

Deep Ensemble Transformers (with Marius Geitle)

Abstract: We propose deep ensemble transformers for dimensionality reduction problems. The method uses cascade ensemble methods as its building block for feature extraction and benefits from the high predictive power of deep neural networks. To sparsify high-dimensional data and prevent overfitting, a flexible number of intermediate layers in deep ensemble transformers sequentially maps the input into decision tree predictions. To boost prediction performance, the final intermediate layer output propagates further through the architecture of feed-forward neural networks and predicts the outcome. Investigation of the large sample theory shows that, under several regularity conditions and distributional assumptions, the method is consistent up to a proportionality constant. Experimental results demonstrate the superiority of the model in terms of prediction performance and representation learning ability. Specifically, the method achieves 100% accuracy in gene expression data and substantially improves upon benchmark methods in other biomedical data.