Machine Learning Review Paper

The objective of dimensionality reduction in machine learning models is to improve its performance through noise reduction by removing redundant features; and minimizing over fitting by minimizing the number of features or dimensions in a dataset without losing the relevant information in the process. This can be done by either feature selection, which is identifying and discarding features based on predefined criteria such as mutual information or feature extraction which is transforming the features into lower dimension representation whilst maintain relevant information. An example can be seen in PCA and LASSO which performs this task by identifying and removing redundant or irrelevant features. This technique is useful in medical imaging.

pros

- Principal component analysis (PCA) can highlight the most important lines of variation in data which helps to detect patterns and anomalies in data set.
- It is computationally efficient in simplifying complex data in many fields such as medical imaging.

Cons

- There is possible loss of information during dimension reduction of a dataset as all dimensionality reduction techniques have their limitations such as identifying the ideal number of constraints to reduce.

The form, scientific style, spelling (review punctuations), and content were easy to follow. It would be nice to have some data sets that can explain the pictures e.g., a relatable use case for PCA. Also, a foundation of what dimensionality reduction is using a step-by-step example in simple terms would be great to ease understanding. Overall, it was a great read.