Conditional Sparse Coding and Multiple Regression for Grouped Data

Abstract:
In multiple regression, we predict a response vector $Y$ from a feature vector $X$ with the model $Y = BX$ where $B$ is a regression parameter matrix. For many applications, the data are naturally grouped and it is unsuitable to assume the same model for each group. For example, when we gather climate data from different geographical regions, we do not expect that all the regions to have the same model of climate changes; when we gather fMRI brain imaging data from different human subjects, we do not believe that all the subjects to have the same model of neural activities; when we gather vector time series data from a long period of time, we do not think that all the time periods to have the same model of autoregressive time evolution.

Yet in all these applications, the groups are still related and these relations can be exploited to improve statistical efficiency. We propose an approach in which a dictionary of low rank parameter matrices is estimated across groups, and a sparse linear combination of the dictionary elements is estimated to form a model within each group. We refer to the method as conditional sparse coding since it is a coding procedure for the response vectors $Y$ conditioned on the feature vectors $X$. This approach captures the shared information across the groups while adapting to the structure within each group. It exploits the same intuition behind sparse coding that has been successfully developed in computer vision and computational neuroscience. We propose an efficient algorithm for conditional sparse coding and analyze its theoretical properties in terms of predictive accuracy. We compare the new technique with reduced rank regression for simulation experiments as well as experiments with equities time series data, fMRI neural activity data, and climate data.

DAP Committee: John Lafferty, Larry Wasserman and Geoff Gordon

Speaker: Min Xu

Date: 5/7/12
Time: 2:00pm
Location: 8102 GHC