XAI: Variational Inference for Cluster

Jack Li

March 2023

Abstract

The purpose of this report is to introduce the Cluster algorithm and experiment with it.

1 Introduction

The primary tasks of Machine Learning encompass a variety of fields, including classification, segmentation, and modern techniques such as generative models. One of the fundamental skills in machine learning is classification. Specifically, data can be categorized into labeled and unlabeled datasets. This encompasses numerous techniques from supervised and unsupervised learning. A key feature of unsupervised learning is clustering, which is an algorithm designed to group data into distinct clusters. Generally speaking, clustering is an unsupervised learning technique. The objective of clustering is to partition the data into different clusters based on inherent similarities. One of the most challege is poepole offen share same experience that it is hard to understand how model works anaylitically. It will cause unpredictable results like medical diagnosis, financial fraud detection, and autonomous driving after industrial adoption. It is important to understand the model and how it works. This is where the concept of explainable AI (XAI) comes in. XAI is a subfield of artificial intelligence (AI) that focuses on making AI models more transparent and understandable to humans.

In this paper, we explore some state of the art XAI techniques and how they can be interpret the features in images. We learn from foundation of machine learning and deep learning that the imporved predicitive accuracy has often been increased model complexity. The obvious drawback is the E_{in} decrease error, however, it might increase the E_{out} error [1, 2, 3]. To better understand the model, we need to interfere the model's decision-making process. While we persuade the idea of XAI, we will introduce the Variational Inference Clustering algorithm and experiment with it. The Variational Inference Clustering algorithm is a popular XAI algorithm that can be used to interpret the features in images.

2 Related Work

...

The GradCAM,[4], algorithm is a popular XAI algorithm.

3 Experiments

...

PyTorch Conv2d Equation

The output size of a Conv2d layer can be calculated using the following equation:

$$\text{Output Size} = \left\lfloor \frac{\text{Input Size} + 2 \times \text{Padding} - \text{Kernel Size}}{\text{Stride}} \right\rfloor + 1$$

Where: - Input Size is the size of the input feature map (height or width). - Padding is the number of zero-padding added to both sides of the input. - Kernel Size is the size of the convolution kernel (height or width). - Stride is the stride of the convolution.
PyTorch ConvTranspose2d Equation

The output size of a ConvTranspose2d (transposed convolution) layer can be calculated using the following equation:

$$\text{Output Size} = (\text{Input Size} - 1) \times \text{Stride} - 2 \times \text{Padding} + \text{Kernel Size} + \text{Output Padding}$$

Where: - Input Size is the size of the input feature map (height or width). - Stride is the stride of the convolution. - Padding is the number of zero-padding added to both sides of the input. - Kernel Size is the size of the convolution kernel (height or width). - Output Padding is the additional size added to the output (usually used to ensure the output size matches a specific value).

4 Conclusion

...

References

[1] Shai Shalev-Shwartz and Shai Ben-David. *Understanding Machine Learning: From Theory to Algorithms.* Cambridge University Press, New York, NY, USA, 2014.

[2] Luc Devroye, László Györfi, and Gábor Lugosi. *A Probabilistic Theory of Pattern Recognition.* Springer, New York, 1996.

[3] Trevor Hastie, Robert Tibshirani, and Jerome H. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer Series in Statistics. Springer, New York, NY, second edition, corrected at 12th printing 2017 edition, 2017.

[4] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. *International Journal of Computer Vision*, 128(2):336–359, February 2020.