Mathematical demonstration of the Multilayer Perceptron and Backpropagation

Rômulo L. Pauliv
Ponta Grossa, Paraná, Brazil
romulopauliv1999@gmail.com

May 17, 2024

Abstract

The aim of this paper is to establish a solid foundation for the mathematical treatment of artificial neural networks in a didactic and clear manner. To achieve this goal, we will mathematically demonstrate the progression from a simple perceptron to the generalization of a multilayer perceptron (MLP), along with the theoretical learning algorithm based on backpropagation.

Keywords ANN · backpropagation · ML · MLP

1 Introduction

Artificial neural networks (ANNs) are widely recognized for their effectiveness in pattern classification and recognition in datasets, owing to their learning capabilities. This raises the fundamental question of how to characterize learning in a non-human environment and represent it mathematically.

When discussing learning, it is often associated with concepts such as the acquisition, retention, and application of information in various contexts. However, a deeper analysis links learning to the concept of error and iterative processes.

This factor arises from the observation that when attempting to achieve a specific outcome without knowing the precise method, we are prone to failure. By repeatedly performing the action, it becomes possible to identify the action that produces the desired result. Once identified, we can discern which actions lead to the expected outcome and which do not. Thus, one of the outcomes of learning is the ability to recognize when something deviates from previously observed reality, i.e., an error.

In this context, it is necessary to evaluate an object based on its observed properties to assign it different states. If a state is found to be inconsistent with previously observed reality, the evaluative action must be adjusted. Therefore, if we attribute learning capability to an object, this object must also be able to assign different states to other objects based on its perception, continually adjusting the mechanism that makes such assignments.

To this end, Frank Rosenblatt coined the term "perceptron" in his work "Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms." We will begin our mathematical approach to ANNs by developing an elementary object endowed with perception, the perceptron. Subsequently, we will enhance this perceptron to increase its complexity, enabling it to manage and correlate more information, thus creating a multilayer perceptron (MLP). Finally, we will endow it with the capacity for judgment through its perception.

2 Perceptron

Our objective is to construct a perceptron described by a function f such that $f : \mathbb{R}^n \to \Theta$, where Θ represents any possible codomain. This implies that in this function, the independent variables or inputs will represent the stimulus to the perceptron, and the dependent variable or output will represent the response to this stimulus.

2.1 Inputs

The input will be denoted by a vector X such that $X \in \mathbb{R}^n$. Therefore, we have $X^T = [x_1, \ldots, x_n]$, where the superscript T represents the transpose of the vector X. Note that the vector does not include the element x_0. This is because we will reserve this term for what will later be introduced as the bias. Consequently, the inputs or independent variables will be listed as $i = 1, \ldots, n$.

2.2 Outputs

O output será denotado por um vetor Y tal que $Y \in \Theta$. Logo, temos que $Y = [y_1]$. No output não haverá a adição do bias futuramente mas evitaremos utilizar o índice zero para evitar enganos futuros. O domínio Θ representa qualquer domínimo possível tal que Θ será ditado pela função de ativação que iremos introduzir em breve. No perceptron teremos o vetor Y com apenas uma dimensão. Futuramente iremos introduzir um vetor de saída com m-outputs. Logo, os outputs ou variáveis dependentes serão listado de $i = 1, \ldots, m$.

2.3 Weights

Os pesos serão parte substancial de nosso perceptron. Tais pesos irão transformar nossos n-inputs a fim de resultar em elementos de Y. Logo, os pesos serão denotados por ϖ tal que $\varpi \in \mathbb{R}^n$.

2.4 Headings: second level

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetuer.

$$\xi_{ij}(t) = P(x_t = i, x_{t+1} = j | y, v, w; \theta) = \frac{\alpha_i(t) a_{ij}^{w_t} \beta_j(t+1) b_j^{v_{t+1}}(y_{t+1})}{\sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i(t) a_{ij}^{w_t} \beta_j(t+1) b_j^{v_{t+1}}(y_{t+1})} \tag{1}$$

2.4.1 Headings: third level

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Paragraph Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

3 Examples of citations, figures, tables, references

Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Donec odio elit, dictum in, hendrerit sit amet, egestas sed, leo. Praesent feugiat sapien aliquet odio. Integer vitae justo. Aliquam vestibulum fringilla lorem. Sed neque lectus, consectetuer at, consectetuer sed, eleifend ac, lectus. Nulla facilisi. Pellentesque eget lectus. Proin eu metus. Sed porttitor. In hac habitasse platea dictumst. Suspendisse eu lectus. Ut mi mi, lacinia sit amet, placerat et, mollis vitae, dui. Sed ante tellus, tristique ut, iaculis eu, malesuada ac, dui. Mauris nibh leo, facilisis non, adipiscing quis, ultrices a, dui. [1, 2] and see [3].

