

Igor Fischer, Katharina A. Faust

igor.fischer@med.uni-duesseldorf.de

KatharinaAngela.Faust@med.uni-duesseldorf.de

Department of Neurosurgery, Medical Faculty and

University Hospital Düsseldorf, Heinrich-Heine-

University Düsseldorf, Germany

00-00

HISTORY AND APPLICATIONS OF LOGISTIC FUNCTION: FROM POPULATION GROWTH TO ARTIFICIAL INTELLIGENCE

Abstract: Logistic function has been known under different names and used in very diverse fields, from quantum physics and chemistry to ecology, statistics and artificial intelligence. Physicists know it as Fermi-Dirac distribution, which describes statistical distribution of quantum particles with half-integer spin over energy levels. Chemists, pharmacologists and toxicologists meet it under the name “Hill equation”, which describes the effects of a reaction in dependence of the substance concentration. Logistic function models the population growth in a limited environment, and in statistics and machine learning it is used to model the probability that the observation belongs to one of the two possible classes. As a “soft” variant of the Heaviside step function it is commonly applied in the activation function in the output layer of neural networks and many practitioners consider it the canonical sigmoid function. The wide applicability of the logistic function is an indirect consequence of the central limit theorem.

Keywords: logistic function, Hill equation, Fermi-Dirac distribution

Introduction

Among the myriad mathematical function which can be constructed, several have historically crystalized as “elementary”. Among them doubtlessly belong logarithms, exponential function, trigonometric functions, and their inverses. These functions are being taught already in school, every scientific calculator can calculate them directly, and in computer programming languages they are, as a rule, immediately accessible or through standard, well known, documented and maintained libraries. The title “elementary functions” should not be taken too literally; it is more a matter of tradition than of their mathematical properties. For example, trigonometric functions follow from the exponential one – in that sense, the exponential function is more “elementary” than trigonometric ones.

Specialized areas of mathematics, science, and engineering often use additional functions which can be considered “elementary” within the area, such as factorial for combinatorics or the Lorentz factor in relativistic physics, in the sense that they appear often in equations or serve as building blocks for further functions. Logistic function is one such function which often, but under different names, appear in many very diverse areas of science and engineering.

In its most simple variant, the *standard* logistic function is defined as:

$$f(x) = \frac{1}{1 + e^{-x}}$$

It is easy to show that the logistic function is simply scaled and shifted hyperbolic tangent:

$$\frac{1}{1 + e^{-x}} = \frac{1}{2} + \frac{1}{2} \tanh \frac{x}{2}$$

For practical purposes, logistic function is usually parametrized in some way.

1. Population growth in limited environment

First historical records about logistic function date back to the period 1838-1847, by the Belgian mathematician Pierre-François Verhulst [1-3]. The phenomenon of exponential growth was already well known at that time, and at first glance it seemed that if each generation in a country had more descendants than the parent generation, the population should grow exponentially. At the same time, it was obvious that for its survival, and especially for growth, the population needed resources, primarily food, which were available only in limited quantities, which in turn limited growth.

Inspired by earlier considerations by Thomas Malthus [4] and Adolphe Quetelet, Verhulst postulated that the growth rate depends on the population P itself, on the maximum possible population K that the environment can support, and on the reproduction factor r , so that it can be modeled by the differential equation:

$$\frac{dP}{dt} = rP \left(1 - \frac{P}{K}\right)$$

In other words, the rate of population growth – the number of additional inhabitants

per unit of time – is proportional to the population size itself, the reproduction factor (a multiplicative term that tells how many times the population would be larger at time $t + 1$ time unit than at time t if growth were not limited), and the relative difference between the population and its limiting size. Verhulst called this differential equation “logistic”, for reasons that cannot be reconstructed any more. It is only certain that this name has nothing to do with logistics in a military or economic context.

