Learning Ordinality in High-Dimensional Data

Jeongyoun Ahn
Department of Industrial and Systems Engineering, KAIST

Numerous real-world applications involve naturally ordinal outcomes, such as cancer stages or tumor grades. Despite the recent surge in high-dimensional statistical methodologies, high-dimensional learning with ordinal outcomes has been largely overlooked in the HDLSS literature. In this talk, I will introduce recent projects on ordinality in high-dimensional data. The first three topics concern supervised learning aimed at predicting ordinal labels. All three ordinal methods assume sparsity and equal covariance population structure, leading us to term them 'ordinal sparse high-dimensional LDA'. They operate on the principle that a classification rule primarily reliant on variables that are monotonically associated with the outcome should be preferable. They all result in a low-dimensional discriminant subspace where classes are sequentially aligned. The first FWOC method weights features based on their rank correlations with class labels, integrating these weights into the LDA framework. The second SOBL method combines sparsity and ordinality regularizations in a high-dimensional generalized eigenvalue problem. The third SODA approach applied regularization to optimal scores within the sparse LDA framework. In addition, in scenarios where an ordinal outcome is unobserved, one may search for an ordinal signal in the data. This leads us to develop 'monotone clustering' that is designed to identify subgroups interpretable in an ordinal manner.