IMTech
FME & CFIS

Artificial intelligences and Mathematics

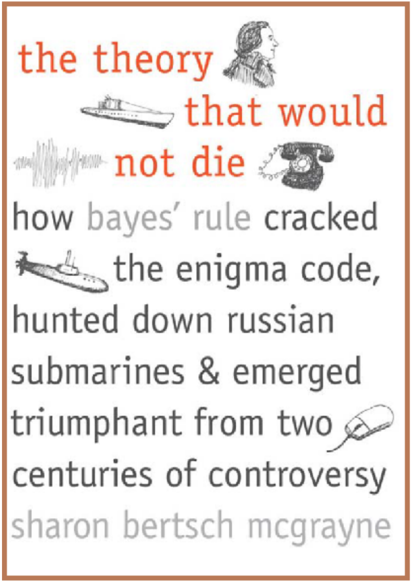
S. Xambó

IMTech & BSC

9/11/2022

Preface

https://web.mat.upc.edu/sebastia.xambo/99/22-11-09.pdf



[1] (mcgrayne-2011) [2] (silver-2012)

Learning from experience is neatly encoded, in probabilistic terms, by the *Bayes-Laplace rule*:

$$P(x|y) = P(x)P(y|x)/P(y) = P(x)K(x,y),$$

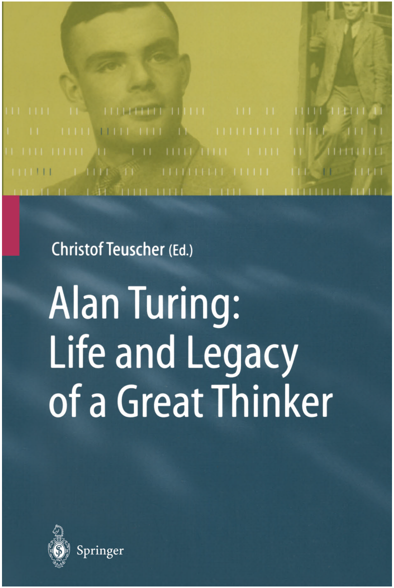
where $K(x,y) = P(y|x)/P(y) = P(x \cap y)/P(x)P(y)$ is symmetric.

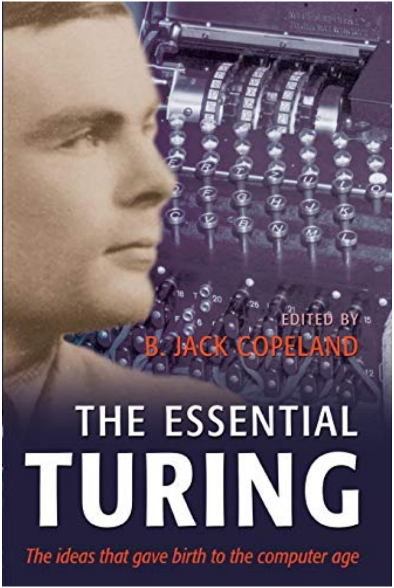
This tells us how to update our *prior* belief in x, $P(x)$, to the belief $P(x|y)$ *posterior* to having observed y.

When $K > 1$ ($K < 1$), our believe is *increased* (*decreased*), and in any case it may be construed as a *learning* on x produced by the observation of y.

The Bayes–Laplace rule is the basis of many sorts of AI models (we provide a few references on page 11 and on page 15).

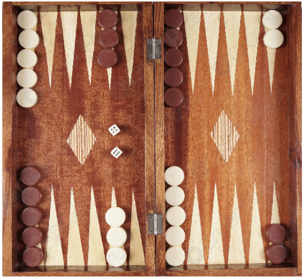
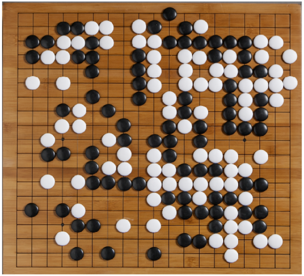
In practice, the trouble arises from the evaluation of $P(y)$, the *evidence* of y, as $P(y) = \sum_x P(y|x)P(x)$, and this sum (or integral in continuous models) is often intractable.

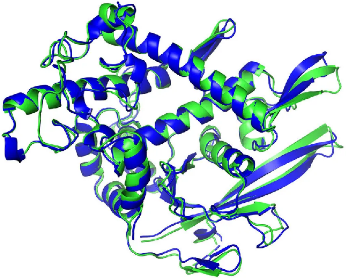




[3] (teuscher-2004) [4] (copeland-2004)

AIs are blowing in the wind



Chess, Backgammon, Go, Console games, Protein folding, ...

Beethoven's Unfinished 10th Symphony Brought to Life by Artificial Intelligence (Scientific American, October 15, 2021).
See Classic *f*M. Sample.

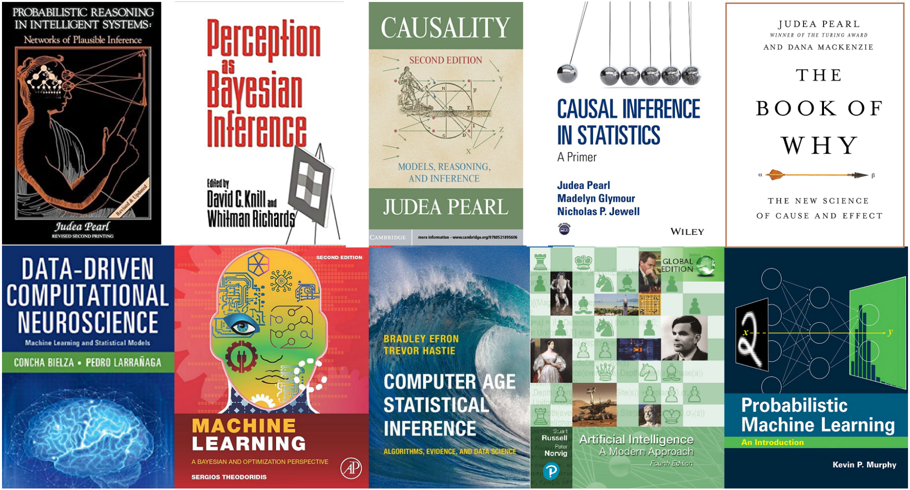
Walter Werzowa: "I dare to say that nobody knows Beethoven as well as the AI did—as well as the algorithm". "I think music, when you hear it, when you feel it, when you close your eyes, it does something to your body. Close your eyes, sit back and be open for it, and I would love to hear what you felt after".

We will see more examples later.

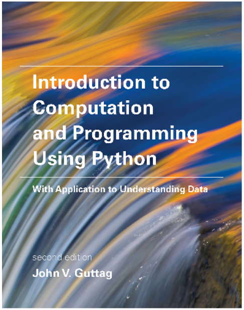

Neural Networks and Learning Machines
Third Edition
Simon Haykin
Shai Shalev-Shwartz and Shai Ben-David
UNDERSTANDING MACHINE LEARNING
FROM THEORY TO ALGORITHMS
DEEP LEARNING
Ian Goodfellow, Yoshua Bengio, and Aaron Courville
Foundations of Machine Learning second edition
Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar
Gopinath Rebala · Ajay Ravi Sanjay Churiwala
An Introduction to Machine Learning
Springer
INTRODUCTION TO Machine Learning
fourth edition
ETHEM ALPAYDIN
Learning Theory from First Principles
DRAFT
October 13, 2022
Francis Bach
francis.bach@inria.fr
THE PRINCIPLES OF DEEP LEARNING THEORY
An Effective Theory Approach to Understanding Neural Networks
Daniel A. Roberts and Sho Yaida
based on research in collaboration with Boris Hanin

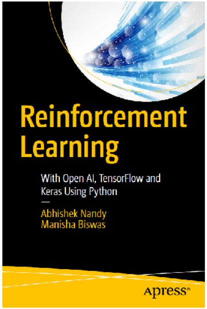
General

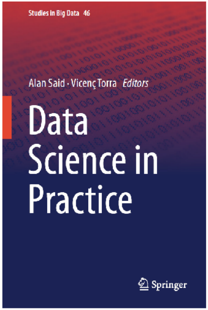
N

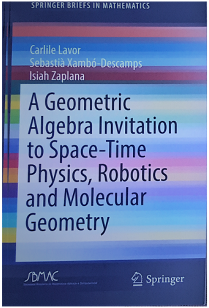


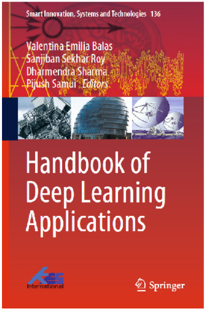
Bayesian approaches

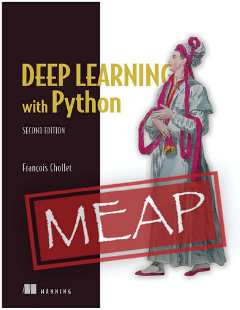


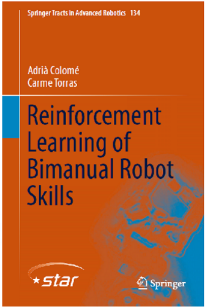


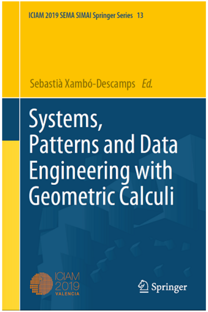












Applications

N

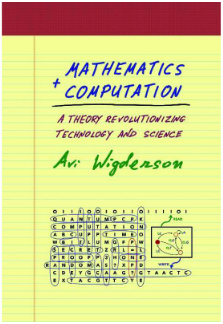
Many facets of the symbiosis *Mathematics & Computation* are presented in [5] (wigderson-2019).

