bel78] R. E. Bellman. *An Introduction to Artificial Intelligence: Can Computers Think?* Boyd & Fraser Publishing Company, Jan. 1, 1978 (cit. on p. 9).

[CMM85] E. Charniak, D. McDermott, and D. V. McDermott. *Introduction to Artificial Intelligence*. Addison-Wesley series in computer science and information processing. Addison-Wesley, 1985 (cit. on p. 10).

[Dad12] E. Dadios, ed. *Fuzzy Logic - Controls, Concepts, Theories and Applications*. IntechOpen, Mar. 28, 2012. 430 pp.

[ENM15] I. El Naqa and M. J. Murphy. "What Is Machine Learning?" In: *Machine Learning in Radiation Oncology: Theory and Applications*. Ed. by I. El Naqa, R. Li, and M. J. Murphy. Cham: Springer International Publishing, 2015, pp. 3–11. DOI: 10.1007/978-3-319-18305-3_1.

[Gac15] L. Gacôgne. *Intelligence artificielle. Cours, exercices corrigés et projets*. Ed. by Ellipses. Nov. 2015. 240 pp.

[GBC16] I. Goodfellow, J. Bengio, and A. Courville. *Deep Learning*. MIT Press Ltd, Nov. 18, 2016. 800 pp.

Further Reading (2/3)

[Hau89] J. Haugeland. *Artificial Intelligence: The Very Idea*. A Bradford book. MIT Press, 1989 (cit. on p. 9).

[JPM21] L. M. John Paul Mueller. *Machine Learning For Dummies*. Wiley John + Sons, Apr. 8, 2021. 464 pp.

[Kur92] R. Kurzweil. *The Age of Intelligent Machines*. Viking, 1992 (cit. on p. 11).

[LS93] G. F. Luger and W. A. Stubblefield. *Artificial Intelligence: Structures and Strategies for Complex Problem Solving*. Artificial intelligence. Benjamin/Cummings Publishing Company, 1993 (cit. on p. 12).

[Mit97] T. M. Mitchell. *Machine Learning*. McGraw-Hill International Editions. McGraw-Hill, 1997.

[RK91] E. Rich and K. Knight. *Artificial Intelligence*. Artificial Intelligence Series. McGraw-Hill, 1991 (cit. on p. 11).

[Rob14] A. Robinson. *Construction Informatics*. Ed. by H. S. of Construction Engineering. 2014.

[Sch90] R. J. Schalkoff. *Artificial Intelligence: An Engineering Approach*. McGraw-Hill Computer science series. McGraw-Hill, 1990 (cit. on p. 12).

[SNK12] T. Sai, D. Nakhaeinia, and B. Karasfi. "Application of Fuzzy Logic in Mobile Robot Navigation". In: *Fuzzy Logic - Controls, Concepts, Theories and Applications*. InTech, Mar. 2012. DOI: 10.5772/36358.

Further Reading (3/3)

[Win92] P. H. Winston. *Artificial Intelligence*. A-W Series in Computer Science. Addison-Wesley Publishing Company, 1992 (cit. on p. 10).

[Woj12] J. Wojtusiak. "Machine Learning". In: *Encyclopedia of the Sciences of Learning*. Springer US, 2012, pp. 2082–2083. DOI: 10.1007/978-1-4419-1428-6_1927.