Assignment 3

Boosting

In this assignment, we experiment with **boosting** as a method for improving the prediction accuracy of learning algorithms.

1. Use scikit-learn but design your own implementation of the AdaBoost algorithm.
   As a bonus, compare this with the implementation of AdaBoost in scikit-learn.
2. Test the boosting algorithm on the MNIST digit database.
   Convert the multiclass dataset to a two-class dataset (for example: even versus odd, prime versus non-prime, etc).
3. Choose at least two "weak" learners for your experiments.
   For one of the weak learners, use the decision tree classifier which is supported by scikit-learn.
   Then, choose another weak classifier of your choice (within scikit-learn) and compare this with the decision tree weak classifier.
4. Provide a plot showing how the training error of the AdaBoost classifier changes during the boosting process.
   Determine whether boosting works effectively and which weak classifier is better.

What to submit

Submit via email your source file(s) and the plot (as PDF). Provide a brief demo of your program. You may work in groups of three.

Bonus

Repeat the above using a different real-world dataset (other than the MNIST digit dataset) of your choice.