

High-dimensional statistical learning: Roots, justifications, and potential machineries

Amin Zollanvari

- School of Engineering

Abstract

High-dimensional data generally refer to data in which the number of variables is larger than the sample size. Analyzing such datasets poses great challenges for classical statistical learning because the finite-sample performance of methods developed within classical statistical learning does not live up to classical asymptotic premises in which the sample size unboundedly grows for a fixed dimensionality of observations. Much work has been done in developing mathematical–statistical techniques for analyzing high-dimensional data. Despite remarkable progress in this field, many practitioners still utilize classical methods for analyzing such datasets. This state of affairs can be attributed, in part, to a lack of knowledge and, in part, to the ready-to-use computational and statistical software packages that are well developed for classical techniques. Moreover, many scientists working in a specific field of high-dimensional statistical learning are either not aware of other existing machineries in the field or are not willing to try them out. The primary goal in this work is to bring together various machineries of high-dimensional analysis, give an overview of the important results, and present the operating conditions upon which they are grounded. When appropriate, readers are referred to relevant review articles for more information on a specific subject.

| | |
|-------------------|---------------------------|
| Original language | English |
| Pages (from-to) | 109-121 |
| Number of pages | 13 |
| Journal | <u>Cancer Informatics</u> |
| Volume | 15 |
| State | Published - Apr 12 2016 |

Zollanvari, A. (2016). *High-dimensional statistical learning: Roots, justifications, and potential machineries*. *Cancer Informatics*, 15, 109-121. DOI: 10.4137/CIN.S30804