When $P \ll K$ or, equivalently, when environmental resources are unlimited ($K \rightarrow \infty$), growth is approximately exponential:

$$\frac{dP}{dt} \approx rP$$

and:

$$P \approx P(0)e^{rt}$$

In the opposite case, when $P \rightarrow K$,

$$\frac{dP}{dt} \approx rK \left(1 - \frac{P}{K}\right)$$

so that:

$$P \approx K \left(1 - \left(1 - \frac{P(0)}{K}\right)e^{-rt}\right)$$

In the general case, the solution to the Verhulst equation is:

$$P(t) = \frac{P(0)e^{rt}}{1 + \frac{P(0)}{K}(e^{rt} - 1)}$$

or, rearranged:

$$P(t) = \frac{K}{1 + e^{\ln\left(\frac{K}{P(0)} - 1\right) - rt}}$$

This is simply the standard logistic function from the introduction, scaled by factor K , where $x = rt - \ln(K/P(0) - 1)$. Its shape is shown in Fig. 1. As we can see – and it can be recognized from the approximative equations for limit cases – the function asymptotically approaches 0 and K as time approaches $-\infty$ resp. $+\infty$, by the same velocity in both cases. Moreover, the function is symmetrical around $P = K/2$, that is, around $t = \frac{1}{r} \ln\left(\frac{K}{P(0)} - 1\right)$.

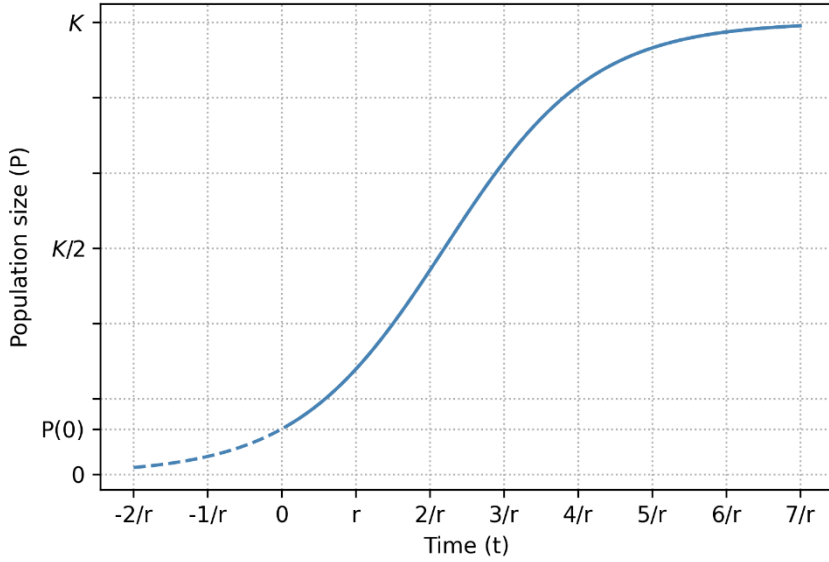


Fig. 1.: Verhulst's logistic function in parametrized coordinate system.

It goes without saying that this model applies not only to countries and human populations, but also to animals in their habitat, microbes in the organism they have invaded, cells grown in a laboratory, in a Petri dish, etc.

The logistic function, graphically, has the shape of an elongated letter “S”, and therefore belongs to the family of so-called “sigmoid functions”, along with the hyperbolic tangent, the cumulative normal distribution, and countless other functions.

2. Chemical reactivity in the presence of inhibitory and excitatory factors

In the early 20th century, the British physiologist and pharmacist Archibald Vivian Hill studied the saturation of hemoglobin dissolved in various salts as a function of oxygen tension [6]. Hill had the measured data available and was trying to find a mathematical function that described them well. Of the three very similar functions he listed in his article, the second one, the “B” function, proved to be the best compromise between precision and complexity. Hill expressed this function in the form:

$$y = 100 \frac{Kx^n}{1 + Kx^n}$$

and this is, basically, the equation known today under his name. Here, y is the percentage of hemoglobin saturation with oxygen, x is the oxygen tension, and K

and n are parameters that depend on the substances in which the reaction is described. For example, Hill found that for hemoglobin in Na_2HPO_4 the model best matches measurements for $n = 1.67$ and $K = 0.0453$, while for NaHCO_3 the best parameters are $n = 1.835$ and $K = 0.0155$.

