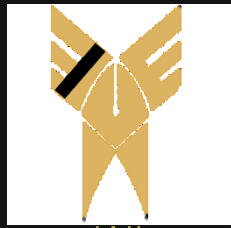


فقط حیدرآباد میں ہے

عاشقِ حیات



I.A.U.
Birjand Branch

Research title of advanced machine learning course

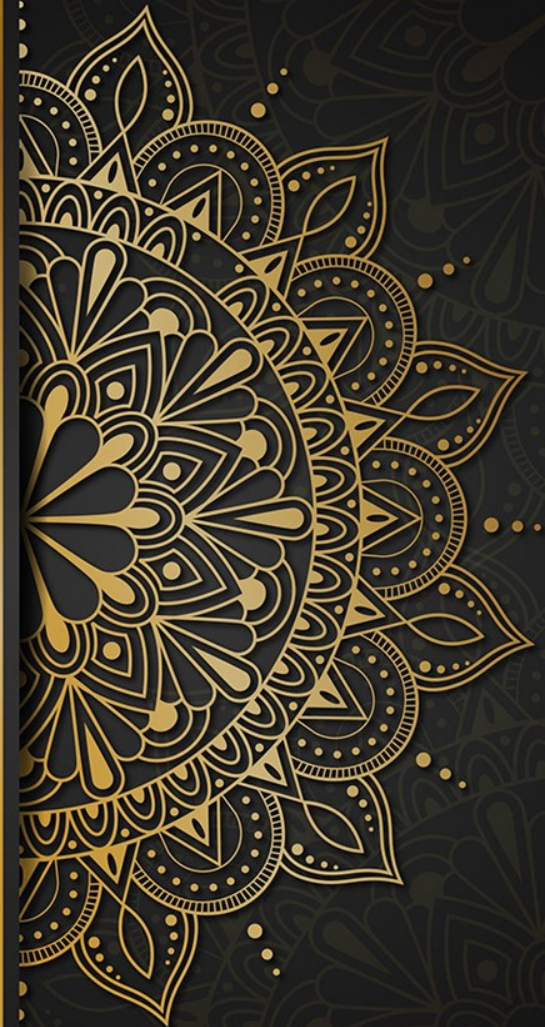
A comprehensive review on ensemble deep learning: Opportunities and challenges

Provider:



Supervisor:

May 2024





A comprehensive review on ensemble deep learning: Opportunities and challenges

Authors: Ammar Mohammed, Rania Kora

Department of Computer Science, Faculty of Graduate Studies
for Statistical Research, Cairo University, Cairo, Egypt

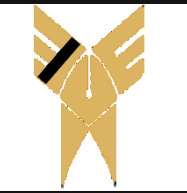
Accepted 19 January 2023

Available online 1 February 2023

DOI Link : <https://doi.org/10.1016/j.jksuci.2023.01.014>



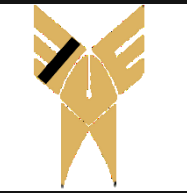
- Introduction
- Search result in Scopus.
- Ensemble Method.
 - Methodology: Voting, Stacking, Bagging, Boosting.
 - Fusion method
- A brief overview of errors in machine learning
- Result & Conclusion



Introduction

In a world full of diverse and varied data sources, machine learning has become one of the most important and dominant branches of artificial intelligence methods, which is applied in many fields. There are many different learning algorithms and methods. Each method's pitfalls and drawbacks are measured in terms of several factors, including performance and scalability. Based on a lot of research in machine learning, two methods dominate learning algorithms: deep learning and ensemble learning.

Deep learning techniques can scale and handle complex problems and offer automatic feature extraction from unstructured data. Additionally, deep learning methods contain several types of network architectures for different tasks, such as feed forward neural networks, convolutional neural networks, and recurrent neural networks. However, the training process of deep learning models requires a massive effort, and tuning the optimal hyper parameters requires expertise and extensive trial, which is a tedious and time-consuming task.



Introduction

On the other hand, ensemble learning refers to a learning methodology that combines several base models to build a larger and more powerful model than its constituents. Additionally, ensemble learning can reduce the risk of over fitting thanks to the diversity of base models. Ensemble learning has been successfully applied in various fields and domains, and it outperforms single models. There are several ensemble learning techniques that differ in how the base models are trained and combined. The most common ensemble techniques include averaging, bagging, stacking, and boosting.

