

Machine Learning II: Assignments #1
14 performance points (max),
email: PDF+code to jan.nagler@gmail.com
due: Wed, April 10, 2024 A.C.

1. Form your group and let me know.
Submission required here.
2. Python-Self-Learning
Go through the Part-1-SelfLearning-Notebook. No submission required here.
3. Random projections

Based on the program developed in the lecture (SparseRandomProjections), analyze 2 databases of your choice (but not exactly the same digits data as in the lecture) using random projections.

Study the accuracy (or a score of your choice that makes most sense for your data) as a function of the number of dimensions / features that survived the random projection.

Try to avoid a zick-zack curve below or around the baseline accuracy curve as your final result for both datasets. At least for one dataset the score is expected to be a smooth-ish curve as a function of the kept number of features. Provide a take-home statement and explain every step.

You may find that data embedded in Euclidian spaces (such as digits) may be more appropriate than data for which Euclidian distances are not excellent distance measures.

Submission required here until deadline.