The purpose of the constant 100 is only to express the saturation as percentage (0% – 100%) and can be ignored for practical purposes. Moreover, if we express x^n as $e^{n \ln x}$, the whole equation becomes:

$$y = \frac{1}{1 + e^{-n \ln x - \ln K}}$$

which is basically also logistic function, but with different parameter names and the argument on the logistic scale.

Although the Hill equation is essentially a heuristic, without any theoretical justification, it soon found application in the broader fields of chemodynamics, toxicology, and pharmacology. Expanding the equation by the multiplicative factor y_{max} , we obtain the three-parameter Hill equation, whose co-domain is not limited to (0, 1), but can describe phenomena on the scale $0 - y_{max}$:

$$y = y_{max} \frac{1}{1 + e^{-n \ln x - \ln K}}$$

The Michaelis–Menten equation [7], which describes the reaction rate v as a function of the substrate concentration S , is a special case of the Hill equation, where $n = 1$ and $y_{max} = V_{max}$:

$$v = V_{max} \frac{S}{K_m + S}$$

K_m is here simply the substrate concentration at which the reaction rate reaches half of its maximum rate.

In pharmacology, the three-parameter Hill equation is used to describe the effect of a drug as a function of its concentration:

$$E = E_{max} \frac{C^\alpha}{EC_{50}^\alpha + C^\alpha}$$

Here, C is the drug concentration, EC_{50} is the concentration at which the drug develops 50% of its maximum possible effect, E_{max} the maximum effect that the drug can develop, and α a parameter that depends on the type and effectiveness of the drug.

In some cases – especially in toxicology, where the survival of subjects exposed to a toxin is modeled – it is desirable to introduce a fourth parameter E_0 that determines the lower limit of action:

$$E = E_0 + (E_{max} - E_0) \frac{C^\alpha}{EC_{50}^\alpha + C^\alpha}$$

The shape of this function is shown in Figure 2. It is important to note that here, unlike the three-parameter equation, EC_{50} is not the concentration at which the drug develops 50% of its maximum effect, but rather at which the effect is $E = E_0 + (E_{max} - E_0)/2$, i.e. halfway between the minimum and maximum effects.

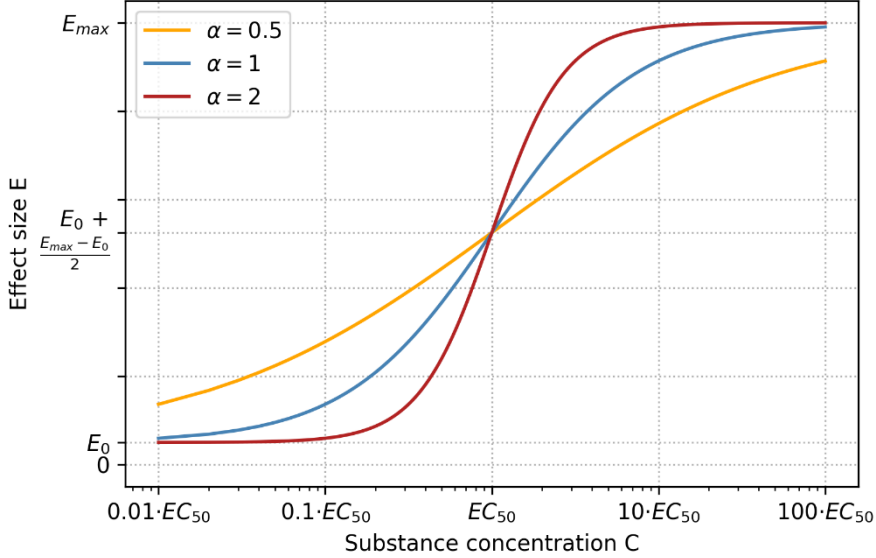


Fig. 2.: Four-parameter Hill equation, for different values of parameter α .