Thus, this paper aims to comprehensively review different strategies for applying deep ensemble learning. It also presents various aspects that influence the success of ensemble methods, including the type of base models used, data sampling techniques in training, the diversity of employing different base classifiers, and methods for fusing deep base models. Additionally, it discusses the advantages and disadvantages of each strategy.

Scopus



Scopus - Trends of Ensemble Learning

Due to the strength and efficiency of the ensemble learning system in improving model predictive performance, ensemble learning has become an important research trend in recent years, leading to an increase in the number of studies utilizing ensemble learning in various application domains. To show the extent to which the number of published articles on ensemble learning has increased each year and the different fields in which ensemble learning has been applied from 2014 to 2021, a search was conducted in this database using the terms "Ensemble Learning" and "Ensemble Deep Learning." These terms were searched in the titles, abstracts, and keywords of the articles.

It appears that ensemble learning and ensemble deep learning are experiencing significant growth, especially in the field of computer science. The presented statistics show that these methods have garnered the attention of researchers due to their ability to enhance predictive performance. The fact that the highest number of articles referencing ensemble learning and ensemble deep learning were published in 2021 indicates that researchers are actively investigating and utilizing these methods across various fields. The high rate of use in computer science suggests that these techniques hold significant importance and practical application in this area.

Scopus

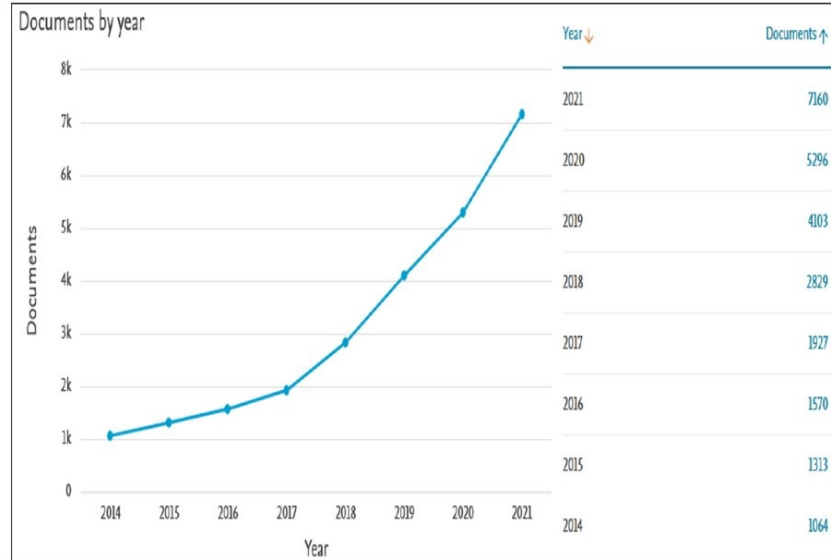
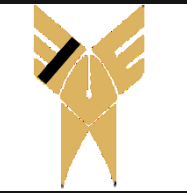


Fig. 1. The trends of search term "Ensemble Learning" in "Scopus" from 2014 to 2021 (Scopus, 2023).

Figure 1 shows the number of articles published for the search term "Ensemble Learning" each year during the mentioned period. This figure indicates that the number of articles found using this term is estimated to be 25,262, demonstrating an increasing trend in ensemble learning over several years.

