Low rank estimation with MNAR data

Aude Sportisse, Claire Boyer
LPSM, Sorbonne Université

Julie Josse
CMAP, Ecole Polytechnique

Abstract

Rubin introduced three types of missing-data mechanism: missing at random (MAR), missing completely at random (MCAR) and missing not at random (MNAR). In the MCAR and MAR cases, we may ignore the underlying mechanism. In the MNAR case, we have to take the missing-data mechanism into account in the inference.

We handle the MNAR data in the following. We consider the low rank model and thus focus on the problem of estimating a parameter matrix supposed to be low rank. We compare three methods to estimate it. First, we ignore the MNAR mechanism and apply classical methods based on Singular Value Decomposition or Principal Component Analysis and threshold. The second method is to concatenate the data matrix and the missing data indicator matrix and to use classical methods, without modeling the missing-data mechanism. Then, we propose a version of the Expectation algorithm by modeling the missing-data mechanism.