A modern look at classical multivariate techniques

Yoonkung Lee
Ohio State University

Technological innovations have made high dimensional complex data widely available. This change in the scale and complexity of data has brought new statistical challenges to modeling, prediction and inference procedures. In this course, we look at some of the standard multivariate techniques for regression, classification and dimensionality reduction, and examine their implicit assumptions on the data in terms of size, type, or regularity. We discuss necessary modifications of the techniques when the data does not conform to those conditions largely due to high dimensionality. Control of model complexity or capacity becomes among the key issues to address when modifying the classical multivariate techniques for high dimensional data. Regularization is generally needed for proper control of complexity. Various forms of regularization can naturally induce such desirable properties of the solutions as smoothness, sparsity, or a low rank structure as well. This perspective of regularization will be illustrated with examples of modern multivariate techniques including lasso, smoothing splines, penalized discriminant analysis, and support vector machines.

Another aspect of classical multivariate techniques that this course will examine is how the implicit normality assumption for data has shaped those techniques and how they can be extended to other types of data. Taking principal component analysis (PCA) as an example, we will present its extension to binary data, which is referred to as logistic PCA, and discuss further generalization of PCA to exponential family data.