Implementation of ensemble learning algorithms

Davide Vettore 868855
Alessio De Luca 919790

Objective:

The objective of this activity is to train and compare the performances of multiple ensemble learning algorithms.

Introduction:

The experiment is conducted on a single dataset, the scatterplot of the training set can be seen in the following figure.

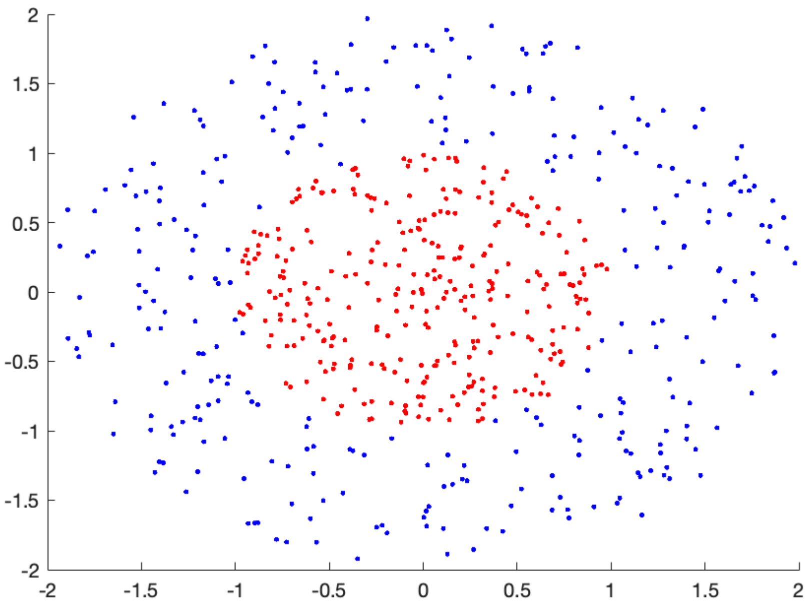

Figure 1: Scatterplot of the training set

As previously mentioned, our methodology incorporates multiple classifiers, as well as multiple meta-classifiers. It's noteworthy that these models are trained over different partitions at different stages. In both approaches, the same five level-1 classifiers are employed:

1. **Linear Support Vector Machine** with kernel scale set to 5.
2. **Polynomial Support Vector Machine** with kernel scale set to 10.
3. **Decision Tree** with maximum number of splits set to 20.
4. **Naïve Bayes** with default hyperparameters.
5. **Ensemble of Decision Trees** with LogitBoost as ensemble aggregation algorithm, one hundred trees (with 10 maximum splits each) compose the ensemble.

Similarly, the **meta-classifiers** are trained with the following hyperparameters:
- Bootstrap aggregation (bagging) as the ensemble aggregation algorithm.
- The number of ensemble learning cycles is set to 200, deviating from default value of 100.

Additionally, a version incorporating optimized hyperparameters was explored; however, this falls outside the scope of the experiment.

Procedure:

Initially, the training set is partitioned into two stratified folds. Five level-1 classifiers are trained on the first fold, and subsequently, their predictions and scores on the second fold are utilized for training respectevly two meta-classifiers. The performances obtained by the meta-classifiers are assessed on a test set.

Following this, an alternative approach is employed: all five level-1 classifiers and the two meta-classifiers are trained on the entirety of the available training set, without any partitioning. We seek to evaluate the effectiveness of the meta-classifiers by comparing their performances under these two distinct approaches.

Results:

The performances obtained in the implementation of the two different approaches can be seen in the following tables:

	Gaussian SVM	Polyno-mial SVM	Decision Tree	Naïve Bayes	Ensemble of Decision Trees	Meta-classifier trained on Scores	Meta-classifier trained on Predictions
Accuracy	0.8683	0.6250	0.9483	0.9783	0.9533	0.9933	0.9700

Table 1: Level-1 classifiers trained on fold-1, Meta-classifiers trained on fold-2

	Gaussian SVM	Polyno-mial SVM	Decision Tree	Naïve Bayes	Ensemble of Decision Trees	Meta-classifier trained on Scores	Meta-classifier trained on Predictions
Accuracy	0.9000	0.6333	0.9667	0.9917	0.9683	0.9700	0.9683

Table 2: Level-1 classifiers and Meta-classifiers are trained on the entire train set

Conclusion:

Notably, the accuracy of each level-1 classifier increases in the second approach. This outcome was expected, as the five models are trained on the entire training set rather than a single fold comprising only half of the observations. Despite the individual classifiers performing better in the second approach, both meta-classifiers achieve higher accuracy in the first approach. This suggests that partitioning the training set into two distinct folds for training the weak classifiers and the meta-classifiers leads to a significant improvement in classification performance.

Another insightful observation is that meta-classifiers perform notably better when trained on the scores rather than predictions. This phenomenon is consistent across both approaches.