# Mathematical Approaches to Automatic Learning: Convolutional Neural Networks

Agnese Adorante

May 28, 2024

## Introduction: The Essential Role of Automatic Learning

Automatic Learning, also known as Machine Learning, is a field of study in Artificial Intelligence focused on the development of algorithms that can learn from data and generalize to unseen data, thereby performing tasks without explicit instructions [1].

A clear example of the revolutionary impacts of ML can be found in the field of Computer Vision[1]. Indeed, teaching a computer how to recognize an object in a digital image is a complex task that traditional programming methods struggle to accomplish effectively.

Let's say we want to teach a computer how to recognize a chair, if we were to do it the 'standard' way we would write a series of if else statements, looking for specific characteristics. For instance, we might instruct our program to identify objects with four legs as a chairs. However, this approach would mistakenly classify any table as a chair unless additional conditions were added. Furthermore, we would need to account for chairs with varying numbers of legs, swivel chairs, rocking chairs, and other variations. A non-ML program able to account for this many exceptions would need millions of lines of code, would be very slow and still have a high error rate.

This simple example shows why we need automatic learning: instead of deciding a priori what are the relevant characteristics to perform our task, we should let the model itself decide those.

In the realm of Computer Vision the most popular architectures are called Convolutional Neural Networks (CNNs), which leverage mathematical properties and intuitions to achieve high performance and accuracy. This essay will explore the mathematical foundations that underpin CNNs, demonstrating how these principles contribute to their effectiveness in automatic learning.

## Brief History of Automatic Learning

Autmatic Learning, despite its recent surge in popularity, has historical roots dating back several decades.

The first model of an artificial neuron was developed by McCullogh & Pitts in 1943 and the first Perceptron[2] dates back to 1958.

After a period of general distrust towards the whole field, called AI winter, advancements in processor technology and algorithmic breakthroughs revived interest in AI. [2]

A pivotal moment came in 2012 with the success of AlexNet[3]. This CNN architecture was the best one developed at the ImageNet Large Scale Visual Recognition Challenge[3]. It achieved a top-5 error

---

[1] Computer vision is an interdisciplinary field that deals with how computers can be made to gain high-level understanding from digital images or videos.

[2] The Perceptron is a binary classifiers. A binary classifier is a function which can decide whether or not an input, represented by a vector of numbers, belongs to some specific class.

[3] The ILSVRC is an annual computer vision competition developed upon a publicly available computer vision dataset called ImageNet.

rate[4] of 15.3%, significantly outperforming previous best results.[4]

Today, image classification models not only match but exceed human accuracy. As we look to the future, the continuous evolution of AI promises to address complex challenges in ways we are only beginning to imagine.

# Mathematical Foundations of Convolutional Neural Networks

### Translation Invariance: Overcoming Positional Dependencies

Basic Neural Networks can be inefficient when processing images due to the inherent structural characteristics of visual data. Distinctive features must be recognized independently of their position within the image. This property, known as translation invariance, is challenging to impart to a network.

The great intuition of Krizhevsky and Sutskever, the minds behind AlexNet (building on LeCun's earlier work), was the following: instead of letting the network learn that it should perform the same operations in a translation-invariant way, define an operation, the convolution, and apply it in a translation-invariant fashion. This principle underlies CNNs and it is the reason why they're also known as Shift or Space Invariant Artificial NNs.

CNNs consist of convolutional layers that function as specialized versions of fully connected layers, optimized to process inputs with known geometrical structures. In these layers the convolution operation is applied to inputs. Small filters, known as kernels, play a crucial role. A kernel is a small matrix of weights that moves across the input image, like a tiny window sliding over a larger picture. Each position of the kernel over the image produces a single number, which is a weighted sum of the pixel values within that window. This process helps the network to recognize patterns, such as edges, or shapes, no matter where they appear in the image. The result of this operation is a set of new matrices called feature maps.

### The Equivariance Property

Kernels are said to be equivariant with respect to translations. Equivariance is defined as follows: suppose that we have some data $x$ that belongs to an input space $\mathcal{X}$ endowed with some symmetry[5]. Call $T \in \mathcal{T}$ one element of the symmetry, for example a particular translation. Call $f : X \to \tilde{\mathcal{X}}$ a map from the input space to some other space $\tilde{\mathcal{X}}$. Suppose that the other space $\tilde{\mathcal{X}}$ also has the same symmetry, and thus that we can operate with the same transformation $T$ on both the $X$ and $\tilde{\mathcal{X}}$ spaces. We say that such a map $f$ is equivariant with respect to a family of transformations $\mathcal{T}$ (the symmetry group) if it commutes[6] with every member of the group: $\forall T \in \mathcal{T} : f \circ T = T \circ f$ [7]

Going back to our use case: in an image classification task, the class of an image should be approximately invariant to translations. It's easy to see that slightly translating an image and then applying a convolutional filter produces approximately the same result as applying the filter first and then translating. Therefore, convolutional filters are approximately equivariant. [2]

# Conclusion: The Future of Convolutional Neural Networks

In summary, CNNs leverage the underlying mathematical properties of visual data, enabling computer programs to accomplish tasks that were previously inconceivable using traditional programming methods. The profound impact of Automatic Learning extends across various fields, particularly in scientific research, where CNNs have demonstrated remarkable applications.

---

[4]The top-5 error is the percentage of times that the target label does not appear among the 5 highest-probability predictions.

[5]The symmetry group of a geometric object is the group of all transformations under which the object is invariant.

[6]An operation is commutative if changing the order of the operands does not change the result.

[7]∘ denotes composition of functions. If you have two functions, say $f$ and $g$, then $f \circ g$ means you first apply $g$, and then apply $f$ to the result of g

CNNs are capable of reconstructing astronomical images and identifying celestial bodies and phenomena. By processing vast amounts of telescope data, they facilitate the discovery of new stars, planets, and galaxies, accelerating scientific breakthroughs by centuries.

Furthermore, these technologies enable the creation of digital twins of various areas of the animal brain, particularly CNNs serve as (and were also inspired by) digital reconstructions of the Visual Cortex. Through such advancements, scientific discoveries can be made in a faster and more ethical manner, potentially reducing the need for experiments on animals.

Reflecting on CNNs as a prime example, it becomes evident that by exploiting the mathematical properties inherent in our tasks, we have the potential to achieve extraordinary milestones and develop even more impactful technologies. Harnessing mathematics, we can forge a path toward a future where technology acts as a driving force for progress and positive development in many fields.

# References

[1]  John R. Koza et al. *Automated Design of Both the Topology and Sizing of Analog Electrical Circuits Using Genetic Programming*. Ed. by John S. Gero and Fay Sudweeks. Dordrecht: Springer Netherlands, 1996, pp. 151–170. URL: https://doi.org/10.1007/978-94-009-0279-4_9.

[2]  Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. http://www.deeplearningbook.org. MIT Press, 2016.

[3]  Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. "ImageNet classification with deep convolutional neural networks". In: (2012).

[4]  Olga Russakovsky et al. "ImageNet Large Scale Visual Recognition Challenge". In: *International Journal of Computer Vision (IJCV)* 115.3 (2015), pp. 211–252. DOI: 10.1007/s11263-015-0816-y.