1 Introduction

The last decade has witnessed an experimental revolution in data science and machine learning, epitomised by deep learning methods. Indeed, many high-dimensional learning tasks previously thought to be beyond reach – such as computer vision, playing Go, or protein folding – are in fact feasible with appropriate computational scale. Remarkably, the essence of deep learning is built from two simple algorithmic principles: first, the notion of representation or *feature learning*, whereby adapted, often hierarchical, features capture the appropriate notion of regularity for each task, and second, learning by local gradient-descent, typically implemented as *backpropagation*.

While learning generic functions in high dimensions is a cursed estimation problem, most tasks of interest are not generic, and come with essential pre-defined regularities arising from the underlying low-dimensionality and structure of the physical world. This text is concerned with exposing these regularities through unified geometric principles that can be applied throughout a wide spectrum of applications.

Exploiting the known symmetries of a large system is a powerful and classical remedy against the curse of dimensionality, and forms the basis of most physical theories. Deep learning systems are no exception, and since the early days researchers have adapted neural networks to exploit the low-dimensional geometry arising from physical measurements, e.g. grids in images, sequences in time-series, or position and momentum in molecules, and their associated symmetries, such as translation or rotation. Throughout our exposition, we will describe these models, as well as many others, as natural instances of the same underlying principle of geometric regularity.

Such a 'geometric unification' endeavour in the spirit of the Erlangen Program serves a dual purpose: on one hand, it provides a common mathematical framework to study the most successful neural network architectures, such as CNNs, RNNs, GNNs, and Transformers. On the other, it gives a constructive procedure to incorporate prior physical knowledge into neural architectures and provide principled way to build future architectures yet to be invented.

Before proceeding, it is worth noting that our work concerns *representation learning architectures* and exploiting the symmetries of data therein. The many exciting *pipelines* where such representations may be used (such as