Reducing Data Storage Requirements for Machine Learning Algorithms using Principle Component Analysis

Saritha Kinkiri & Wim J.C. Melis

Faculty of Engineering and Science, University of Greenwich,
Chatham Maritime ME4 4TB, UK
k.sarithareddy20@gmail.com, Wim.J.C.Melis@greenwich.ac.uk

Abstract

While current computers have shown to be particular useful for arithmetic and logic implementations, their accuracy and efficiency for applications such as e.g. face, object and speech recognition, are not that impressive, especially when compared to what the human brain can do. Machine learning algorithms have been useful, especially for these type of applications, as they operate in a similar way to the human brain, by learning the data provided and storing it for future recognition. Until now, there has been a strong focus on developing the process of data storage and retrieval, merely neglecting the value of the provided information and the amount of data required to store. Hence, currently all information provided is stored, because it is difficult for the machine to decide which information needs to be stored. Consequently, large amounts of data are stored, which then affects the processing of the data. Thus, this paper investigates the opportunity to reduce data storage through the use of differentiation and combine it with an existing similarity detection algorithm. The differentiation is achieved through the use of, Principal Component Analysis (PCA), which not only reduces the data storage requirements by about 80%, but also improves the overall detection accuracy around 50 to nearly 80%.

Keywords: Machine Learning, Data Storage Efficiency, Principle Component Analysis (PCA)