The documentation for `natbib` may be found at

`http://mirrors.ctan.org/macros/latex/contrib/natbib/natnotes.pdf`

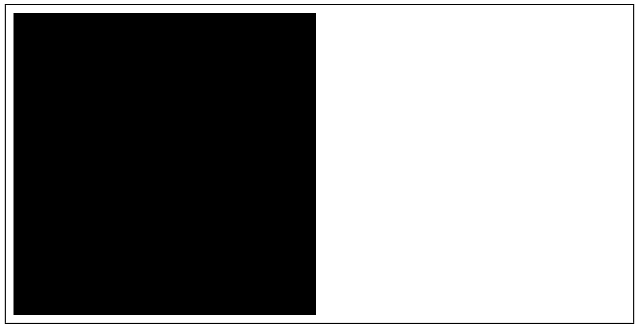

Figure 1: Sample figure caption.

	item_name	item_id	item_category_id
0	! ВО ВЛАСТИ НАВАЖДЕНИЯ (ПЛАСТ.) D	0	40
1	!ABBYY FineReader 12 Professional Edition Full...	1	76
2	***В ЛУЧАХ СЛАВЫ (UNV) D	2	40
3	***ГОЛУБАЯ ВОЛНА (Univ) D	3	40
4	***КОРОБКА (СТЕКЛО) D	4	40

Of note is the command `\citet`, which produces citations appropriate for use in inline text. For example,

```
\citet{hasselmo} investigated\dots
```

produces

Hasselmo, et al. (1995) investigated…

`https://www.ctan.org/pkg/booktabs`

3.1 Figures

Suspendisse vitae elit. Aliquam arcu neque, ornare in, ullamcorper quis, commodo eu, libero. Fusce sagittis erat at erat tristique mollis. Maecenas sapien libero, molestie et, lobortis in, sodales eget, dui. Morbi ultrices rutrum lorem. Nam elementum ullamcorper leo. Morbi dui. Aliquam sagittis. Nunc placerat. Pellentesque tristique sodales est. Maecenas imperdiet lacinia velit. Cras non urna. Morbi eros pede, suscipit ac, varius vel, egestas non, eros. Praesent malesuada, diam id pretium elementum, eros sem dictum tortor, vel consectetuer odio sem sed wisi. See Figure 1. Here is how you add footnotes. [1] Sed feugiat. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Ut pellentesque augue sed urna. Vestibulum diam eros, fringilla et, consectetuer eu, nonummy id, sapien. Nullam at lectus. In sagittis ultrices mauris. Curabitur malesuada erat sit amet massa. Fusce blandit. Aliquam erat volutpat. Aliquam euismod. Aenean vel lectus. Nunc imperdiet justo nec dolor.

3.2 Tables

Etiam euismod. Fusce facilisis lacinia dui. Suspendisse potenti. In mi erat, cursus id, nonummy sed, ullamcorper eget, sapien. Praesent pretium, magna in eleifend egestas, pede pede pretium lorem, quis consectetuer tortor sapien facilisis magna. Mauris quis magna varius nulla scelerisque imperdiet. Aliquam non quam. Aliquam porttitor quam a lacus. Praesent vel arcu ut tortor cursus volutpat. In vitae pede quis diam bibendum placerat. Fusce elementum convallis neque. Sed dolor orci, scelerisque ac, dapibus nec, ultricies ut, mi. Duis nec dui quis leo sagittis commodo. See awesome Table 1.

[1] Sample of the first footnote.

Table 1: Sample table title

Part		
Name	Description	Size (μm)
Dendrite	Input terminal	$\sim$100
Axon	Output terminal	$\sim$10
Soma	Cell body	up to 10^6

3.3 Lists

- Lorem ipsum dolor sit amet
- consectetur adipiscing elit.
- Aliquam dignissim blandit est, in dictum tortor gravida eget. In ac rutrum magna.

References

[1] George Kour and Raid Saabne. Real-time segmentation of on-line handwritten arabic script. In *Frontiers in Handwriting Recognition (ICFHR), 2014 14th International Conference on*, pages 417–422. IEEE, 2014.

[2] George Kour and Raid Saabne. Fast classification of handwritten on-line arabic characters. In *Soft Computing and Pattern Recognition (SoCPaR), 2014 6th International Conference of*, pages 312–318. IEEE, 2014.

[3] Guy Hadash, Einat Kermany, Boaz Carmeli, Ofer Lavi, George Kour, and Alon Jacovi. Estimate and replace: A novel approach to integrating deep neural networks with existing applications. *arXiv preprint arXiv:1804.09028*, 2018.