Tutorial 4 Ensemble learning

1. Please, prepare a short .ppt presentation about ensemble learning methods such as: bagging, random subspace, AdaBoost, RandomForest. Up to 10 slides. Please, base on my lecture, and also on the blog: https://towardsdatascience.com/ensemble-methods-in-machine-learning-what-are-they-and-why-use-them-68ec3f9fef5f

2. Transform using PCA MNIST and FMNIST files to 30-D space. Choose from the MNIST set two the most overlapping classes (e.g. from MNISTA 3 and 8; or 5.3; 4.9; 9.7). These will be the data sets for the next exercises.

3. Show how classic "bagging" works (choose one simple linear classifier) for e.g. MNIST (3.8) for a set of 5 simple classifiers. Students has to answer:

a. how the accuracy of classification and F1 changes for the increasing number of classifiers in the ensemble and the number of examples on which simple classifiers are trained (create a Table of results).

b. How does the classification accuracy change for different methods of decisin fusion (normal average, majority voting, etc.)

4. The same as in 3, but use "random subspace" method (create Table --> increasing number of classifiers, increasing numbers and characteristics).

5. Same as 3 and 4 but SIMULTANEOUSLY for "bagging" and "random subspace" (create respective Table). Choose the optimal number of classifiers on the basis of previous results.

6. Same as 3 and 4 but for the AdaBoost and RandomForest algorithms.

7. Individual PROJECT:

a. Do the same as in points 3-6 but for different pairs of sets in FMNIST. Please, redistribute "difficult" pairs of FMNISTA classes among students.