3. Distribution of quantum particles over energy levels

The development of quantum mechanics in the first decades of the 20th century forced physicists to reexamine many of the established concepts and models. Among other things, it was shown that the Maxwell-Boltzmann model of the distribution of particles – originally gas atoms and molecules – over energy levels was inconsistent with the fact that energy in a quantum system can only have discrete values. To explain the distribution of elementary particles with half-integer spin, such as electrons or protons, Enrico Fermi and Paul Dirac independently developed a new model in 1926, the distribution that now bears their name.

Although quantum mechanics has the reputation of being quite complex, for deriving the Fermi-Dirac distribution, it suffices to know only its basics and some elementary probability and combinatorics. In his publication [8], Fermi skipped the derivation and stated the result directly. Below we list the main steps of the derivation, as presented in Dirac's paper [9].

Taking into account the Pauli exclusion principle, the number of different ways in which a fixed number N_s of identical particles, e.g. electrons, can be distributed into an equally fixed number A_s of quantum states (wave functions) with identical energy E_s , where $N_s \leq A_s$, corresponds to the number of combinations N_s out of A_s :

$$W_s = \binom{A_s}{N_s} = \frac{A_s!}{N_s! (A_s - N_s)!}$$

The total number of combinations, over all energy levels, is:

$$W = \prod_s W_s = \prod_s \frac{A_s!}{N_s! (A_s - N_s)!}$$

Assuming that all energy levels are equally probable and with the conditions that the total number of particles is $N = \sum_s N_s$ and that the total energy is given by $E = \sum_s N_s E_s$, and by applying Stirling approximation, we find that the distribution with the maximum likelihood is given by:

$$N_s = \frac{A_s}{1 + e^{\alpha + \beta E_s}}$$

which is again logistic function with different parameter names. In quantum physics, α and β have the meanings:

$$\alpha = -\frac{\mu}{kT}$$

and

$$\beta = \frac{1}{kT}$$

where μ is the chemical potential of the system, T its absolute temperature, and k Boltzmann's constant. For semiconductors, μ lies between the valent and the conductive band. As the number of particles can be very large and isn't important by itself, it is common to express Fermi-Dirac distribution as the probability of occupying the energy level E_s :

$$p(E_s) = \frac{N_s}{A_s} = \frac{1}{1 + e^{\frac{E_s - \mu}{kT}}}$$

It is interesting to note that the argument E_s in the exponent's nominator here comes with a positive sign, while in Verhulst's and Hill equation t and $\ln x$, respectively, come with a negative sign. Therefore, $p(E_s)$ is a monotonically decreasing function, from 1 to 0 for increasing values of the argument, while

Verhulst's and Hill are mirrored, monotonically increasing.

4. Regression for binary dependent variables

4.1. Motivation: Linear regression

Linear regression is one of the oldest, best known, and most important statistical methods. It models the relationship between one or more independent variables (correlates, predictors, or simply x) and one real dependent variable (y). Independent variables usually also have real values, but this is not necessary. It suffices that they are numerical in nature.

Probably the most common application is when we have a set of measured values or observations, where each observation is described by two or more numbers, for example, the mass m with which a coil spring is loaded and the elongation y of the spring. In this simple example, it is reasonable to assume that the mass is the independent variable, and that the elongation of the spring depends on it. Based on a large number of such measurements, with different loads on the spring, linear regression allows us to determine the principle by which the elongation depends on the mass. According to Hooke's law, the elongation is proportional to the mass, perhaps with the addition of a constant term y_0 which denotes the elongation of the unloaded spring:

$$y = y_0 + km$$

where k is the elastic coefficient of the spring. In the general case, with arbitrarily many independent variables, statisticians usually write a linear model as:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_m$$

or, in vector form,

$$y = \beta_0 + \boldsymbol{\beta}^T \mathbf{x}$$

where β_0 is the constant term (intercept on the y -axis), $x_1 \dots x_m$ independent variables (components of the vector $\mathbf{x}$), and $\beta_1 \dots \beta_m$ the corresponding coefficients, i.e. the components of the vector $\boldsymbol{\beta}$. It is important to note that this is just a model; each individual measurement is subject to error ε_i , and linear regression assumes the errors to be additive, normally distributed around zero and to have a constant variance σ^2 :

$$y_i = \beta_0 + \boldsymbol{\beta}^T \mathbf{x}_i + \varepsilon_i, \quad \varepsilon \sim N(0, \sigma)$$

The assumption of normality for the errors might appear arbitrary, but in practice, it is often a good approximation. Normal distribution is very common in

nature, due to the central limit theorem.