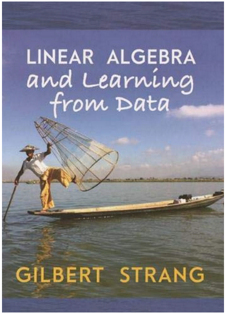
In particular, Chapter 17 is devoted to *computational learning theory*.

Another useful reference is [6] (strang-2019).

See also the extensive *survey* [7] (nguyen-dlugolinsky-bobak-et-5-2019) and the popular approach [8] (Marr's blog).

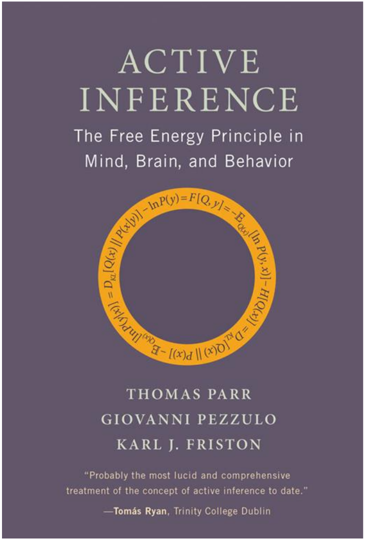
And, of course, articles in the Wikipedia, such as `Machine_learning`, keep growing in number and substance.





How do they work?

Bayesian brain
Clustering methods
Supervised learning
Reinforcement learning



[9] (parr-pezzulo-friston-2022)

- $H(Q(x))$: $-\sum_x Q(x) \ln Q(x)$
  (*Entropy* of $Q(x)$)
- $KL(Q(x)||P(x))$: $\sum_x Q(x) \ln \frac{P(x)}{Q(x)}$
  (*Kullback-Leibler divergence*)
- $F(Q,y)$: *Free energy*

The golden ring

$$
\begin{aligned}
F(Q,y) &= \underbrace{-\mathbb{E}_{Q(x)}[\ln P(y,x)]}_{\textit{Energy}} - \underbrace{H[Q(x)]}_{\textit{Entropy}} \\
&= \underbrace{KL[Q(x)||P(x)]}_{\textit{Complexity}} - \underbrace{\mathbb{E}_{Q(x)}[\ln P(y|x)]}_{\textit{Accuracy}} \\
&= \underbrace{KL[Q(x)||P(x|y)]}_{\textit{Divergence}} - \underbrace{\ln P(y)}_{\textit{Evidence}}
\end{aligned}
$$

A shorter account is provided by the tutorial paper [10] (smith-friston-whyte-2022) on "active inference and its application to empirical data".

Aimed at finding *hidden structure in data*, $\mathcal{D} = \{x^1, \dots, x^m\}$.

k-**Means.** This algorithm groups unlabeled data $\mathcal{D}$ in k classes:

(1) Select k vectors $z^1, \cdots, z^k \in \mathcal{D}$ at random.

(2) Assign each $x^j \in \mathcal{D}$ to the first z^i nearest to x^j (initial groups).

(3) Update each z^i to the centroid (or *mean*) of the z^i group.

(4) Iterate (2) and (3) until the z^i are stable (up to a *tolerance*).

The associated *cluster predictor* assigns x to the first nearest z^i.

k-**NN** (nearest neighbors). Let $\mathcal{D} = \{(x^1, y^1), \cdots, (x^m, y^m)\}$ be a labeled set and k a positive integer. The *label predictor* of the k-NN algorithm assigns a vector x to the mode of $y^{j_1}, \cdots, y^{j_k}$, where $x^{j_1}, \cdots, x^{j_k}$ are the nearest neighbors of x from among $x^1, \cdots, x^m$.

Hypothesis space[3]

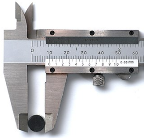
$$\mathcal{H} = \{h\}$$

Data, expert tables[2]

$$\mathcal{D}$$
$$\{(x^1, y^1), \ldots, (x^m, y^m)\}$$
$$x^j \sim P, \quad y^j = f^*(x^j)$$

World – Observer[1]

Error measure[4]



Learner/LA[5]

$\Longrightarrow$

Prediction rule[6]

$$y = h(x)$$

[1] *Urbi et Orbi* engraving (Flammarion). Observer produces data (**Tycho Brahe**).
[2] Tables of planet and comet positions over time (**Ephemeris**)
[3] Hypothesis space (*Inductive bias*). Greeks/Copernicus: circles around the Earth/Sun.
[4] Error measure (*loss*, *risk*, *regret*): How close are predictions to observations?
[5] Learning algorithm (**Kepler**). New bias: Ellipses with a focus at the Sun.
[6] The algorithm supplies a predicton h. Hopefully, $h \approx f^*$.

Set of data $x^j \in \mathbf{R}^n$, $j \in [m]$ (*dataset*).

Want to predict values y^j provided by a *supervisor* or *expert* in such a way that for objects x not in the dataset the value y corresponding to x is predicted with high probability (*generalization* capacity).

Hypothesis space: a space of parameterized functions,

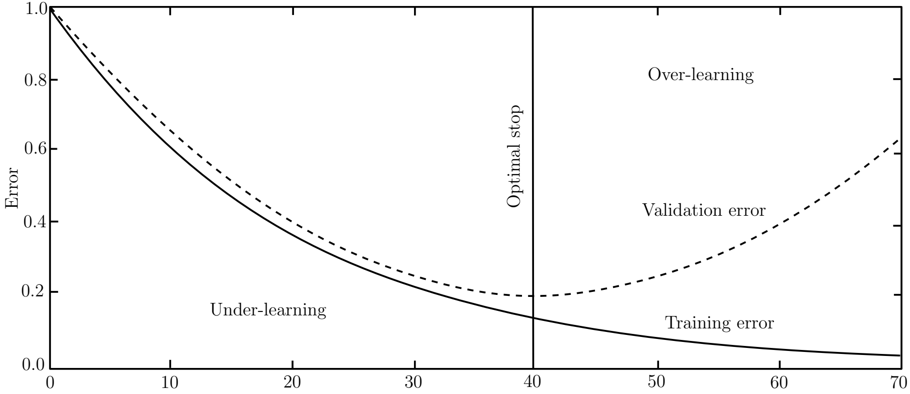
$$\mathcal{H} = \{h_w\}_{w \in W}.$$

Problem: to find $w \in W$ such that $h_w(x^j) \approx y^j$.

Method: If the criterion for fitness depends on a function $\ell(h(x), y)$ (*local cost*), the problem amounts to finding w that minimizes $\sum_j \ell(f_w(x^j), y^j)$ (*empirical cost* or *risk*):

$$\mathrm{argmin}_w \sum_j \ell(h_w(x^j), y^j).$$

This is the *empirical risk minimization* rule, ERM.



Scheme of a basic training process. In one *epoch* (one step in the training loop), f_w is applied to the dataset. The proportion of errors is the *training error* for that epoch. The learning algorithm determines an update $w = w - \Delta w$ (for example with some variant of gradient descent) and a new epoch is run. The *validation error* is the error incurred by f_w on a *validation dataset*. The training error is decreasing, while the validation error decreases until the *optimal stop* and thereafter it increases.

Supervised learning has two main modalities:

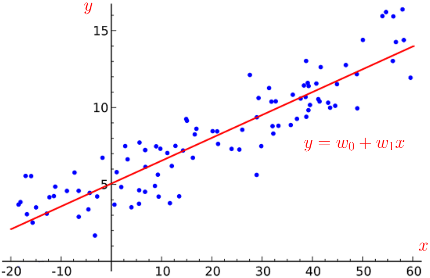
Classification: When the set $\mathcal{Y}$ is finite. In this case its elements are usually called *labels* or *classes*.

Regression: When the set $\mathcal{Y}$ is the set $\mathbf{R}$ of real numbers.

Example. In *linear regression*, $\mathcal{H}$ is the space of functions of the form

$$h(x) = w_0 + w_1x_1 + \cdots + w_nx_n$$

and the local cost is usually $(h(x) - y)^2$.



$\mathcal{D} = \{(x^1, y^1), \ldots, (x^m, y^m)\}$ $(x^j, y^j \in \mathbf{R})$.

Problem. Find a polynomial map $p : \mathbf{R} \to \mathbf{R}$ of degree r,

$$p(x) = w_0 + w_1 x + \cdots w_r \cdot x^r, \quad w_0, w_1, \cdots, w_r \in \mathbf{R},$$

such that $\hat{y}^j = p(x^j)$ are as close as possible to the y^j (*polynomial approximation of degree r*).

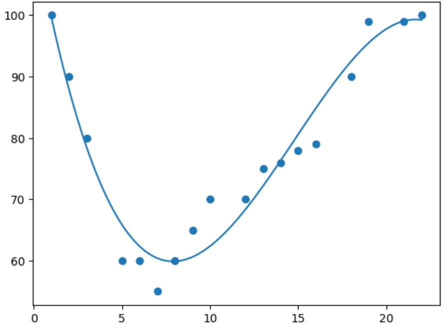


Figure 6.1: Cubic approximation of a dataset in $\mathbf{R} \times \mathbf{R}$.

The *logistic regression* is linear regression of $\log p/(1-p)$.

