

GSD2016

Outlier Pursuit: Robust PCA and Collaborative Filtering

Huan Xu, Georgia Institute of Technology

Principal Component Analysis is one of the most widely used techniques for dimensionality reduction. Nevertheless, it is plagued by sensitivity to outliers; finding robust analogs is critical. In the standard form, PCA involves organizing data into a matrix where columns represent the points, and rows the features. As such, outliers can be modeled as some columns that are completely arbitrary. We propose Outlier Pursuit - a recovery method based on convex optimization, and provide analytical guarantees on when it is able to both (a) recover the low-rank matrix, and (b) identify the outliers. Similar results can be obtained in the more challenging case where on top of having outliers, most entries of the matrix are missing. This can be used in the task of collaborative filtering where some *users* are malicious.