Classification with a Random Geometric Graph Family

Elvan Ceyhan

SAMSI
4501 Research Commons, Suite 300
79 T.W. Alexander Drive
Durham, NC 27709

NC State University
Dept. of Statistics, College of Sciences
4264 SAS Hall, 2311 Stinson Dr.
Raleigh, NC 27695

We employ a geometric graph family called class cover catch digraphs (CCCDs) for classification of data in low to high dimensions. CCCDs are constructed based on spherical proximity regions --- regions that determine the presence and direction of the arcs in the digraph--- and emerged as a graph theoretic solution to the class cover problem. We assess the classification performance of CCCD classifiers by extensive Monte Carlo simulations, comparing them with other classifiers commonly used in the literature. In particular, we show that CCCD classifiers perform well when one class is more frequent than the other in a two-class setting, an example of the class imbalance problem. That is, CCCD classifiers are robust to the class imbalance problem in statistical learning. We also point out the relation between class imbalance and class overlapping problems, and their influence on the performance of CCCD classifiers and other classification methods including some of the state-of-the-art algorithms which are also robust to the class imbalance. CCCDs --- by construction --- tend to substantially under-sample from the majority class while preserving the information on the discarded points during the under-sampling process. While many state-of-the-art methods keep this information by means of ensemble classifiers, CCCDs yield only a single classifier with the same property, making it both simple and fast.