4.2. Logistic regression

In 1957-58, David Roxbee Cox considered the problem of regression when the dependent variable is not real, but binary – that is, when it can only have the values 0 or 1. Such problems often arise when observations can be classified into two discrete classes, e.g. "healthy" versus "sick", or "wheat" versus "barley", which we can numerically code as "0" and "1", and when we can assume that the probability of belonging to one of the classes is conditional on the independent variables. In this case, the assumptions of linear regression are clearly not satisfied: The probability is mathematically limited to values between 0 and 1 and therefore cannot be proportional to the independent variables, which can have arbitrary real values. Likewise, the errors cannot be normally distributed, especially not around zero and with constant variance. When the independent variables model the probability of belonging to class "0" as low but nonzero, observations from class "0" with those values of the independent variables have a smaller error than observations from class "1". For the same reason, the variance at probabilities that are close to 0 or 1 must be much smaller than around probabilities close to 0.5.

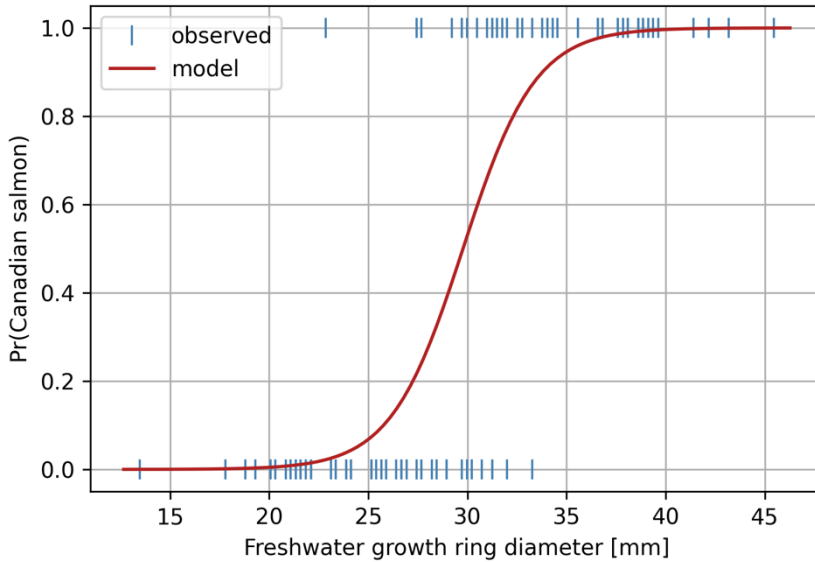
To circumvent these problems, Cox [10] proposed that the probability of belonging to class "1" be modeled by a logistic function. Cox did not explicitly justify this choice, stating only that "*in the absence of special considerations for a particular problem, the best form seems to be the logistic law*". However, from further derivation it is evident that the logistic function, compared to other sigmoid functions, has one attractive property. If we can represent the probability of belonging to class "1" as:

$$p = \frac{1}{1 + e^{-(\beta_0 + \boldsymbol{\beta}^T \mathbf{x})}}$$

then:

$$\ln \frac{p}{1 - p} = \beta_0 + \boldsymbol{\beta}^T \mathbf{x}$$

In other words, *log-odds* for belonging to class "1" are a linear model! Unlike linear regression, where the optimal coefficients β_j can be found analytically, by solving a linear system of equations, logistic regression is a nonlinear problem and the solution must be sought numerically. Figure 3 shows the result of a simple logistic regression, with only one independent variable, based on data from [11].



