

Cannings, Timothy I.; Samworth, Richard J.

Random-projection ensemble classification. (With discussion). (English) Zbl 1373.62301
J. R. Stat. Soc., Ser. B, Stat. Methodol. 79, No. 4, 959-1035 (2017).

Summary: We introduce a very general method for high dimensional classification, based on careful combination of the results of applying an arbitrary base classifier to random projections of the feature vectors into a lower dimensional space. In one special case that we study in detail, the random projections are divided into disjoint groups, and within each group we select the projection yielding the smallest estimate of the test error. Our random-projection ensemble classifier then aggregates the results of applying the base classifier on the selected projections, with a data-driven voting threshold to determine the final assignment. Our theoretical results elucidate the effect on performance of increasing the number of projections. Moreover, under a boundary condition that is implied by the sufficient dimension reduction assumption, we show that the test excess risk of the random-projection ensemble classifier can be controlled by terms that do not depend on the original data dimension and a term that becomes negligible as the number of projections increases. The classifier is also compared empirically with several other popular high dimensional classifiers via an extensive simulation study, which reveals its excellent finite sample performance.

MSC:

62H30 Classification and discrimination; cluster analysis (statistical aspects)
62P10 Applications of statistics to biology and medical sciences; meta analysis

Cited in **1** Review
Cited in **19** Documents

Keywords:

aggregation; classification; high-dimensional classification; random projection

Full Text: [DOI](#) [arXiv](#)