DATA 604, Winter, 2021, Final Project

Due on Feb 16th, 2020, at 11:59PM.

For the final project please provide a comparative analysis of the usefulness of using representation methods for classification, based on the example of the USPS handwritten digits data set. The representation methods must include at a minimum: raw data, PCA (Principal Component Analysis), kPCA (kernel PCA) and LE (Laplacian Eigenmaps). Classification must include at least one of the nonlinear methods presented in this class, like kNN or kSVM. The report (no longer than 10 pages) must be well composed, containing your opinion, supported by the numerical results also presented in the document. The report must include a detailed description of the comparison method you use, description of the experiments performed, their results, and conclusions.

- Proper analysis of each of the 4 methods (raw data, PCSA, kPCA, LE), using a nonlinear classifier for training, validation, and testing, with respect to at least two different measures of success, is worth up to 15 points.
- Comparative analysis and proper conclusions are worth up to 20 points.
- The professionally looking writeup, well written and properly structured, with all necessary supporting evidence, is worth up to 20 points. (Please no longer mixing Matlab printouts with text.)