Stockholm University

Mathematical Statistics
Stockholm University
Master Thesis **2022:6**
http://www.math.su.se

Unveiling the inner mechanisms of deep convolutional neural networks through the lens of unsupervised learning

Hilding Köhler*

June 2022

Abstract

To understand the world it is vital to find answers to the questions how, why and where. This is also true when modelling data where these questions are asked to better understand the model's working mechanisms. By understanding model mechanisms, it is possible to explain what causes its results. In recent years the rise of complex models such as convolutional neural networks (CNN:s) has made it possible to produce high performance models, but they are difficult to understand at first sight. CNN:s are the models in focus for this thesis because they are the model standard for image classification. CNN:s are implemented in more fields making it important to understand their working mechanisms. To understand the working mechanisms of CNN:s, this thesis aims to unveil the mechanisms of the CNN, Residual neural network 18 (ResNet-18). Studying ResNet-18 is interesting due to its complex architecture, wide use and high accuracy. Understanding its working mechanisms provides insights into the results of this and other CNN:s. ResNet-18 is in this thesis trained and tested on the benchmark dataset CIFAR-10. The unveiling of the working mechanisms is done using cluster analysis and shape aware distances called commute time distances. This method analyses image clusters based on their characteristics using a distance that respects the underlying data structure. These techniques belong to the field of unsupervised learning, a field providing methods for analysing high dimensional data without needing the data labels. The power of the methods in this thesis is that they can be applied to different mechanisms of models while also avoiding complicated model fitting. The conclusion of the thesis is that all working mechanisms in ResNet-18 serve a purpose. However, the last convolutional operation is the cause for the good image classification which exhibits itself in the low test error of 5%.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: hildingkohler@gmail.com. Supervisor: Chun-Biu Li.