STACKING CLASSIFIERS - ASSIGNMENT 2

Mirko Morello
920601
m.morello11@campus.unimib.it

Andrea Borghesi
916202
a.borghesi1@campus.unimib.it

January 3, 2025

1 INTRODUCTION

In this task, our objective was to test the performance of an ensemble classifier. The tests that we are asked to perform regard a couple of rules that we should respect when training an ensemble of classifiers. We will test how the performance changes when we do not take into account these rules.

2 CLASSIFIERS & ENSEMBLE

2.1 Classifiers

The classifiers used, that have later been stacked for the ensemble of the meta-classifiers are the following:

- SVM Gaussian: `KernelScale = 5`
- SVM Polynomial:`KernelScale = 5`, `Degree = 3`
- Classification Tree: `MaxNumSplits = 15` `SplitCriterion = 'gdi'`
- Naïve Bayes
- Ensemble Decision Tree

2.2 Ensemble

The type of ensemble used is a stacked classifier, a type of ensemble that involves combining multiple learners to improve the overall performance compared to using just the models on their own. These learners are also called *base learners* and their prediction, usually made on a section of the training that they've not seen yet, is used to train the next layer's *meta-classifier*. The meta-classifier makes the final prediction based on the predictions made by the base classifiers. We paired this technique with the Bagging algorithm to generate base classifiers and aggregate their output, used as input for the meta-classifier.

3 FIRST TASK: META-CLASSIFIER TRAINED WITH PREDICTIONS

3.1 Description

The first task was to try to disrupt the behaviour of the Meta-classifier by ensembling it with the predictions of the classifiers instead of their scores, the infima of the confidence intervals given by every classifier.
The test set has not been modified, otherwise, we wouldn't be able to understand if there has been any changes performance-wise. We performed a 2-fold over the training set: the first fold was used to train the base classifiers while the second fold was used to test them and therefore train the meta-classifier.

3.2 Results

By following the proper methodology we achieved an accuracy of 0.99. This first experiment led to an accuracy of 0.9817, slightly worse than the first result. An interpretation of this behaviour might regard the nature of the values onto which we are training our classifiers. The CI refers to an interval where the estimated value can be with a high probability, bringing more information than the simple prediction, leading to a higher accuracy of the prediction.

4 SECOND TASK: META-CLASSIFIER TRAINED ON SAME TRAINING DATA AS THE BASE CLASSIFIERS

The second task was also to try to disrupt the behaviour of the Meta-classifier by training it on the same training set as the base classifiers. That has been accomplished by not performing the 2-fold on the training set.

4.1 Results

As expected, training and testing the base learners using the same dataset leads to a biased model and, if we were to test their accuracy at this stage, also to an overly optimistic performance estimate. This leads to a suboptimal performance as **table 1** shows.

5 Results

All the results of the various classifiers and experiments can be seen in the **table 1**, the accuracies of the base learners that are later stacked are referred to the accuracies when trained on the first split of the training dataset, while the accuracies that they have when trained on the whole training dataset (see 4) are not reported.

Classifier	**Accuracy**
SVM Gaussian	0.868333
SVM Polynomial	0.625000
Classification Tree	0.948333
Naive Bayes	0.978333
Ensemble Decision Tree	0.953333
Meta-Classifier	0.990000
Meta-Classifier trained on predictions	0.970000
Meta-Classifier trained on same training data as base learners	0.968333

Table 1: Accuracy of Different Classifiers

6 Conclusion

As expected, not following the suggested methodology leads to suboptimal results. This tests are not saying that we shouldn't ever go for these procedures, as there are cases where they perform better, for instance, when the dataset is small, using the entire dataset for training both the base learners and the meta-classifier could be necessary due to the limited amount of data available. On the contrary, it is safer to say that the suggested methodologies work best as rules of thumb.