Explanation-Oriented Partitioning

In many applications of predictive analytics, the ability to explain the context of predictions is often as important as prediction accuracy. We introduce an algorithm called Explanation-Oriented Partitioning (EOP), a method that uses a small number of low-dimensional projections of data, each with its own discriminator, to explain the classification decision. This meta-algorithm can work with discriminators of various types, leveraging point-wise performance of the classifier to identify low-dimensional regions of the feature space where data is easily classifiable. EOP picks out multiple clusters of well-classified points from informative projections of data, maximizing expressiveness while maintaining compactness of the resulting models.

EOP iteratively selects projections where an accurate classifier can be found. In each of the selected projections, EOP identifies contiguous areas in which predictions are consistently correct. We have elaborated several heuristic methods of defining regions. The parametric version encases the points in polyhedra such that each polyhedron is maximal. The nonparametric regions are determined through a score based on proximity to well classified points. All regions are calibrated using a hold-out set to prevent overfitting. Ultimately, regions serve as explanations for predictions made for data inside their bounds, allowing the user to trace how the system made the prediction and to visually confirm its validity. At every iteration, EOP selects, support of the classification task, the most effective projection – discriminator among all subspaces of low dimensionality. The data that cannot be accurately explained – i.e. placed in a classifiable region – using the current model becomes the focus of the next iteration. The resulting model is hierarchical - each level specifying a low-dimensional projection, an associated set of regions and an accurate discriminator for the set.

We have compared EOP against Boosting, since, like Boosting, our algorithm seeks to apply classifiers to the tasks they perform best at. However, EOP is a white box model, so we also compared it to CART – a decision tree model. The evaluation was performed on artificial data with injected Gaussian patterns as well as five datasets from the UCI repository. Our results showed that the performance of EOP is comparable to that of Boosting and CART, while the models are more compact and easy to evaluate – we measure compactness by considering the information captured by the model and ease of evaluation by calculating the expected number of decisions required. Notably, we have remarked the high performance of EOP compared to CART for low-complexity models – after one projection, it uniformly outperforms CART on all datasets. We have also assessed the models in terms of expressiveness, as given by four metrics used to identify relevant rules – J-Score, Bayes Factor, Lift and Normalized Mutual Information. The empirical results show that EOP indentifies more interesting regions according to all metrics for both the real-world and synthetic data used.

To conclude, EOP algorithms are capable of finding expressive projections while maintaining accuracy high. The resulting models are compact and capture the essence of data in a format that is intuitive to the human users.