

Machine Learning II: Assignment #3  
14 performance points (max),  
email: PDF+code to jan.nagler@gmail.com  
due: Wed, May 5, 2021

1. Comments

Please submit only a zip file (no tar please) to my email that unzipped creates a folder called A3GroupX, where X is your group number. The folder should contain the PDF and the code (and dataset, aux files, if necessary).

2. 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 will find that data that is embedded in Eukledian spaces (such as digits) may be more appropriate than data for which Eukledian distances are not an excellent distance measure.