Scopus

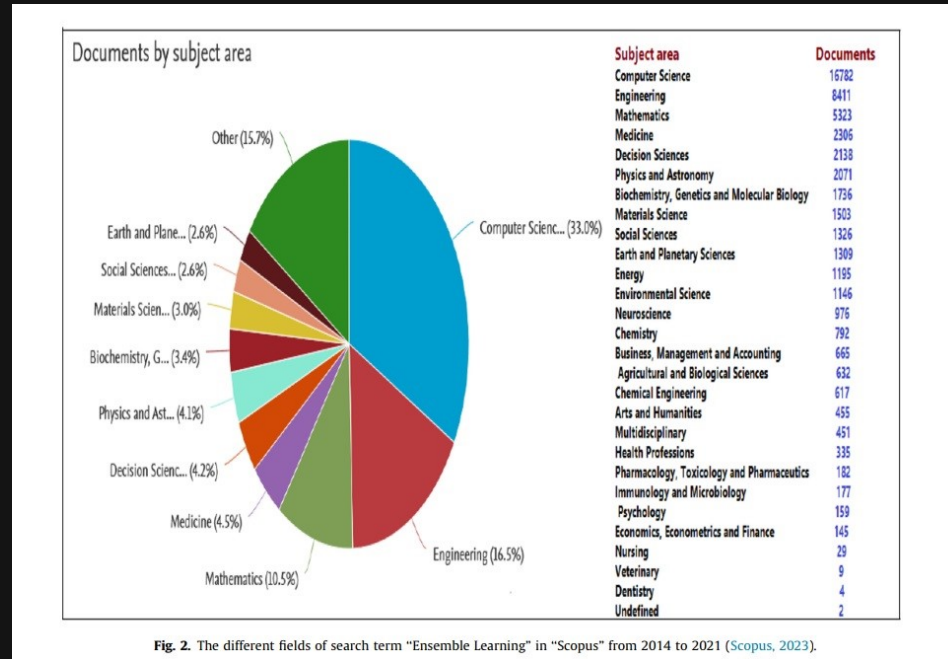


Figure 2 shows the number of articles that addressed the search term "Ensemble Learning" across all fields. From this figure, it can be observed that the field of computer science has the highest number of articles, estimated at approximately 16,782 documents.

Scopus

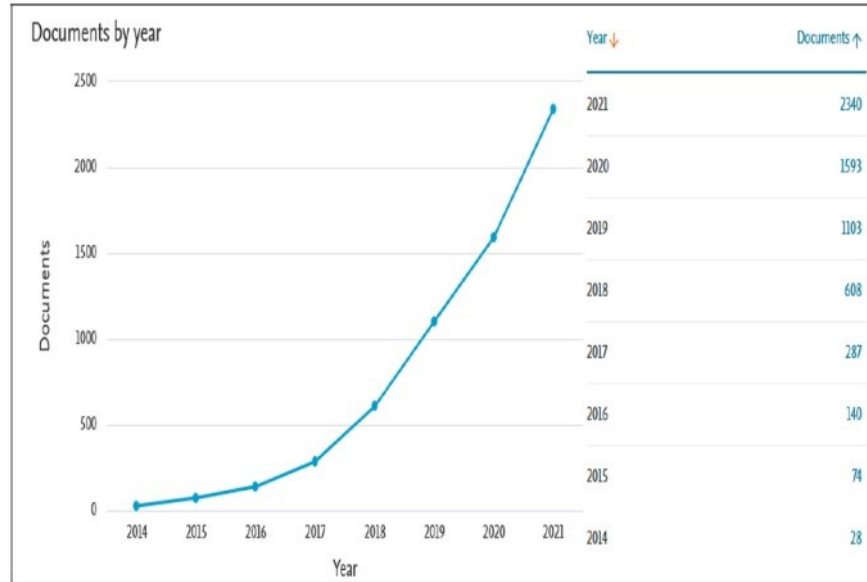


Fig. 3. The trends of search term "Ensemble Deep Learning" in "Scopus" from 2014 to 2021 (Scopus, 2023).

Figure 3 shows the number of articles published for the search term "Ensemble Deep Learning" each year during the mentioned period. This figure indicates that the number of articles found using this term is estimated to be 6,173, demonstrating the increasing interest of researchers in this trend.

