Project: Utilizing Voronoi Diagrams for Multi-Label K-Nearest Neighbor (ML-KNN) Algorithm

Description:

Traditional supervised learning task is single-label, i.e. an instance is classified into one label among many. Multi-label learning is, on the other hand, a learning paradigm in which an instance can be classified into many labels at the same time. It is considered to be a hard task by its nature due to the fact that number of possible classifications for an instance is exponential in terms of the number of labels (having $2^L$ subsets for L labels).

ML-KNN is one of the most cited papers in the multi-label learning community (currently having 1342 citations in Google Scholar), due to the following reasons: (1) it is intuitive and easy to understand once a person knows the single-label counterpart of this algorithm, which is the traditional kNN, (2) it is a lazy learner, i.e. it does not learn a discriminatory function with the training data but rather just stores the training data and does its calculations for classification whenever a query is received. Hence, it is a widely-used baseline method for any multi-label algorithm.

In this project, I'm planning to make use of computational geometric objects such as Voronoi Diagrams to fasten the process of finding k-nearest neighbors of a given instance. For this purpose, First Order or Higher Order Voronoi Diagrams can be utilized. However, the process may become less and less tractable as k increases. Also, the data points have more than three attributes (in fact, some data have >30 attributes in some datasets), which is in very high dimensional space and prone to the curse of dimensionality.

Therefore, the following must be done for the project: (1) Dimensionality reduction on the data to embed them into a plane (methods such as PCA can be utilized for this purpose), (2) conducting experiments on how Voronoi Diagrams behave as the parameter k changes, (3) comparison of standard brute-force ML-kNN with the one that uses Voronoi Diagrams in terms of execution time and memory.

References