MOAD'2024

The Sixth International Colloquium on Methods and Tools for Decision Support

October 20th–22nd 2024, Tizi-Ouzou, Algeria

A prediction model based on multi-layer stacking Ensemble learning

Abdouche Safia
Mouloud Mammeri University of Tizi-Ouzou, ALGERIA
Laboratoire de Mathématiques pures et appliquées
safia.abdouche@ummto.dz

Abstract

The idea of ensemble learning is to try to reduce the bias and/or variance of the base models in order to reduce the generalization error of the global model. One of these methods is stacking which consists of using a meta-model to combine the predictions of several basic models. In this study we propose a prediction models based on ensemble learning by stacking with multi-layers which we compare to the best-known ML methods on several datasets.

Introduction

In Machine learning, ensemble methods combine multiple models to obtain better predictive performance [1]. Combining multiple models has been theoretically and experimentally shown to provide significantly better performance than their single base learners in a predictive system [2]. There are several techniques to construct an ensemble. They differ mainly in how the individual learners are trained and how their predictions are combined. The most common ensemble learning methods are: bagging, boosting and stacking. In the stacking, multiple base learners (often heterogeneous) are combined using a meta-model to create a more accurate and robust model [3]. The meta-learner, which is trained to predict the target variables, uses the predictions made by the base models, which were trained on the training data as input features. In this paper a set of stacking ensemble models with two and three layers are constructed . These stacking models are compared to the usual machine learning methods, namely: linear regression (LR), regularized regression (Lasso, Ridge and Elastic net), k-nearest neighbor algorithms (KNN), decision trees (DT), support vector machines (SVM), multi-layer perceptron (MLP), and includes representatives of bagging and boosting, random forests (RF) and the extreme gradient boosting algorithm (XGB), and in particular compare them with each other An analysis of the performance of the different stacking models is done.

Methodology

Let $D = \{(x_i, y_i), i = 1, \ldots, N\}$ denote the dataset with N observations with input x_i and output y_i value.

The stacking process is as follows:

- Step 1
  - Given a data set D
  - Given n base learners $\mathcal{L}_t$, $t = 1$ to n
  - Given a meta learner $\mathcal{L}$
- step 2
  - Split the data set into train and test set: $D = D_{train} \cup D_{test}$
  - perform a k-fold cross validation on D_{train}, $D_{train} = D_1 \cup D_2 \cup \ldots \cup D_k$ (the same k-folds must be used for each base learner)
- Step 3
  - Train each base learner $\mathcal{L}_t$ on $D_{-j} = D_{train} - D_j$ to produce the n models h_t
  - For each $x_i \in D_j$ get the prediction: $\hat{y}_{it} = h_t(x_i)$
  - the new data set is generated from the n learners as:

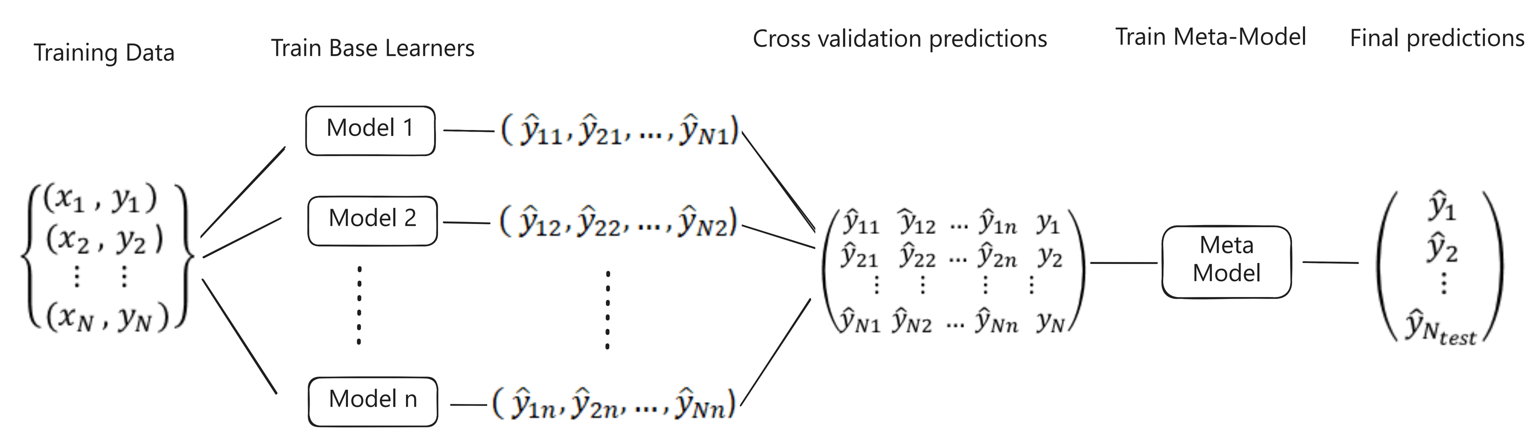
$$D' = \{(h_1(x_i), h_2(x_i), \ldots, h_n(x_i), y_i), i = 1, \ldots, N\}$$