Scopus

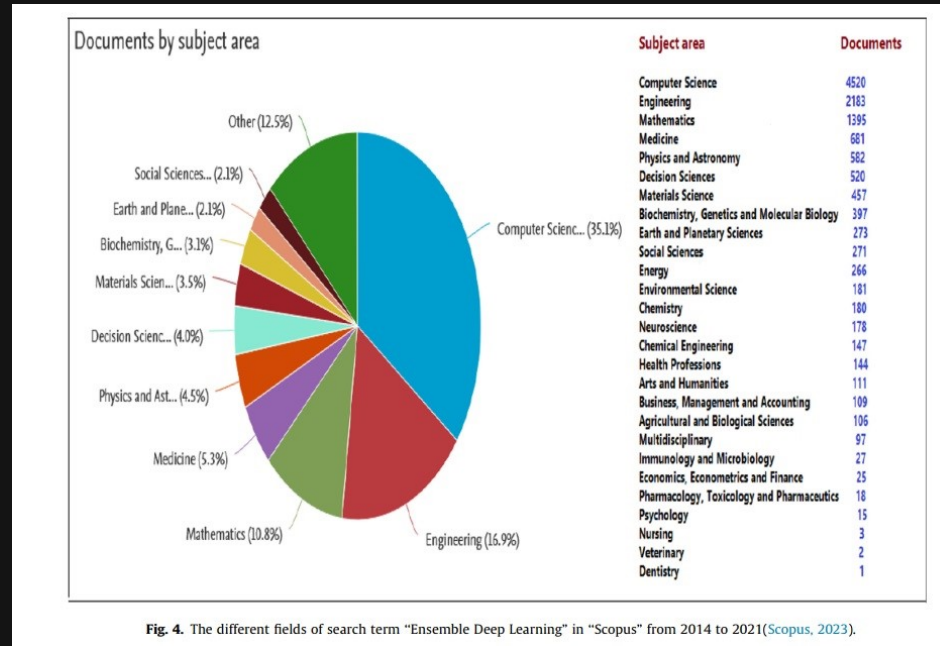
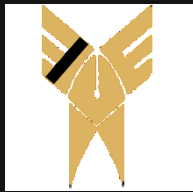


Figure 4 shows the number of articles that addressed the search term "Ensemble Deep Learning" across all fields. From this figure, it can be observed that the field of computer science has the highest number of articles.

Ensemble Learning Method



Ensemble learning is another method in machine learning that improves the final accuracy and performance of the model by considering multiple models. The goal of this machine learning method is to reduce errors or biases that may exist in each of the individual machine learning models. In this domain, multiple models are used in combination to enhance the model's ability to predict output data.

A brief overview of errors in machine learning

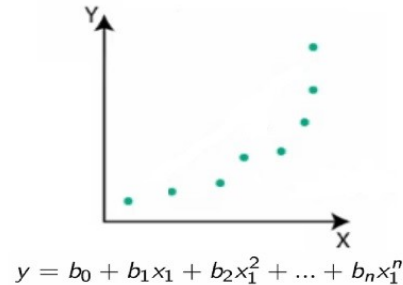
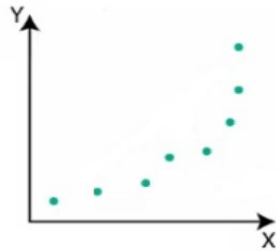
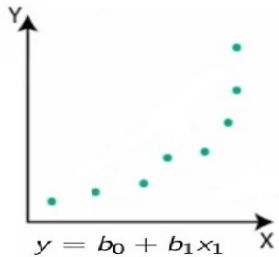


Simple
Linear
Regression

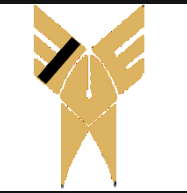
$$y = b_0 + b_1x_1$$

Polynomial
Linear
Regression

$$y = b_0 + b_1x_1 + b_2x_1^2 + \dots + b_nx_1^n$$



Important Points in Collaborative Learning



Data Sampling:

The selection of a data sampling method is one of the most critical factors influencing the performance of a collaborative system. There are two methods of data sampling in collaborative learning.

1- Independent Dataset Strategy:

In this strategy, subsets are not dependent on each other. This means that the performance of one subset does not affect the performance of other subsets.

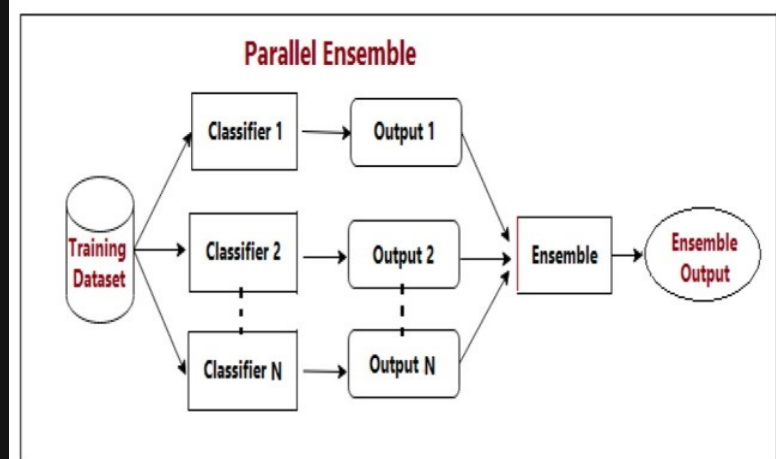
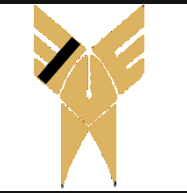


Fig. 8. General framework of parallel ensemble.