Sl. 3.: Logistic regression for predicting the species of salmon – Alaskan or Canadian. The x -axis shows the measured diameters of the freshwater growth rings, which are visible on the scales, and the y -axis shows the class to which each salmon belongs (blue, 1 = Canadian) and the modeled probability (red) of belonging to that class.

The function $x = \ln(p/(1 - p))$ is in statistics known under the name „*logit*“. As it is the inverse of the simple, standard logistic function, $p = 1/(1 + e^{-x})$, some practitioners, by analogy call the logistic function „*expit*“. It is the name under which standard logistic function can be found e.g. in Python’s `scipy.special` library, which also holds the function *logit*.

5. Folded functions

When summarizing empirical data, it is often useful to be able to express not only what values the data have, but also how much those values deviate from what is expected or common. For example, in a normal distribution, observations that are within one standard deviation of the center of the distribution are relatively common, unsurprising, while observations that are more than three standard deviations away are quite rare.

In [12], John Tukey considers a family of functions for – potentially nonlinear – scaling proportions p , i.e. values in the interval between 0 and 1, in order to quantify their deviation from some threshold value. According to Tukey, all functions in this family should satisfy the following conditions:

1. be monotonically increasing: $f'(p) > 0$;
2. map the proportion 0.5 onto 0: $f(0.5) = 0$;

3. be symmetrical around it's zero: $f(p) = -f(1 - p)$.

Additionally, for easier interpretation, it is desirable that the function around its zero grows as fast as the proportions themselves: $f'(p = 0.5) = 1$.

It is interesting that any monotonically increasing function g that is differentiable around 0.5 can be transformed into a function that satisfies the above conditions.:

$$f(p) = \frac{g(p) - g(1 - p)}{2g'(0.5)}$$

In the special case, for $g = \ln p$, we get

$$f(p) = \frac{1}{4} \ln \frac{p}{1 - p}$$

which is simply linearly scaled *logit*. Tukey gave that function the name “folded logarithm”, analogously to the “folded root”, which is obtained for $g = \sqrt{p}$.

6. “Soft” step function

6.1. Machine Learning as an alternative for statistical methods

Machine learning and pattern recognition have historically developed independently of statistics, with somewhat different goals, and differ from it in their approach to problems. Statistics, traditionally, seeks to explain data and summarize them in a form which allows for easy interpretation by humans, so that they can make decisions based on these explanations and summaries. On the other hand, machine learning was developed with the aim of enabling machines to make decisions autonomously, without human assistance. The ability to interpret the model plays a negligible role, and data summarization is only a method that attempts to remove noise (errors) from the data and distill the signal, for the purpose of optimizing the decisions.

Logistic regression is a method by which statisticians approach the problem of classification. The result of a logistic model is not a binary answer to which class an unknown observation belongs to – e.g. the salmon in the example in Fig. 3 – but only the probability of belonging to different classes. If a decision needs to be made or an action taken (e.g. to throw the salmon back into the sea), that decision is made by a human, based on many factors, the probability being only one of them.

The problem with logistic regression (and similar statistical methods) is that it requires very precise knowledge of the relationship between the independent and dependent variables. Everything except a finite and, as a rule, a very small number of parameters is assumed to be known. This, as we have seen, works well for simple problems, such as classifying salmon, but much worse for recognizing handwritten digits, and especially for recognizing faces or other multidimensional data.

Machine learning is less ambitious, but requires knowing only the general properties of the data. This is usually not enough to model probabilities, but that is not the aim. It is enough that the machine (computer) can give a clear-cut answer. Probabilities may or may not play a role in the background. Vladimir Vapnik expressed this in the words [13]: *“If you are limited to a restricted amount of information, do not solve the particular problem you need by solving a more general problem.”*

6.2. From logistic to Heaviside function and back

For the above reasons, early machine learning algorithms did not use the logistic or a similar sigmoid function, although it was, of course, already known and although logistic regression was discovered at about the same time. In these algorithms, especially the perceptron [14] and its variants, the machine internally calculated with real values, but, in the last step, to give an unambiguous answer, passed the real value through the Heaviside step function. When, in the mid-1980s, multilayer neural networks appeared, it turned out that for their training the output function of the neuron needed to be differentiable. Since the Heaviside function is simply a limiting case of the sigmoid – in the case of the logistic when the coefficient with the independent variable is infinite, or, in the Fermi-Dirac formulation, when the temperature $T = 0$ – in neural networks, the logistic function soon began to be used as the so-called “activation function” of neurons. The logistic function can be seen as a “soft variant” of the Heaviside function, because it has a similar shape, but avoids sharp transitions.

