
CLAIRE BOYER, LPSM, Sorbonne Université Paris

Some ways of missing data handling in machine learning

One of the big ironies of data sciences is that the more data we have, the more missing data are likely to appear. After discussing the various issues presented by the missing data in daily-life machine learning, we will present different ways to tackle them for different purposes: (i) one may want to infer the missing values, this is what is called imputation. Imputation can be performed with low-rank techniques, but with optimal transport as well; (ii) one may want to handle missing values in regression, both for performing model estimation or for predictive concern; (iii) one may want to cope with missing values in a non-supervised learning scenario such as for clustering data. We will present some insights and works trying to address the previous issues.