Important Points in Collaborative Learning



2- Dependent Dataset Strategy:

In contrast, in this strategy, subsets are dependent on each other. This means that the performance of one subset is influenced by the results of previous subsets.

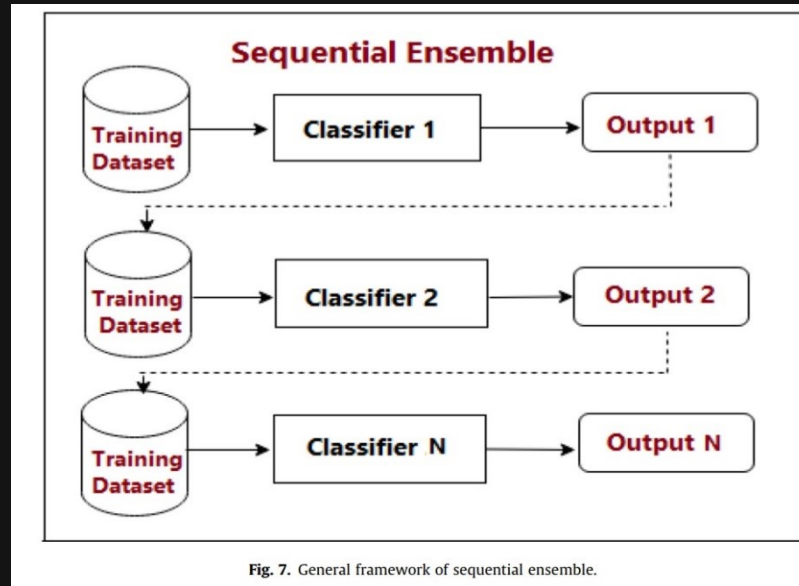
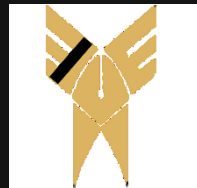


Fig. 7. General framework of sequential ensemble.

Training baseline classifiers



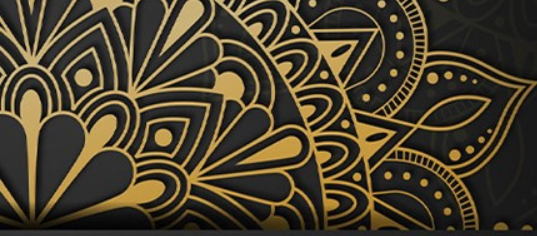
The diversity of baseline classifiers is the second influential factor in a group system. In the core of each group-based system, there are two techniques for training individual group members: sequential group technique and parallel group technique.

- ❖ Sequential Group Technique:

In this technique, different learners are trained sequentially because of data dependency. Therefore, the errors of the first model are corrected sequentially by the second model, as shown in Figure 7. The main advantage of sequential methods is that they exploit the dependency between baseline learners.

- ❖ Parallel Group Technique:

In this technique, baseline learners are created simultaneously as there is no data dependency. Therefore, each data point in the baseline learner is generated independently, as shown in Figure 8. The main advantage of this technique is exploiting independence among baseline learners. Thus, errors generated by one model differ from errors present in another independent model, allowing the group model to compute the average errors.