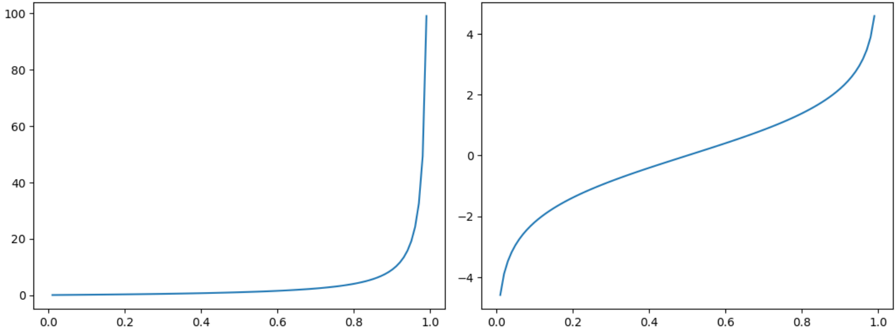


Figure 6.2: For probability values $p \in [0,1]$ it makes no sense to apply linear regression procedures. Left: graph of the *odds* function, $p/(1-p)$, for $p \in [0,1]$. Right: graph of $\log(p/1-p)$, with symmetry about the point $(1/2, 0)$, so linear regression of its values is in principle possible.

If $\log(p/1-p) = w \cdot x$ $(x, w \in \mathbf{R}^n)$, then $p = p(x) = 1/(1+e^{-w\cdot x})$ estimates the probability of observations x.

The function $1/(1+e^{-t})$ is the *logistic function*. Its range is $(0,1)$. A variation is the function $(1-e^{-t})/(1+e^{-t})$, with range $(-1,1)$.

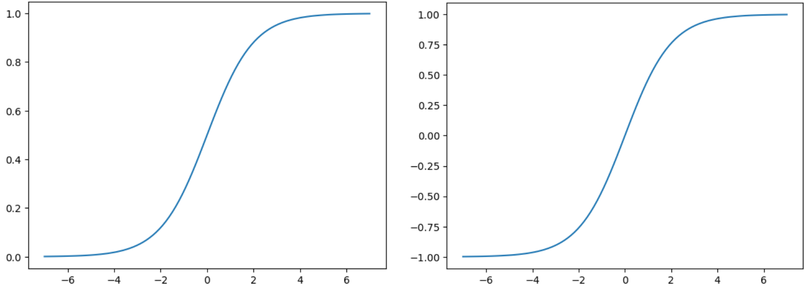
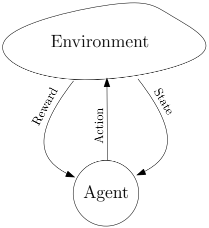
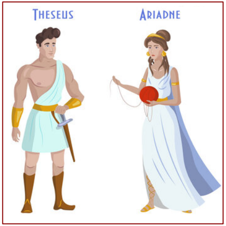
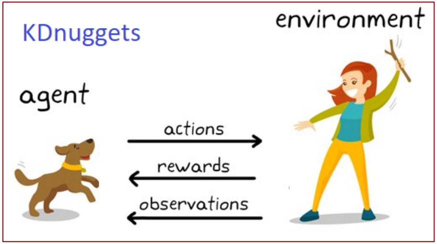


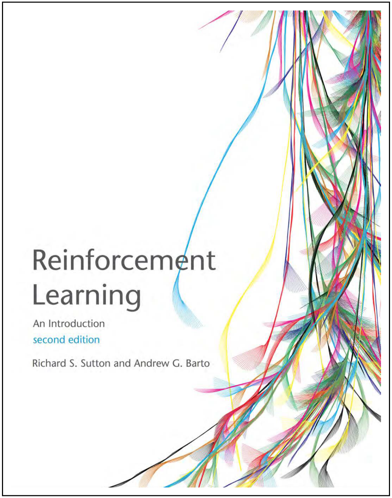
Figure 6.3: Logistic (or *sigmoid*) functions.

Algorithm learns to *react* to an *environment* that provides *rewards*









[11] (sutton-barto-2018)

N

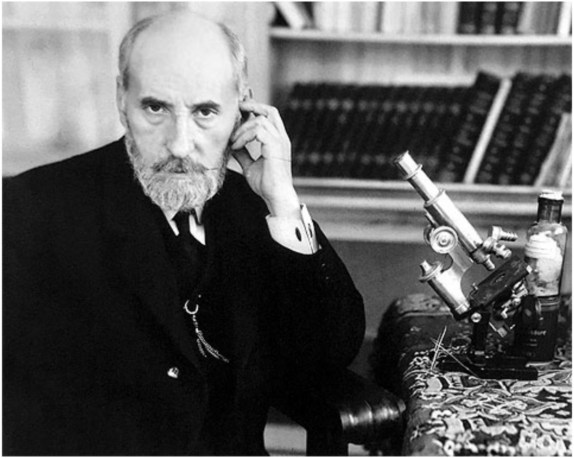
Neural networks

Cajal, father of neuroscience

Artificial neurons

Neural networks: the layered model

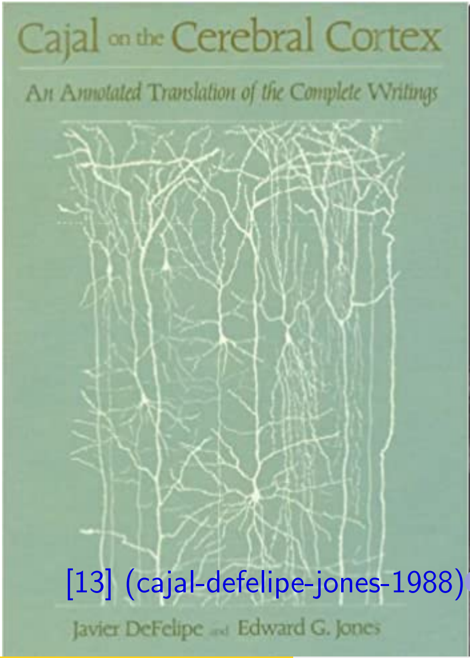
Computational resources and techniques

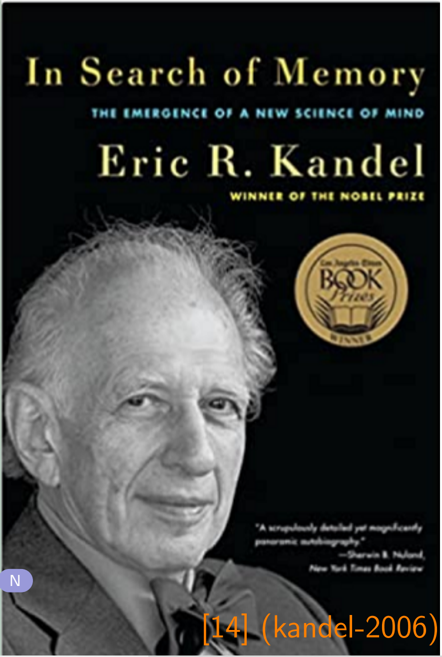
Santiago Ramón y Cajal (1852-1934). Nobel Prize of Physiology or Medicine (1906, shared with Camilo Golgi) for his discoveries about the structure of the nervous system and the role of the neuron. The father of modern neuroscience. The (core of) the *Neural Doctrine* was discovered in 1888 at his *home lab* in the Notariado street 7 (Barcelona), 200 m away from the "Reial Acadèmia de Medicina de Barcelona".

[12] (cajal-pasifik-pasifik-99-00-02)

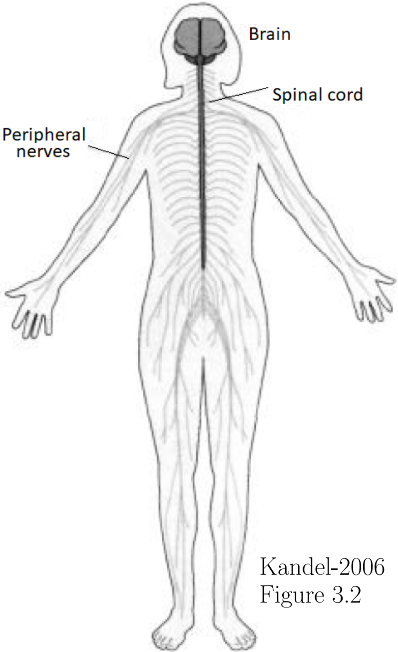
"His studies on the *architectural organization of the brain*, and his *prophetical predictions* of its functions became the basis of neuroanatomy, neurophysiology, neuropathology, and what he named as *'rational psychology'*. This monumental work justifies his well deserved title of founder of modem Neuroscience."



[13] (cajal-defelipe-jones-1988)



[14] (kandel-2006)



Kandel-2006
Figure 3.2

The central and peripheral nervous systems

• The central nervous system, which consists of the **brain** and the **spinal cord**, is bilaterally symmetrical.

• The spinal cord receives sensory information from the skin through bundles of long axons that innervate the skin. These bundles are called peripheral nerves.

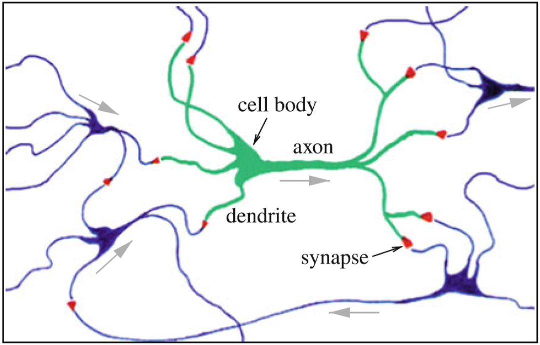
• The spinal cord also sends motor commands to muscles through the axons of the motor neurons.

• These sensory receptors and motor axons are part of the peripheral nervous system.

