

PROJECT2: ADA-BOOST

Vaibhav Choudhary

IMPLEMENTATION

1. Adaboost algorithm allows us to combine weak classifiers to give a strong classifier. It is based on combining one-level decision trees called stumps and allocating them to weight alpha, to produce a model.
2. In the code, the module AdaBoost implements this technique. The class train goes through each haar feature at each of the coordinates allowing us to find the stump that gives the min weighted error at each step.
3. At first, the code was written to try each value of the stump as a threshold. But this made the algorithm scale in n^3 for each stump making it computationally very expensive. Hence, the threshold was decided to be the mean for each of the features.
4. The CBCL dataset was chosen for training and testing and an average accuracy of 89 percent was observed.
5. Another function, harr_without_ada returns the top 10 features with the least error along with their coordinates before without using Adaboost.
6. The ensemble model obtained is saved using joblib and submitted as adaboost_model.

Results

The ROC curve obtained is as follows:

