

GSD2016

Low Rank Structure in Highly Multivariate Models

Iain Johnstone, Stanford University

We start with an overview of some high-dimensional phenomena seen in principal components analysis. More generally, back in 1964 Alan James gave a remarkable classification of many of the eigenvalue distribution problems of multivariate statistics, including PCA. We show how the classification readily adapts to contemporary ‘spiked models’ – high dimensional data with low rank structure. In particular we approximate likelihood ratios when the number of variables grows proportionately with sample size or degrees of freedom. High dimensions bring phase transition phenomena, with quite different likelihood ratio behavior for small and large spike strengths. James’ framework allows a unified approach to problems such as signal detection, matrix denoising, regression and canonical correlations.