- Cajal formulated the *neuron doctrine*, the basis for all modern thinking about the nervous system. He is arguably the most important brain scientist who ever lived (Kandel 2006)
- Cajal is considered the founder of modern neuroscience. Historians have ranked him alongside Darwin and Pasteur as one of the greatest biologists of the 19th century and among Copernicus, Galileo and Newton as one of the greatest scientists of all time. His masterpiece, *Texture of the Nervous System of Man and the Vertebrates*, is a foundational text for neuroscience, comparable to *On the Origin of Species* for evolutionary biology (SciAM, April 2022)
- To know the brain is equivalent to ascertaining the material course of *thought* and *will* (Cajal)
- The highest ideal for a biologist is to clarify the *enigma of the self* (Cajal)
- The mysterious butterflies of the soul, [the neurons] whose beating of wings may one day reveal to us the *secrets of the mind* (Cajal)

- Every man can be, if he wants to, a sculptor of his own brain (Cajal)
- Nothing inspires me more reverence and awe than an old man who is willing to change his mind (Cajal)
- The car of Spanish culture lacks the wheel of science (Cajal)

  [15] (sporns-2011), [16] (sporns-2018)



In AL, a useful model of a *neuron* is depicted in Fig. 8.1:

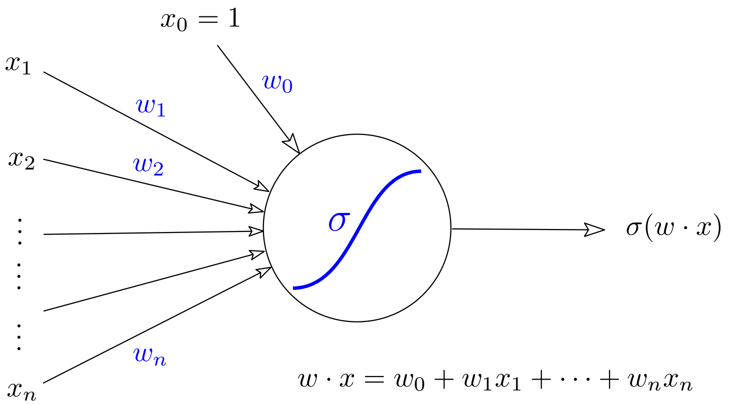


Figure 8.1: Model of a neuron. The neuron's output depends on the inputs x, the *weights* w, and on σ (*activation function*), and this functionality is represented by the decorated circle.

In mathematical terms, a neuron is a function

$$x \mapsto f_w(x) = \sigma(x \cdot w), \tag{1}$$

where $w \in \mathbf{R}^n$ (*weights* or *parameters*) and σ is a *sigmoid* function (*activation function*), like for instance the *logistic function* $\sigma(t) = (1 + e^{-t})^{-1}$, in which case the neuron computes a *logistic regression*. Other widespread choices are $\mathrm{ReLU}(t) = \max(t, 0)$ (Rectified Linear Unit) and $\tanh(t)$.

Augmenting x with $x_0 = 1$ and providing an extra weight w_0 (called the *bias*), the neuron computes $\sigma(w_0 + w_1x_1 + \cdots + x_nw_n)$.

A *neural network* (NN) can be construed as a *composition of neurons* according to a directed graph of connections called the *architecture* of the net.

The standard architecture of a NN is a directed graph structured in *layers* L_j, as illustrated in Figure 9.1, and its *functional signature* can be condensed as a chain:

$$\mathcal{N}: \quad \text{Input} \to L_0 \xrightarrow{f_1} L_1 \xrightarrow{f_2} \cdots \to L_{d-1} \xrightarrow{f_d} L_d \to \text{Output} \tag{2}$$

The integer d is the *depth* of the net. Conventionally, the net is *deep* if $d > 2$, and *shallow* otherwise. The layers $L_1, \ldots, L_{d-1}$ are considered to be *hidden*, while the input and output layers (L_0 and L_d), are *visible*.

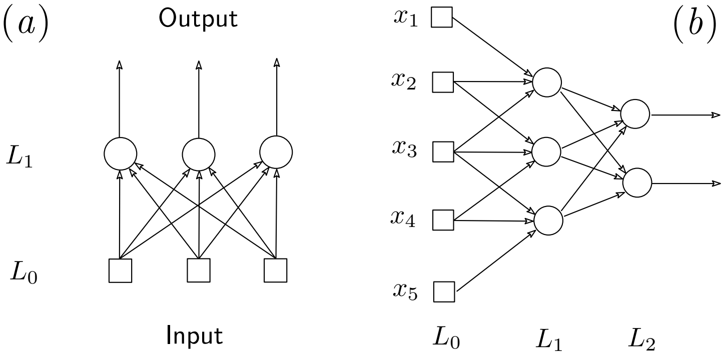


Figure 9.1: (*a*) Neural network with no hidden neurons and *fully connected*. (*b*) Network with a hidden layer L_1 of three neurons fully connected to the two output neurons of L_2. The input layer, L_0, is only partially connected to L_1.

Currently, there is a wealth of software (frameworks) for deep learning (see `Comparison_of_deep-learning_software` in Wikipedia).

For example, *Tensorflow* (see [17]) provides

> ... an interface for expressing machine learning algorithms, and an implementation for executing such algorithms.

Most of them offer a Python interface and increasingly also a *Julia* interface, as for instance *Tensorflow*. An interesting case is Flux (2017), which is pure Julia (framework and interface).

Critical appraisals. [18] (marcus-2018), [19] (marcus-2020)

Interactions AIs–Mathematics

Maths applied to ML:

–Generalization bounds

–Recent work of A. Figalli at al.

–The double descent phenomenon

DL applied to Mathematics

A *generalization bound* for the hypothesis $h \in \mathcal{H}$ returned by a *learning algorithm* has the form

$$L(h) \leqslant_\delta L_{\mathcal{D}}(h) + \mathsf{F}(\mathsf{C}, m, \delta),$$

where $L(h)$ is the *expected loss* (or *error*) of h; $L_{\mathcal{D}}(h)$ is the empirical error of h on the data set $\mathcal{D}$; $\mathsf{C} = \mathsf{C}(\mathcal{H})$ is some measure of the *capacity* (or *complexity*) of $\mathcal{H}$; δ is a positive (small) real number; $\mathsf{F}(\mathsf{C}, m, \delta)$ is a mathematical expression involving the parameters C, $m = |\mathcal{D}|$, and δ; and where $\leqslant_\delta$ means that the inequality is true with probability at least $1 - \delta$ relative to the data samples $\mathcal{D}$.

Example [20, Cor. 4.6] Let $\mathcal{H}$ be a finite set of binary hypothesis. Then

$$L(h) \leqslant_\delta L_{\mathcal{D}}(h) + \sqrt{\frac{\ln|\mathcal{H}| + \ln\frac{2}{\delta}}{2m}}.$$

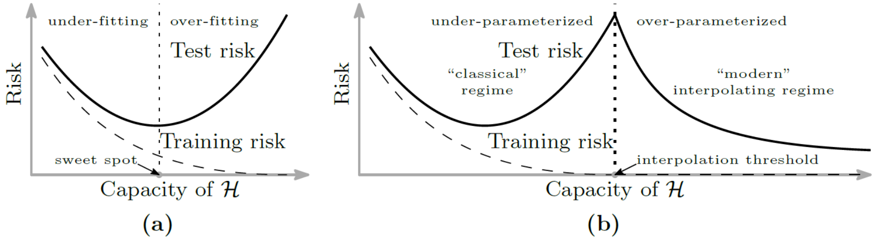
In this case $\mathsf{C} = |\mathcal{H}|$.

N

[21] (abbas-sutter-figalli-woerner-2021), *Effective dimension of machine learning models* (17 pages): "we propose the local effective dimension as a capacity measure which seems to *correlate well with generalization error* on standard data sets. Importantly, we prove that *the local effective dimension bounds the generalization error* and discuss the aptness of this capacity measure for machine learning models". Table 1:

	VC-dimension	Rademacher complexity	Margin-based	Norm-based	Sharpness-based	Local ED
1. Generalization bound	✓	✓	✓	✓	✓	✓
2. Correlation generalization	✗	✗	✗	✗	✗	✓
3. Scale invariant	✗	✓	✗	✗	✗	✓
4. Data dependent	✗	✓	✓	✓	✓	✓
5. Training dependent	✗	✗	✓	✓	✓	✓
6. Finite data	✗	✓	✓	✓	✓	✓
7. Efficient evaluation	✗	✗	✓	✓	✓	✓

The paper [22] (abbas-sutter-zoufal-lucchi-figalli-woerner-2021) establishes that "well-designed *quantum neural networks offer an advantage over classical neural networks* through a higher effective dimension and faster training ability, which we verify on real quantum hardware".



[23] (bach-2022), §11.2, and references therein.

Other applications:

Graph theory: [24] (calvo-tyukin-makarov-2020)

Topology: [25] (gardner-et7-2022).

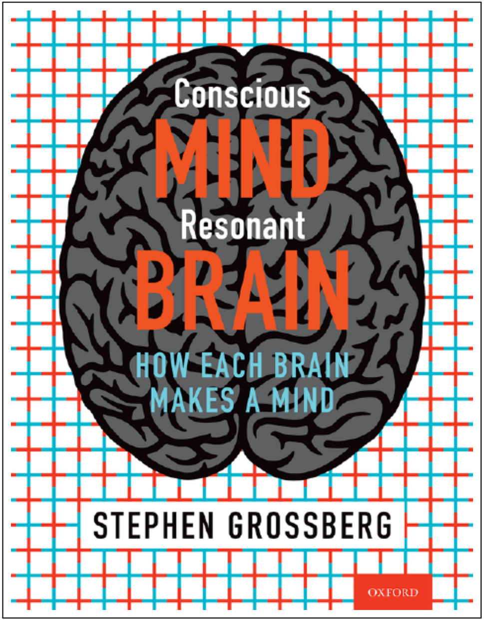
- [26] (he-2021) (general). Publicity. Y-H He.
- [27] (he-2021-CY).
- [28] (lamb-garcez-gori-prates-avelar-vardi-2020) (graph NNs).
- [29] (wagner-2021) (see NL01, Reviews, page 20). A nice application of reinforcement learning.

N

Brain science

Philosophiæ Neuronalis Principia Mathematica?

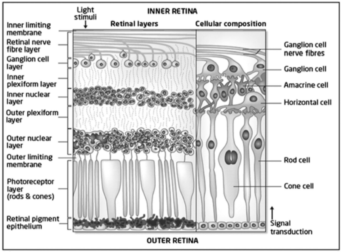
Further Readings



[30] (grossberg-2021) (OUP flyer)

Knuth sampling method

- Newton's Principia: new intuitions and new mathematics.
- Einstein's dark hole: *cognitive impenetrability*.
- Feynman and human vision:



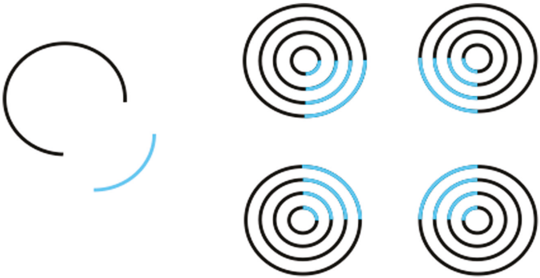
http://brain.oxfordjournals.org/content/early/2011/01/20/brain.awq346

- von Neumann: *The Computer and the Brain*.
- Shun-Ichi Amari: *Back propagation algorithm*.

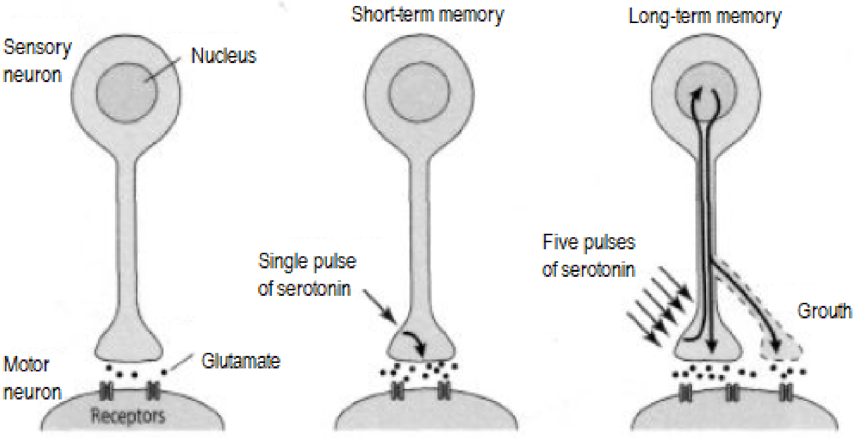
N

- Neocortical circuits: Towards a synthesis of biological and artificial intelligence.
- Bayesian networks and DL: *Not sufficient*.
- Frances Crick: *The Astonishing Hypothesis*.
- Philosophers, D. Dennett in particular: *Consciousness Explained* (1991).

FIGURE 3.10. The visual illusion of neon color spreading. Neither the square nor the blue color that are perceived within it are in the image that defines a neon color display. The display consists only of black and blue arcs.



- Cajal: *Gospel*.



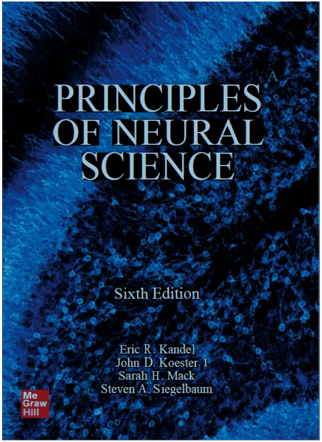
18-4 Changes underlying short- and long-term memory in a single sensory and motor neuron.

Functional change: Synapse strenghthened via enhanced release of glutamate. The nucleus is not involved

Anatomical change: Synthesis of proteins in the nucleus and growth of new synaptic connections, as well as enhanced release of glutamate.

[14] (kandel-2006)

For a thorough treatise on Neural Science, see

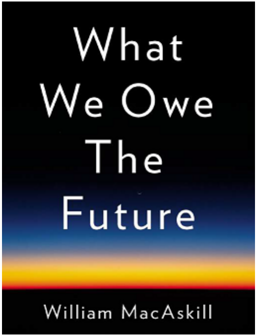


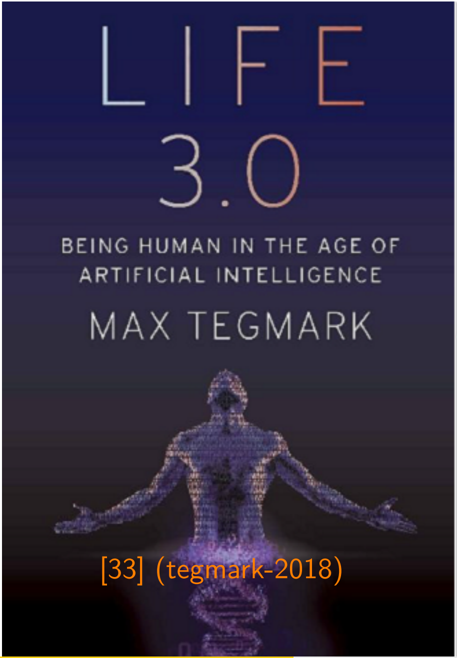
[31] (kandel-koester-mack-siegelbaum-2021)

Some Nobel Prizes in Physiology or Medicine: N

Futurims

[32] (macaskill-2022)




LIFE
3.0
BEING HUMAN IN THE AGE OF
ARTIFICIAL INTELLIGENCE
MAX TEGMARK
[33] (tegmark-2018)


BESTSELLING AUTHOR OF AI SUPERPOWERS
KAI-FU LEE
AI
/// TEN VISIONS FOR OUR FUTURE ///
2041
[34] (lee-qiufan-2021)
CHEN QIUFAN
AUTHOR OF WASTE TIDE

We can only see a short distance ahead,
but we can see plenty there that needs to be done
(Alan Turing)

For example, producing a comprehensive

Philosophiæ Neuronalis
Principia Mathematica!

Thank you!

Notes

This image was the entry icon of the AI section included in the exhibit Imaginary-2021 held at the FME in the Fall of 2021.

It symbolizes the dialectics between two approaches to model intelligent behaviors. One, represented by the artificial hand, relies on *computational sciences* in a broad sense. The other, represented by the human hand, relies on *brain sciences*, also in a broad sense.

In my talk I will consider both approaches, not as separate domains, but rather as *intertwined disciplines*. This allows a sort of *in crescendo* unfolding that will culminate with a glimpse over possible syntheses of the two. All along, I will try to emphasize the two-way bridges with mathematics. I will also provide a good number of reading and study references.

P

Brandon Walker, 2020. Games that AI have played have thus captured the interest of news, here are the most important victories it has won.
1996–1997 Kasparov v. IBM Deep Blue.
2011 IBM Watson Wins Jeopardy!
2013 DeepMind Beats Atari.
2016 AlphaGo v. Lee Sedol.
2017 AlphaZero Masters Chess, Go, and Shogi.
2019 AlphaStar to play Starcraft.

In February 1996 Kasparov won Deep Blue 4-2 (lost match 1, won matches 2,5,6, tie 3,4). In May 1997 (3-11), IBM's Deeper Blue wins Kasparov $3\frac{1}{2} - 2\frac{1}{2}$ (won 1, lost 2, 6, draw 3, 4, 5. Many materials for reflection: the 2003 movie Game Over, and Kasparov's 2017 book (with Greengard) *Deep Thinking: Where Machine Intelligence Ends and Human Creativity Begins*, [35].

AlphaGo (2016). After defeating the European Go champion Fan Hui, AlphaGo was tasked with playing Lee Sedol, one of the highest ranking players ever. Beating Lee prompted him to retire, saying "Even if I become the number one, there is an entity that cannot be defeated".

AlphaZero (2017) was a generalized version of AlphaGo, built with the intention of winning Chess, Go and Shogi (a Japanese version of chess). Not only did AlphaZero beat AlphaGo, it was *done by only playing simulated games against itself, having no examples of expert's games to look at*. *At the start of these simulations it knew absolutely nothing*. *It mastered Chess after 9 hours of training, Shogi after 2, and Go after 34*.

AlphaStar. AlphaZero was then transitioned into AlphaStar with the intention of beating the real-time strategy game Starcraft. In 2019 AlphaStar achieved a ranking in the top 0.2 percent of human players. This was the first time that an AI had ever topped an e-sport.

"Starcraft," released in 1998 by Blizzard Entertainment, is a real-time strategy game where players build a military base, mine resources, and attack other bases.

Other games introduce demands of handling incomplete information (e.g. StarCraft), understanding narrative (e.g. Skyrim), or very long-term planning (e.g. Civilization).

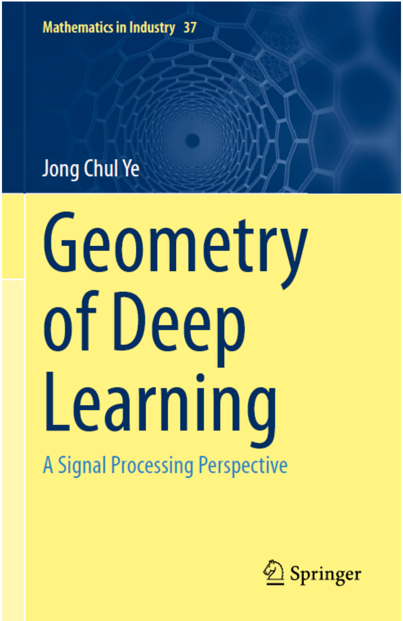
More info: [36] (general), [37] and [38] (backgammon), [39] (Go). See also [35] (an account by Kasparov about his 'chess defeat' by IBM Deep Blue). P

General

[40] (haykin-2009)
[20] (shalevshwartz-bendavid-2014)
[41] (goodfellow-bengio-courville-2016)
[42] (mohri-rostamizadeh-talwalkar-2018)

[43] (rebala-ravi-churiwala-2019)
[44] (alpaydin-2020)
[23] (bach-2022)
[45] (roberts-yaida-hanin-2022)

P



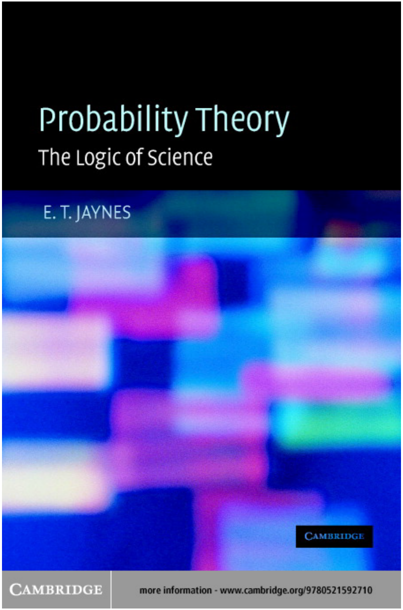
[46] (ye-2022)

Bayesian approaches

[47] (pearl-1988)
[48] (knill-richards-1996)
[49] (pearl-2009)
[50] (pearl-glymour-jewell-2016)
[51] (pearl-mackenzie-2018)

[52] (bielza-larranaga-2020)
[53] (theodoridis-2020)
[54] (efron-hastie-2021)
[55] (russell-norvig-2022)
[56] (murphy-2022)

P



[57] (jaynes-2003)

Applications

[58] (guttag-2016)
[59] (nandy-biswas-2018)
[60] (said-torra-2019)
[61] (lavor-X-zaplana-2018)

[62] (balas-roy-sharma-samui-2019)
[63] (chollet-2020)
[64] (colome-torras-2020)
[65] (X-2021)

29 case studies. Example: Supermarket giant Kroger and 84.51° : *138 different machine learning models in production*; over ten petabytes of customer data analyzed; AutoML (makes it possible for those without traditional data science training to create ML models); forecasts for each item in each of more than 2,500 stores for each of the subsequent fourteen days. **Insights**. Example: The platforms that make AI work (209-215). P

[66] (davenport-miller-2022)

Karl Friston (born 12 July 1959): According to his profile in GoogleScholar, he has published more than 500 articles and has more than 300000 citations (consulted 23 October 2022). Fields of work: Neuroscience, Mathematical and theoretical biology, Variational Bayesian methods.

Some recent papers: [67] (sajid-ball-parr-friston-2021), [68] (dacosta-lanillos-sajid-friston-khan-2022), [69] (barp-et2-friston-et3-2022)

K. Friston G. Pezzulo T. Parr

P

Brahe experienced the solar eclipse of 21 August 1560 [he was 15], and was greatly impressed by the fact that it *had been predicted, although the prediction based on current observational data was a day off*. He realized that *more accurate observations* would be **the key to making more exact predictions**.

Kepler relied on Brahe's data to study the orbit of Mars. The circle was discarded, as it produced errors that were incompatible with Brahe's observational precision. Then he tried eccentric circles, and again the errors were two large. Finally he tried ellipses with a focus at the Sun, and with this he succeeded (Kepler's first law). With this model, he could discover how fast a planet moved (in the form of areal velocity, the second law) and to compare the periods of different planets (third law: a_P^3/T_P^2 has the same value for all planets P, hence $a_P \propto \sqrt[3]{T_P^2}$).

P

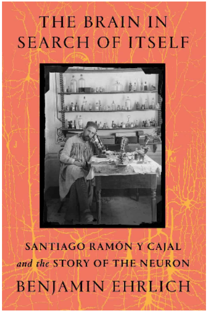
"Conversely, in *active inference* an agent's interaction with the environment is determined by action sequences that *minimize expected free energy* (and not the expected value of a reward signal). Additionally, unlike in reinforcement learning, the reward signal is not differentiated from other types of sensory outcomes. That is any type of outcome may be more or less preferred. This means that *the implicit reward associated with any outcome is a feature of the creature seeing the observation* – not the environment they inhabit. This may be different for different agents, or even for the same agent at different points in time. This highlights that the two frameworks have fundamentally different objectives: *reward-maximization in reinforcement learning* and *free energy minimization in active inference*" [67] (sajid-ball-parr-friston-2021).

Definition of reinforcement learning in [11] (sutton-barto-2018):

"Reinforcement learning is *learning what to do* – how to map situations to actions – *so as to maximize a numerical reward* signal". P

Warmflash-2016: By 1891, Ramón y Cajal decided that the expansion of nerve cells occurred through growth of dendrites, which he referred to as "protoplasmic branches." Within three years, he adapted some ideas from other researchers concerning learning, and *speculated that intelligence was related to the number and efficiency of connections between different types of cells with the part of the brain known as the cerebral cortex*. It was a remarkable idea for the time, and it dovetails with what neuroscientists have learned in recent years.

Full story: [70] (ehrlich-2022)



P

A learning algorithm has good *generalization rate* if the hypothesis $h \in \mathcal{H}$ it supplies is close to the expert who produced the examples, that is, if the expected loss $L(h)$ is small. How to guarantee this condition if the LA only knows the $\mathcal{D}$ examples and somehow the space $\mathcal{H}$? The problem may seem impossible if we consider that $\mathcal{D}$ is always finite, and that $\mathcal{H}$, as a rule, is infinite.

The result of the example tells us that the generalization rate increases if the empirical risk minimization ensures that $L_{\mathcal{D}}(h)$ is small and that m is sufficient large so that the second sum of the expression is also small.

P

he-2021: We review, for a general audience, a variety of recent experiments on extracting structure from machine-learning mathematical data that have been compiled over the years. Focusing on supervised machine-learning on labeled data from different fields ranging from geometry to representation theory, from combinatorics to number theory, we present a comparative study of the accuracies on different problems. The paradigm should be useful for conjecture formulation, finding more efficient methods of computation, as well as probing into certain hierarchy of structures in mathematics. Based on various colloquia, seminars and conference talks in 2020, this is a contribution to the launch of the journal *Data Science in the Mathematical Sciences*.

lamb-et5-2000: if one aims to build richer AI systems, i.e. semantically sound, explainable, and reliable, one has to add a sound reasoning layer to deep learning.

P

■ [71] (hestenes-1987-brain) “In spite of the enormous complexity of the human brain, there are good reasons to believe that only a few basic principles will be needed to understand how it processes sensory input and controls motor output. In fact, the most important principles may be known already! These principles provide the basis for a definite mathematical theory of learning, memory, and behavior.”

Two main points about the book:

• Explores how your mind works, notably how you learn to consciously see, hear, feel, and know things.

• Creates a computational foundation for the next generation of autonomous, adaptive, and intelligent algorithms, devices, and mobile agents in engineering, technology, and AI.

Newton: I hope that this brief summary [split between physical sciences and psychology and neuroscience at the beginning od the XX century] helps to put into historical perspective why it has taken so long to begin to theoretically understand how a brain can give rise to a mind. Such progress required the introduction of a new scientific paradigm that simultaneously discovered new conceptual intuitions and new mathematics with which to understand the nonlinear, nonlocal, and nonstationary laws that link brain to mind. In many scientific revolutions, such as relativity theory and quantum mechanics, once the new physical intuitions were discovered, relevant mathematics was available with which to convert them into rigorous theoretical science. In the mind and brain sciences, we have not been so lucky, since both new intuitions and new mathematics needed to be developed. The Newtonian revolution was also of this kind, since Newton had to both derive the laws of celestial mechanics and to discover calculus with which to mathematically analyze them. My own scientific work since 1957, when I was a Freshman in Dartmouth College, has been devoted to introducing and developing foundational intuitions, mathematics, and the behavioral and neural models built upon them into the mind- brain sciences. I could never have imagined then how much I would be able to discover and understand with the help of many gifted PhD student, postdoctoral fellow, and faculty collaborators.

Einstein: In a letter to Queen Elizabeth of Belgium, 1933, Einstein wrote: “Most of us prefer to look outside rather than inside ourselves; for in the latter case we see but a dark hole, which means: nothing at all.” SG: The neural computations that regulate our daily experiences are inaccessible to us due to the property of *cognitive impenetrability*. Because of this impenetrability, we can behave in a world of interacting percepts, feelings, and ideas without having to worry about, or to be distracted by, our individual nerve cells, electrical potentials, and chemical transmitters. We experience the apparent simplicity of our behavioral mastery rather than the actual complexity of its generative neural machinery.

Feynman: This inversion of the retinal layers may make sense from the viewpoint of how the retina develops, but it seems very peculiar from the viewpoint of how the brain sees. Richard Feynman, one of the most brilliant physicists of the twentieth century, got very interested in how the brain sees until he became aware of these facts. Feynman was a famously honest man who realized that he could not think of a sensible explanation for how the brain sees so well despite the retina's peculiarities of having a blind spot as well as retinal veins and retinal layers in front of the photoreceptors. Every scientist needs heuristics, or intuitive design principles, on which to base a principled scientific theory. These heuristics need to make sense, and in the best cases are even parsimonious and beautiful, for a top theorist to be comfortable with them. Feynman realized that he could not find intuitively plausible principles for how the brain copes with its noisy retinas, so he got out of vision. Of course, a man with Feynman's abundant intellectual and personal resources could not be long deterred by this disappointment. He rapidly moved on to successfully pioneer the fields of quantum computing and nanotechnology.

von Neumann: The McCulloch- Pitts model [of neural networks, 1943] had an influence far beyond the field of neural networks. One of its most enduring influences was upon John Von Neumann, one of the greatest mathematicians of the twentieth century, when he was at the Institute for Advanced Study in Princeton, developing the digital computer that has totally revolutionized our lives. This influence is illustrated by a famous series of Silliman Lectures that Von Neumann gave at Yale University in 1956, which he called *The Computer and the Brain*.

Shun-Ichi Amari: Variants of the back propagation model were introduced independently by several authors, including Shun- Ichi Amari in 1972 in Japan (Amari, 1972), Paul Werbos in 1974 as part of his PhD thesis at Harvard University (Werbos, 1974, 1994), and David Parker in 1982 in California (Parker, 1982, 1985, 1986, 1987). Paul Werbos seems to be the first person to have published the algorithm in its modern form and use it successfully in applications. Back propagation finally became popular in response to an oft- cited article by David Rumelhart, Geoffrey Hinton, and Ronald Williams that was published in

1986 (Rumelhart, Hinton, and Williams, 1986). Their 1986 article has often been incorrectly cited as the source of the algorithm by people who are unaware of its history.

P

Neocortical circuits: *Towards a synthesis of biological and artificial intelligence*. As I noted at the beginning of this chapter, the same canonical design for neocortical circuits can be specialized to accomplish many different perceptual and cognitive tasks. This claim has been partially realized by the 3D LAMINART model for 3D vision and figure- ground perception; the cARTWORD model for conscious speech perception, learning, and recognition; and the LIST PARSE model for cognitive working memory and planning. These models provide a foundation for a future unified theory of biological intelligence. They also illustrate how software and hardware with a similar canonical circuit design can be specialized to carry out different types of intelligence in future engineering and technological applications. Because of their compatible design, these software and hardware circuits can be assembled into the "brain" of increasingly autonomous adaptive agents. If and when this happens, biological and artificial intelligence will be able to seamlessly interact, and complement each others' strengths and weaknesses.

Bayesian networks: While most mind and brain experimentalists ignored theory, and most theorists looked for more hospitable frontiers, there arose the widespread tendency to fill this theoretical vacuum by interpreting brain function in terms of whatever technology happened to be current. The ever-expanding list of technological metaphors to which the brain has been compared includes telegraph circuits, hydraulic systems, information processing channels, digital computers, linear control systems, catastrophes (in the mathematical sense of René Thom (1977), holograms, spin glasses, *Bayesian networks*, and *Deep Learning*. *All of these metaphors have been unable to explain substantial databases that link mind and brain*, since none of them arose from a sustained analysis of mind and brain data.

Francis Crick: Francis Crick, who is most famous for his classical Nobel Prize–winning work on understanding DNA, published a book in 1994, with the arresting title *The Astonishing Hypothesis*, in which he discussed some aspects of consciousness (Crick, 1994). The astonishing hypothesis refers to the fact that *all of our mental phenomena arise from activities of the nerve cells, or neurons, of our brains*. Crick claimed that “this hypothesis is so alien to the ideas of most people alive today that it can truly be called astonishing”

Philosophers, Daniel Dennett in particular: Various philosophers, notably the talented and prolific Daniel Dennett, have argued otherwise. I was gratified when Dennett attended lectures of mine about neon color spreading in the 1980s, since

I believed then, as I do now, that theoretical discoveries about how each brain makes a mind offer a treasure trove of new ideas for philosophical discussion and analysis. I was therefore surprised when he later wrote in his 1991 book *Consciousness Explained* (Dennett, 1991): "The fundamental flaw in the idea of 'filling- in' is that it suggests that the brain is providing something when in fact the brain is ignoring something" (p. 356). "The brain doesn't have to 'fill in' for the blind spot . . . We don't notice these gaps, but they don't have to be filled in because we're designed not to notice them" (p. 355). In other words, Dennett argued that a physical process of filling-in does not occur. Given that Dennett put an example of neon color spreading on the back cover of his book, he clearly viewed this claim as an important part of his proposals about consciousness.

There were problems with Dennett's viewpoint even when he first published it. First and foremost, it explained no data about how filling-in works and what its properties are. Although Dennett heard me give a lecture in which I offered a detailed neural explanation of a lot of data about neon color spreading, in his book, he did not attempt to explain any of these data. *He offered a personal opinion, not a scientific theory*. Any viable scientific proposal about experimental data, however, must explain these data at least as well as competing theories. A reader of Dennett's book might come away thinking that there were no theories available to explain how filling-in happens, and Dennett's proposal was at least a step in that direction. This was not, unfortunately, the case.

Cajal. *Gospel*: Cajal's Neuron Doctrine is now taken for granted as gospel. However, on the day that Cajal and Golgi shared the Nobel Prize, Golgi used his own Nobel lecture to deny that individual neurons exist, even as Cajal asserted that they do. Controversy existed then in psychology and neuroscience, and it continues to exist today. How could it be otherwise, given the complexity of the subject matter, and its importance to human identity and society?

P

1906: Camillo Golgi and Santiago Ramón y Cajal: *structure of the nervous system*.

1932: Sir Charles Scott Sherrington and Edgar Douglas Adrian: *functions of neurons*.

1936: Henry Hallett Dale and Otto Loewi: *chemical transmission of nerve impulses* (neurotransmitters)

1944: Joseph Erlanger and Herbert Spencer Gasser: functions of single nerve fibers (*action potentials*)

1949: Walter Rudolf Hess (interbrain) and António Caetano Egas Moniz (lobotomy)

1962: Francis Harry Compton Crick, James Dewey Watson, and Maurice Hugh Frederick Wilkins; *structure nucleic acids*.

1963: Sir John Carew Eccles (*synapse*), Sir Alan Lloyd Hodgkin, and Sir Andrew Fielding Huxley: *ionic mechanisms involved in excitation and inhibition in the peripheral and central portions of the nerve cell membrane*.

1970: Julius Axelrod, Ulf von Euler, and Sir Bernard Katz: *transmittors in the nerve terminals and the mechanism for their storage, release and inactivation*.

1981: Roger W. Sperry (*functional specialization of cerebral hemispheres*), and David H. Hubel – Torsten N. Wiesel (*information processing in the visual system*).

2000: Arvid Carlsson (*dopamine*), Paul Greengard (*molecular and cellular functions of neurons*), and Eric R. Kandel (*physiological basis of memory storage in neurons*): *signal transduction in the nervous system*.

2014: John O'Keefe, May-Britt Moser, and Edvard I. Moser: *discoveries of cells that constitute a positioning system in the brain*.

P

References I

[1] S. B. McGrayne, *The theory that would not die*.
Yale University Press, 2011.

[2] N. Silver, *The signal and the noise: why so many predictions fail–but some don't*.
Penguin, 2012.

[3] C. Teuscher, *Alan Turing: Life and legacy of a great thinker*, vol. 54.
Springer, 2004.
Foreword by Douglas Hofstadter.

References II

[4] B. J. Copeland, *The essential Turing*.
Oxford University Press, 2004.
AMS Notices Review,by A. Hodges.

[5] A. Wigderson, *Mathematics and computation*.
Princeton University Press, 2019.
https://www.math.ias.edu/files/mathandcomp.pdf, 372 p.

[6] G. Strang, *Linear algebra and learning from data*.
Wellesley-Cambridge Press, 2019.

References III

[7] G. Nguyen, S. Dlugolinsky, M. Bobák, V. Tran, Á. L. García, I. Heredia, P. Malík, and L. Hluchỳ, "Machine Learning and Deep Learning frameworks and libraries for large-scale data mining: a survey," *Artificial Intelligence Review*, vol. 52, no. 1, pp. 77–124, 2019.

[8] B. Marr, "The Key Definitions Of Artificial Intelligence (AI) That Explain Its Importance," Consulted 2021.

Key definitions of AI.

[9] T. Parr, G. Pezzulo, and K. J. Friston, *Active inference: the free energy principle in mind, brain, and behavior*.

MIT Press, 2022.

[10] R. Smith, K. J. Friston, and C. J. Whyte, "A step-by-step tutorial on active inference and its application to empirical data," *Journal of mathematical psychology*, vol. 107, p. 102632, 2022.

References IV

[11] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction (2nd edition)*.

MIT press, 2018.

[12] S. Ramón y Cajal, P. Pasifik, and T. Pasifik, *Texture of the Nervous System of Man and the Vertebrates (Volumes I, II, III)*.

Springer-Verlag, 1999, 2000, 2002.

An annotated and edited translation of the original Spanish text (1904) with the additions of the French version by Pedro Pasifik and Tauba Pasifik.

[13] S. Ramón y Cajal, J. DeFelipe, and E. G. Jones, *Cajal on the cerebral cortex: An annotated translation of the complete writings*.

Oxford University Press, 1988.

XVII+654p.

References V

[14] E. R. Kandel, *In search of memory: The emergence of a new science of mind*.

W. W. Norton & Company, 2006.

[15] O. Sporns, *Networks of the Brain*.

MIT press, 2011.