7. The connection between logistic function and normal distribution

Logistic function and normal distribution are directly related, which is probably one of the reasons why the logistic function works well in pattern recognition and why it forms the basis for logistic regression. If we have two classes with a normal distribution and identical variance, or, in the multidimensional case, with identical covariance matrix, which differ only in the coordinates of their centers, the logistic function is exactly the conditional probability of belonging to one of the two classes.

Independently of the distribution, if we have observations from two classes, A and B, with feature $\mathbf{x}$, which can be modeled as a random variable, the conditional probability that an unknown observation with the feature value $\mathbf{x}$ belongs to class B is:

$$P(B|\mathbf{x}) = \frac{P(\mathbf{x}|B)P(B)}{P(\mathbf{x})} = \frac{P(\mathbf{x}|B)P(B)}{P(\mathbf{x}|A)P(A) + P(\mathbf{x}|B)P(B)} = \frac{1}{1 + \frac{P(\mathbf{x}|A)P(A)}{P(\mathbf{x}|B)P(B)}}$$

Assuming $\mathbf{x}$ is numeric (potentially a vector), so that the classes can be described

by their corresponding probability density functions, $f_A(\mathbf{x})$ i $f_B(\mathbf{x})$, the ratio $P(\mathbf{x}|A)/P(\mathbf{x}|B)$ can be represented as:

$$\frac{P(\mathbf{x}|A)}{P(\mathbf{x}|B)} = \frac{f_A(\mathbf{x})}{f_B(\mathbf{x})}$$

and

$$\frac{P(\mathbf{x}|A)P(A)}{P(\mathbf{x}|B)P(B)} = \frac{f_A(\mathbf{x})P(A)}{f_B(\mathbf{x})P(B)}$$

If both classes follow a normal, potentially multivariate distribution around their centers, $\boldsymbol{\mu}_A$ i $\boldsymbol{\mu}_B$, having a common covariance matrix $\boldsymbol{\Sigma}$, then:

$$\begin{aligned} \frac{f_A(\mathbf{x})P(A)}{f_B(\mathbf{x})P(B)} &= \frac{e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_A)^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_A)} P(A)}{e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_B)^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_B)} P(B)} \\ &= e^{-\frac{1}{2}(\boldsymbol{\mu}_A+\boldsymbol{\mu}_B)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_A-\boldsymbol{\mu}_B)-\ln\frac{P(B)}{P(A)}-(\boldsymbol{\mu}_B-\boldsymbol{\mu}_A)^T\boldsymbol{\Sigma}^{-1}\mathbf{x}} \end{aligned}$$

so that $P(B|\mathbf{x})$ is again logistic function:

$$P(B|\mathbf{x}) = \frac{1}{1 + e^{-(\beta_0 + \boldsymbol{\beta}^T \mathbf{x})}}$$

with

$$\beta_0 = \frac{1}{2}(\boldsymbol{\mu}_A + \boldsymbol{\mu}_B)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_A - \boldsymbol{\mu}_B) + \ln\frac{P(B)}{P(A)}$$

and

$$\boldsymbol{\beta} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_B - \boldsymbol{\mu}_A)$$

In one-dimensional case, we have only a scalar variable x_1 , and instead of the covariance matrix we just have the variance σ^2 , so the coefficients of the logistic function can be written as:

$$\beta_0 = \frac{\mu_A^2 - \mu_B^2}{2\sigma^2} + \ln\frac{P(B)}{P(A)} \quad \text{i} \quad \beta_1 = \frac{\mu_B - \mu_A}{\sigma^2}$$

Here too, the conclusion that, as a consequence of the central limit theorem, normally distributed classes are a common occurrence applies, so that logistic regression is generally a good choice in practice.

