

Supervised Clustering via an Implicit Network for High Dimensional Data

Anand Vidyashankar

*George Mason University
E-mail: avidyash@gmu.edu*

Abstract: In high dimensional data analysis, where the number of parameters exceeds the sample size, it is critical to identify features that are significantly associated with the response variable. Also, it is important to detect groups of features, referred to as clusters or hubs, which have similar effects on the response variable. This allows one to provide summarized information about the relationship between the clusters and the response variable. In this presentation, we introduce a new network-based approach for a high dimensional data analysis that addresses these issues. Specifically, we describe a method for constructing an implicit network and describe a new supervised clustering algorithm based on the network-wide metrics. We study the properties of the network-wide metrics and establish theoretical guarantees for the consistency of the supervised clustering algorithm in a high dimensional setting.