

Probabilistic symmetry and convolutional neural networks

Project Outline

1 Title

The working title of my project is *Probabilistic symmetry and convolutional neural networks*.

2 Background

Deep learning has revolutionized the research field of machine learning due to its powerful accuracy and generalizability on large, high dimensional datasets. The convolutional neural network (CNNs) architecture has become the mainstay for constructing computer vision platforms. The chief design principle of CNNs is the incorporation of convolutional layers that allow the network to detect certain features in the input image. In particular, these convolutional layers are equivariant under translation of the image, in the sense that translations of the input image cause the activations of the convolutional layers to translate in the same way. This design feature is the key reason why convolutional neural networks have achieved excellent generalizability on a variety of datasets.

In the last decade, there has been much work in expanding mathematical understanding of deep learning algorithms. Deep learning algorithms can be formalized under two separate (but not mutually exclusive) points of views: they can be viewed as deterministic functions $f : \mathcal{X} \rightarrow \mathcal{Y}$ where \mathcal{X} is the set of input features and \mathcal{Y} is the set of labels, or they can be viewed from a probabilistic point of view so that $Y = f(X, \eta)$, where X and Y are the input and output random variables, and η is some outsourced noise random variable.

Viewing deep learning algorithms deterministically, Kondor and Trivedi [KT18] proved that f containing a convolution operation $\int f(uv^{-1})g(v)$ is both necessary and sufficient for the layers of a neural network $y = f(x)$ to be equivariant. On the other hand, when viewed from a probabilistic point of view, Bloem-Reddy and Teh [BT19] provided necessary and sufficient conditions for a neural network $Y = f(X, \eta)$ to be invariant or equivariant under the actions of a compact group.

The conditions posed by Bloem-Reddy and Teh on the structure of $Y = f(X, \eta)$ are quite general and do not, by themselves, restrict f to be a function involving convolution, as is the case in Kondor and Trivedi's results. Then, a natural question to ask is: what further conditions on $Y = f(X, \eta)$ are necessary so that f is restricted to functions involving convolution?

3 Technical aspects

The project will draw on technical aspects of the following areas: probabilistic symmetry, machine learning, group theory, fourier analysis, and representation theory.

4 Literature

The key references for this project are [KT18] and [BT19], as stated previously.

5 Plan

I will carry out this project with the following sequence of steps:

1. I will summarize the key points from [BT19] and [KT18], and provide links between the concepts discussed in both papers.
2. I will attempt to derive necessary conditions on $Y = f(X, \eta)$ so that f is a function involving a convolution operator.
3. If my attempts in deriving necessary conditions are not successful, I will explain why my candidate conditions do not invoke necessity.

6 Why I'm interested in this topic

I am interested in the mathematical foundations of machine learning, and I'm particularly excited in understanding how group theory, a topic I learned in my undergraduate degree in mathematics, relates to statistical learning theory.

References

- [KT18] R. Kondor and S. Trivedi. *On the Generalization of Equivariance and Convolution in Neural Networks to the Action of Compact Groups*. 2018. arXiv: 1802.03690 [stat.ML].
- [BT19] B. Bloem-Reddy and Y. W. Teh. *Probabilistic symmetry and invariant neural networks*. 2019. arXiv: 1901.06082 [stat.ML].