[16] O. Sporns, "Graph theory methods: applications in brain networks," *Dialogues in clinical neuroscience*, vol. 20, no. 2, 2018.

[17] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, *et al.*, "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," *arXiv preprint arXiv:1603.04467*, 2016.

References VI

[18] G. Marcus, "Deep learning: A critical appraisal," 2018.

https://arxiv.org/ftp/arxiv/papers/1801/1801.00631.pdf.

[19] G. Marcus, "The next decade in ai: four steps towards robust artificial intelligence," 2020.

https://arxiv.org/vc/arxiv/papers/2002/2002.06177v2.pdf.

[20] S. Shalev-Shwartz and S. Ben-David, *Understanding machine learning: From theory to algorithms*.

Cambridge University Press, 2014.

440 pp.

[21] A. Abbas, D. Sutter, A. Figalli, and S. Woerner, "Effective dimension of machine learning models," 2021.

https://arxiv.org/pdf/2112.04807.

References VII

[22] A. Abbas, D. Sutter, C. Zoufal, A. Lucchi, A. Figalli, and S. Woerner, "The power of quantum neural networks," *Nature Computational Science*, vol. 1, no. 6, pp. 403–409, 2021.

https://arxiv.org/pdf/2011.00027.

[23] F. Bach, "Learning theory from first principles, draft *(13-Oct-22)*," 2022.

https://www.di.ens.fr/~fbach/ltfp_book.pdf. xiv+360 p.

[24] C. Calvo Tapia, I. Tyukin, and V. A. Makarov, "Universal principles justify the existence of concept cells," *Scientific reports*, vol. 10, no. 1, pp. 1–9, 2020.

[25] R. J. Gardner, E. Hermansen, M. Pachitariu, Y. Burak, N. A. Baas, B. A. Dunn, M.-B. Moser, and E. I. Moser, "Toroidal topology of population activity in grid cells," *Nature*, vol. 602, no. 7895, pp. 123–128, 2022.

References VIII

[26] Y.-H. He, "Machine-learning mathematical structures," 2021.
https://arxiv.org/pdf/2101.06317.pdf.

[27] Y.-H. He, *The Calabi–Yau Landscape: From Geometry, to Physics, to Machine Learning*, vol. 2293 of *Lecture Notes in Mathematics*.
Springer, 2021.

[28] L. Lamb, A. Garcez, M. Gori, M. Prates, P. Avelar, and M. Vardi, "Graph neural networks meet neural-symbolic computing: A survey and perspective," 2020.
https://arxiv.org/pdf/2003.00330.pdf.

[29] A. Z. Wagner, "Constructions in combinatorics via neural networks," 2021.
https://arxiv.org/pdf/2104.14516.pdf.

References IX

[30] S. Grossberg, *Conscious mind, resonant brain: How each brain makes a mind*.

Oxford University Press, 2021.

[31] E. R. Kandel, J. D. Koester, S. H. Mack, and S. A. Siegelbaum (editors), *Principles of Neural Science (sixth edition)*.

McGraw-Hill, 2021.

[32] W. MacAskill, *What we owe the future*.

Basic books, 2022.

[33] M. Tegmark, *Life 3.0: Being human in the age of artificial intelligence*.

Vintage, 2018.

References X

[34] K.-F. Lee and C. Qiufan, *AI 2041: Ten visions for our future*.
Penguin Random House, 2021.

[35] G. Kasparov and M. Greengard, *Deep thinking: where machine intelligence ends and human creativity begins*.
Hachette UK, 2017.

[36] J. Togelius, "Better playing through algorithms," 2022.
Togelius blogspot.

[37] G. Tesauro, "Programming backgammon using self-teaching neural nets," *Artificial Intelligence*, vol. 134, no. 1-2, pp. 181–199, 2002.
pdf.

References XI

[38] N. Papahristou, *Decision making in multiplayer environments: application in backgammon variants*.

PhD thesis, University of Macedonia, Department of Applied Informatics, 2015.

[39] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, *et al.*, "Mastering the game of go without human knowledge," *Nature*, vol. 550, pp. 354–359, 2017.

[40] S. Haykin, "Neural networks and learning machines," 2009.

xxx + 906 pp.

[41] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*.

MIT press, 2016.

http://www.deeplearningbook.org, 781 p.

References XII

[42] M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Foundations of machine learning*.

MIT press, 2018.

xvi + 486 pp.

[43] G. Rebala, A. Ravi, and S. Churiwala, *An Introduction to Machine Learning*.

Springer, 2019.

xxii + 263 p.

[44] E. Alpaydin, *Introduction to machine learning (fourth edition)*.

Adaptive computation and machine learning, MIT press, 2020.

xxiv + 682 pp. 1st edition: 2004; 2nd, 2010; 3rd, 2014.

References XIII

[45] D. A. Roberts, S. Yaida, and B. Hanin, *The Principles of Deep Learning Theory: An Effective Theory Approach to Understanding Neural Networks*.

Cambridge University Press, 2022.

Draft: `https://arxiv.org/pdf/2106.10165.pdf`. viii+441 p.

[46] J. C. Ye, *Geometry of Deep Learning: A Signal Processing Perspective*.

No. 37 in Mathematics in Industry, Springer, 2022.

[47] J. Pearl, *Probabilistic reasoning in intelligent systems: networks of plausible inference*.

Morgan kaufmann, 1988.

[48] D. C. Knill and W. Richards (editors), *Perception as Bayesian inference*.

Cambridge University Press, 1996.

Paper #1 is *Pattern theory: a unifying perspective*, by D. Mumford.

References XIV

[49] J. Pearl *et al.*, *Causality: Models, reasoning and inference (2nd edition)*, vol. 19.

Cambridge University Press, 2009.

The first edition appeared in 2000.

[50] M. Glymour, J. Pearl, and N. P. Jewell, *Causal inference in statistics: A primer*.

John Wiley & Sons, 2016.

[51] J. Pearl and D. Mackenzie, *The book of why: The new science of cause and effect*.

Basic Books, 2018.

References XV

[52] C. Bielza and P. Larrañaga, *Data-Driven Computational Neuroscience: Machine Learning and Statistical Models*.

Cambridge University Press, 2020.

xiv+747 p.

[53] S. Theodoridis, *Machine learning: a Bayesian and optimization perspective (2nd edition)*.

Academic Press, 2020.

xxvii+1131 p.

[54] B. Efron and T. Hastie, *Computer Age Statistical Inference, Student Edition: Algorithms, Evidence, and Data Science*, vol. 6.

Cambridge University Press, 2021.

References XVI

[55] S. Russell and P. Norvig, *Artificial intelligence: A modern approach (Fourth edition)*.

Pearson, 2022.

1166 p. Contributing writers: Ming-Wei Chang Jacob Devlin Anca Dragan David Forsyth Ian Goodfellow Jitendra M. Malik Vikash Mansinghka Judea Pearl Michael Wooldridge.

[56] K. P. Murphy, *Probabilistic Machine Learning: An introduction*.

MIT Press, 2022.

[57] E. T. Jaynes, *Probability theory: The logic of science*.

Cambridge University Press, 2003.

References XVII

[58] J. Guttag, *Introduction to computation and programming using Python: With application to understanding data*.

MIT Press, 2016.

xvii + 447 p.

[59] A. Nandy and M. Biswas, *Reinforcement Learning: With Open AI, TensorFlow and Keras Using Python*.

Apress, 2018.

xiii + 167 p.

[60] A. Said and V. Torra, *Data Science in Practice*.

Springer, 2019.

viii + 195 p.

References XVIII

[61] C. Lavor, S. Xambó-Descamps, and I. Zaplana, *A Geometric Algebra Invitation to Space-Time Physics, Robotics and Molecular Geometry*.
SBMA/Springerbrief, Springer, 2018.

[62] V. E. Balas, S. S. Roy, D. Sharma, and P. Samui, *Handbook of deep learning applications*, vol. 136.
Springer, 2019.
vi + 383 p.

[63] F. Chollet, *Deep learning with Python (2nd edition, V4)*.
Manning Early Access Program, 2020.
210 p.

References XIX

[64] A. Colomé and C. Torras, *Reinforcement Learning of Bimanual Robot Skills*, vol. 134 of *Springer Tracts in Advanced Robotics*.

Springer, 2020.

xx + 182 p.

[65] S. Xambó-Descamps (ed), *Systems, Patterns and Data Engineering with Geometric Calculi*.

ICIAM2019 SEMA SIMAI Springer Series, Springer, 2021.

[66] T. H. Davenport and S. M. Miller, *Working with AI: Real Stories of Human-Machine Collaboration*.

MIT Press, 2022.

References XX

[67] N. Sajid, P. J. Ball, T. Parr, and K. J. Friston, "Active inference: demystified and compared," *Neural computation*, vol. 33, no. 3, pp. 674–712, 2021.

arXiv pdf.

[68] L. Da Costa, P. Lanillos, N. Sajid, K. Friston, and S. Khan, "How active inference could help revolutionise robotics," *Entropy*, vol. 24, no. 3, p. 361, 2022.

[69] A. Barp, L. Da Costa, G. França, K. Friston, M. Girolami, M. I. Jordan, and G. A. Pavliotis, "Geometric methods for sampling, optimisation, inference and adaptive agents," 2022.

`https://arxiv.org/pdf/2203.10592`.

References XXI

[70] B. Ehrlich, *The Brain in Search of Itself: Santiago Ramón y Cajal and the story of the neuron*.

Farrar, Straus and Giroux, 2022.

[71] D. Hestenes, "How the brain works: the next great scientific revolution," in *Maximum-entropy and Bayesian Spectral Analysis and Estimation Problems*, pp. 173–205, Reidel, 1987.