Formulation of Multiple Kernel Learning Using Composite Architectures

A thesis submitted

in partial fulfillment for the award of the degree of

Doctor of Philosophy

by

Shiju S. S.

Department of Mathematics

Indian Institute of Space Science and Technology

Thiruvananthapuram, India

February 2019

Abstract

Multiple Kernel Learning (MKL) algorithms deal with learning the optimal kernel from training data along with learning the function that generates the data. Generally in MKL, the optimal kernel is defined as a combination of kernels under consideration (which are usually termed as base kernels). There are different approaches to learning the optimal combination of kernels. The main objective of the thesis is to develop models for finding the optimal combination of kernels suitable for different types of problems.

This thesis describes the formulation of optimal kernel using four different approaches. The first approach is based on the selection of kernel suitable for the features (attributes) of the data. The features having similar characteristics are clustered together and a suitable kernel is found for each cluster. The optimal kernel is defined as a linear combination of the kernels defined over the cluster subspaces. We formulated a methodology for clustering the features and applying a separate kernel over each cluster.

The second approach is based on Kumar et. al. (2012) [1] in which the problem of learning the kernel for a binary classification is designed as another binary classification problem and hence the data modeling problem involves the computation of two decision boundaries of which one is related with that of MKL and the other with that of input data. They used two different cost functions for finding the optimal function related with kernel learning and the classification task. We modified this work in such a way that in our approach, the optimal functions are found with the aid of a single cost function by constructing a global reproducing kernel Hilbert space (RKHS). This global RKHS is defined as the direct sum of the RKHSs corresponding to the decision boundaries of MKL and that of input data. Hence the optimal function can be represented as the direct sum of the decision boundaries under consideration.

Kumar et. al. (2012) framework has been extended to regression problems also. We also developed a nonlinear formulation, in which the optimal kernel is represented as a nonlinear combination of kernels. Such a combination of kernels results in an indefinite symmetric matrix and hence makes use of the concepts of Krein Space for finding the optimal functions.

Finally, we formulated the MKL using composite kernel functions (MKLCKF). In this MKLCKF the optimal kernel is represented as a linear combination of composite kernel

functions. Corresponding to each data point a composite kernel function is designed whose domain is constructed as the direct product of the range space of base kernels. In this way, the composite kernels make use of the information of all the base kernels for finding their image. Thus MKLCKF has three layers in which the first layer consists of base kernels, the second layer consists of composite kernels and the third layer is the optimal kernel which is a linear combination of the composite kernels. For making the algorithm more computationally effective, we formulated one more variation of the algorithm in which the coefficients of the linear combination are replaced with a similarity function that captures the local properties of the input data. With the aid of data compression techniques, the models have been applied on large data. In the case of large scale classification and regression, dictionary learning and pre-clustering approaches have been used respectively.

The efficiency of all the developed approaches was verified by applying them on real world problems and the results were found to be promising. The comparative study of all the models we developed had also been conducted.