- Step 4
  - Train the meta learner $\mathcal{L}$ on D' to produce the model h
- Step 5
  - Get the final prediction of the stacking ensemble model H on D_{test}

$$H(x_i) = h(h_1(x_i), h_2(x_i), \ldots, h_n(x_i)) \quad , x_i \in D_{test}$$

  - Evaluate the stacked model H on D_{test} by calculating the different errors and the determination coefficient

These different steps are illustrated in the following figure



To evaluate the two proposed three-layer stacking ensemble models and compare them with each other and with the other ML models, we used three metrics: mean absolute error (MAE), root-mean-square error (RMSE), and R^2 score.

Let $D_N = \{(x_i, y_i), i = 1, \ldots, N\}$ denote the dataset with N observations with input x_i and output y_i value. Let $\hat{y}_i = \hat{f}(x_i)$ be the estimation obtained from x_i using the ML model $\hat{f}(\cdot)$:

$$\hat{f} = argmin_{f \in F} R_{D_N}(f)$$

- The R^2 score, also called coefficient of determination is defined as

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}$$

  The R^2 corresponds to the average error of the model divided by the average error of a base model that always predicts the average of the variable to be predicted.

- The RMSE is defined by:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}.$$

- The MAE is the average of the absolute values of the errors, defined by the formula:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|\,.$$

These metrics are calculated on unseen data during training.

Main result

The following table shows the performance measures in terms of MAE, RMSE and R2 score and the complexity (training time) , of the proposed stacking models and the other ML models.

Models	MAE	MAESD	RMSE	RMSESD	R2	R2SD	time
LR	4111	107	5933	204	0.7698	0.035	2.60
Lasso	4104	158	5935	262	0.7600	0.020	3.07
Ridge	4246	175	6022	316	0.7601	0.038	1.55
E.net	4104	184	5935	424	0.7600	0.040	1.40
KNN	3190	191	5427	427	0.7988	0.041	5.54
SVM	2500	129	4774	238	0.8448	0.018	3436
MLP	2774	251	5208	661	0.8149	0.047	53.49
DT	2716	131	4937	205	0.8365	0.021	1.53
RF	2607	111	4716	151	0.8488	0.018	23.92
XGB	2363	134	4391	198	0.8679	0.015	461.56
Stack1	2612	192	4503	333	0.8618	0.028	26.25
Stack2	2720	155	4570	304	0.8573	0.026	21.50
Stack3	2554	155	4509	352	0.8617	0.027	11.33
MLS1	2249	147	4531	631	0.8605	0.035	64.06
MLS2	2318	132	4453	131	0.8650	0.033	69.53

Conclusion

In this paper we propose a set of multi-layer stacking models, the proposed models are compared to the most known ML models, the results showed that the proposed models outperform the considered ML models even other ensemble models like bagging and boosting.
We also compared the proposed models with each other and the results show an enhancement in performance by going from 2 to 3 layers

References

[1] Dong, X., Yu, Z., Cao, W., Shi, Y., & Ma, Q. (2019). A survey on ensemble learning. Frontiers of Computer Science, 14, 241 - 258.

[2] Mienye, I.D., & Sun, Y. (2022). A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects. IEEE Access, 10, 99129-99149.

[3] Pavlyshenko, B.M. (2018). Using Stacking Approaches for Machine Learning Models. 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP), 255-258.