Mind the Gap: A Generative Approach to Interpretable Feature Selection and Extraction

Abraham Jose
ID :5068109, CAP6614
abraham@knights.ucf.edu

1 Summary

Mind the Gap Model(MGM) is a an approach to place interpretable criteria directly into the model. It helps to help interpret the model and to optimize parameter related to interpretability and helps us to understand and to assist further data analysis and hypothesis.

2 Strengths of the proposal

1. MGM offers a machine learning model that helps us to understand the principal components that creates a cluster which will aid the classification which is interpretable at the same time.
2. MGM offers a model that is well hypothesised and well explained by the theory.
3. Many of the models offers poor interpretability, leaving neural models as a black box. These approaches can shed some light and inspiration to create interpretable clustering in neural networks.

3 Weaknesses of the proposal

1. Eventhough MGM offers interpretable classifier, we will need to model better methodolgies to fetch data rather than binarization of input data. Binarization means poor word count or less information at once, hence we cannot consider fine graded classifications of any sort.
2. In real world, none of the entities or features has a logical linear relationship like simple AND and OR. It tends to have more complex relationship, influenced by many other features. Also, the computational requirement may be high.

4 Results

The idea to combine extractive and selective approaches for interpretable model has better performance compared to other classification algorithms like HFS. Even though MGM is second to HFS in Face dataset, it clearly dominates every other algorithm through the interpretable cluster details that the model offers. It can be used when performing unsupervised learning on high dimensional data sets.

5 Discussion

How the MGM model can be modified to take non-binary valued data and thus create a extractive and generative model with any types of input and without compromising the input data's fidelity.

How these systems can be incorporated to neural network models which will improve the interpretability of the model hence reduce the randomness associated with decision making.