Conclusion

Logistic function appears in many different branches of science and in many of them plays a very important role. Although it is only one of an infinite number of sigmoid functions, in today's artificial intelligence, especially in neural networks, it has been so widely used that many users think that they are synonyms. Interestingly, the logistic function can be obtained in many different ways, which is also one of the factors that makes it “elementary”.

References

- [1] P.-F. Verhulst, Notice sur la loi que la population poursuit dans son accroissement. *Corresp. Math. Phys.* 10, 1838, 113–121.
- [2] P.-F. Verhulst, Recherches mathématiques sur la loi d’accroissement de la population. *Nouv. Mém. Acad. R. Sci. B.-lett. Brux.* 18, 1845, 1–45.
- [3] P.-F. Verhulst, Deuxième mémoire sur la loi d’accroissement de la population. *Mém. Acad. R. Sci. Lett. B.-Arts Belg.* 20, 1847, 142–173.
- [4] T. Malthus, *An essay on the principle of population as it affects the future improvement of society*, J. Johnson, London, 1798.
- [5] A. Quetelet, *Sur l’homme et le développement de ses facultés*. Bachelier, Paris, 1835.
- [6] A.V. Hill, The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. *Proceedings of the physiological society*, 22.01.1910 iv–vii.
- [7] L. Michaelis, M.L. Menten, Die kinetic der Invertinwirkung, *Biochem. Z.*, 49, 1913 333–369.
- [8] E. Fermi, Sulla quantizzazione del gas perfetto monoatomico, *Rendiconti Lincei* 3, 1926, 145–149.
- [9] P.A.M. Dirac, On the Theory of Quantum Mechanics, *Proceedings of the Royal Society A*. 112 (762), 1926, 661–677.
- [10] D.R. Cox, The Regression Analysis of Binary Sequences, *Journal of the Royal Statistical Society B*, XX (2), 1958, 215–232.
- [11] R.A. Johnson, D.W. Wichern, *Applied Multivariate Statistical Analysis*, Pearson Prentice Hall, Upper Saddle River, NJ, USA, 2007, 604–605.
- [12] J.W. Tukey, *Exploratory Data Analysis*, Addison-Wesley Publishing Company, Reading, MA, USA, 1977, 498–502.

- [13] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag New York, Inc, New York, NY, USA, 1995, 169.
- [14] F. Rosenblatt, The perceptron: A probabilistic model for information storage and organization in the brain, *Psychological Review*, 65 (6), 1958, 386-407.

POVIJEST I PRIMJENA LOGISTIČKE FUNKCIJE: OD RASTA POPULACIJE DO UMJETNE INTELIGENCIJE

Sažetak: Logistička funkcija je poznata pod raznim imenima i koristi se u brojnim vrlo različitim područjima, od kvantne fizike i kemije do ekologije, statistike i umjetne inteligencije. Fizičarima je poznata kao Fermi-Diracova razdioba, koja opisuje statističku razdiobu čestica sa polovičnim spinom po energetske razinama. Kemičari, farmakolozi i toksikolozi je poznaju pod imenom „Hillova jednačba“, kojom opisuju efekt reakcije u ovisnosti o koncentraciji tvari. Logistička funkcija modelira rast populacije u ograničenom okolišu, a u statistici i strojnom učenju koristi se za modeliranje vjerojatnosti da promatrani uzorak pripada jednom od dva razreda. Kao „meka“ varijanta Heavisideove stepenaste funkcije koristi se često kao aktivacijska funkcija u zadnjem sloju neuronskih mreža i mnogi praktičari je smatraju kanoničkom sigmoidnom funkcijom. Široka primjenjivost logističke funkcije je posredna posljedica centralnog graničnog teorema.

Ključne riječi: logistička funkcija, Hillova jednačba, Fermi-Diracova razdioba