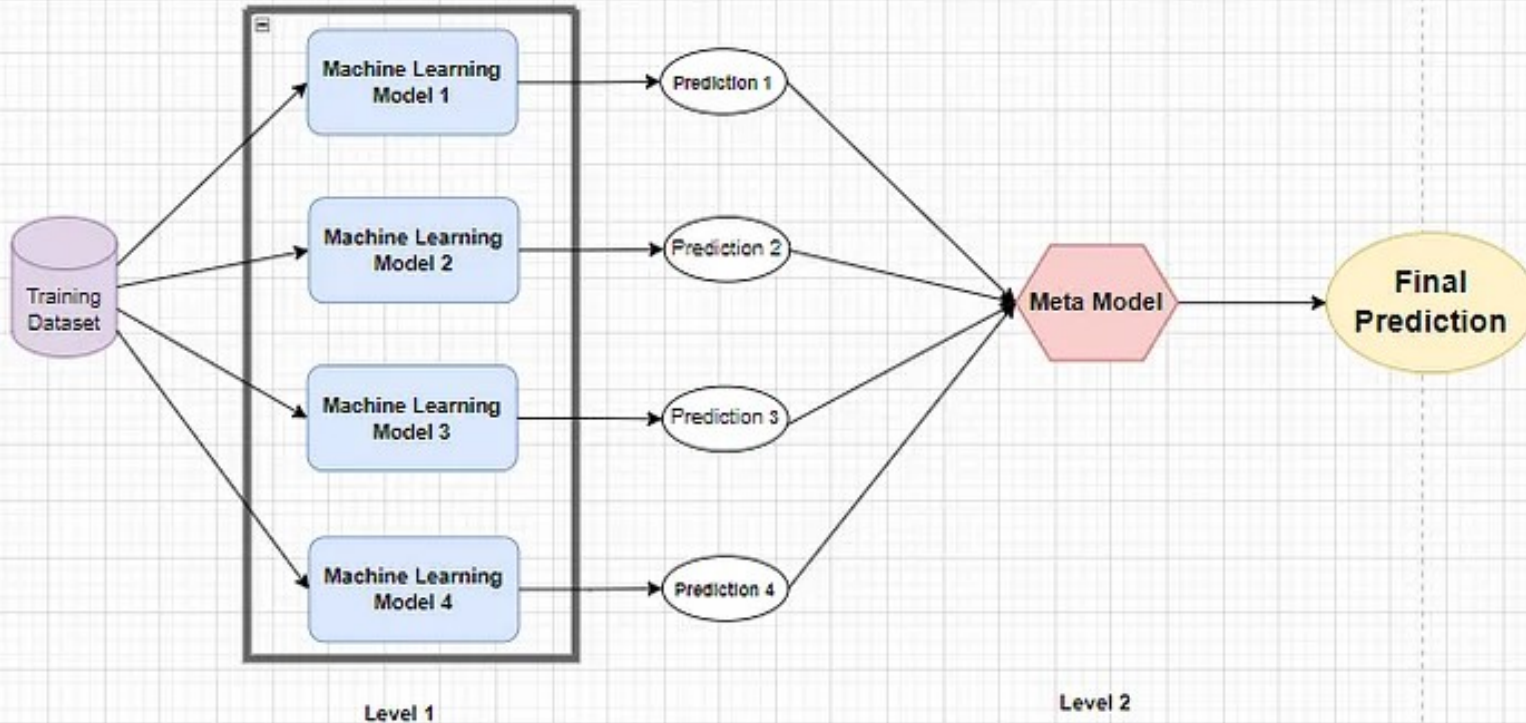
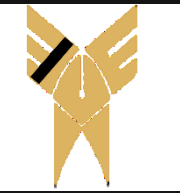


Ensemble Learning Method

روش Method	وابستگی Dependent	روش تجميع Fusion method	همگنی Heterogeneity
Stacking	موازی Parallel	فراگیری متا Meta Learning	ناهمگن Heterogeneous
Bagging	موازی Parallel	وزن دهی Weight Voting	همگن Homogenous
Boosting	ترتیبی Sequential	وزن دهی Weight Voting	همگن Homogenous

- Table 1: Classification of Group Methods Investigated in the Article

Machine Learning using Stacking





Advantages of Stacking



Improved Performance:

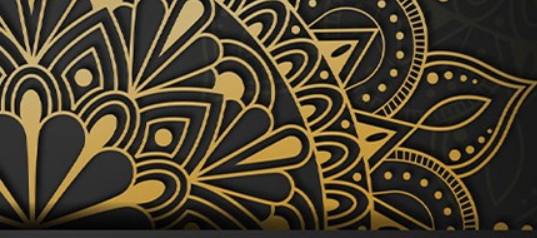
1- By combining multiple models, stacking often leads to better predictive performance than any single model.

Flexibility:

2- Allows the use of diverse base models and a flexible choice of meta-models.

Reduction of Over fitting:

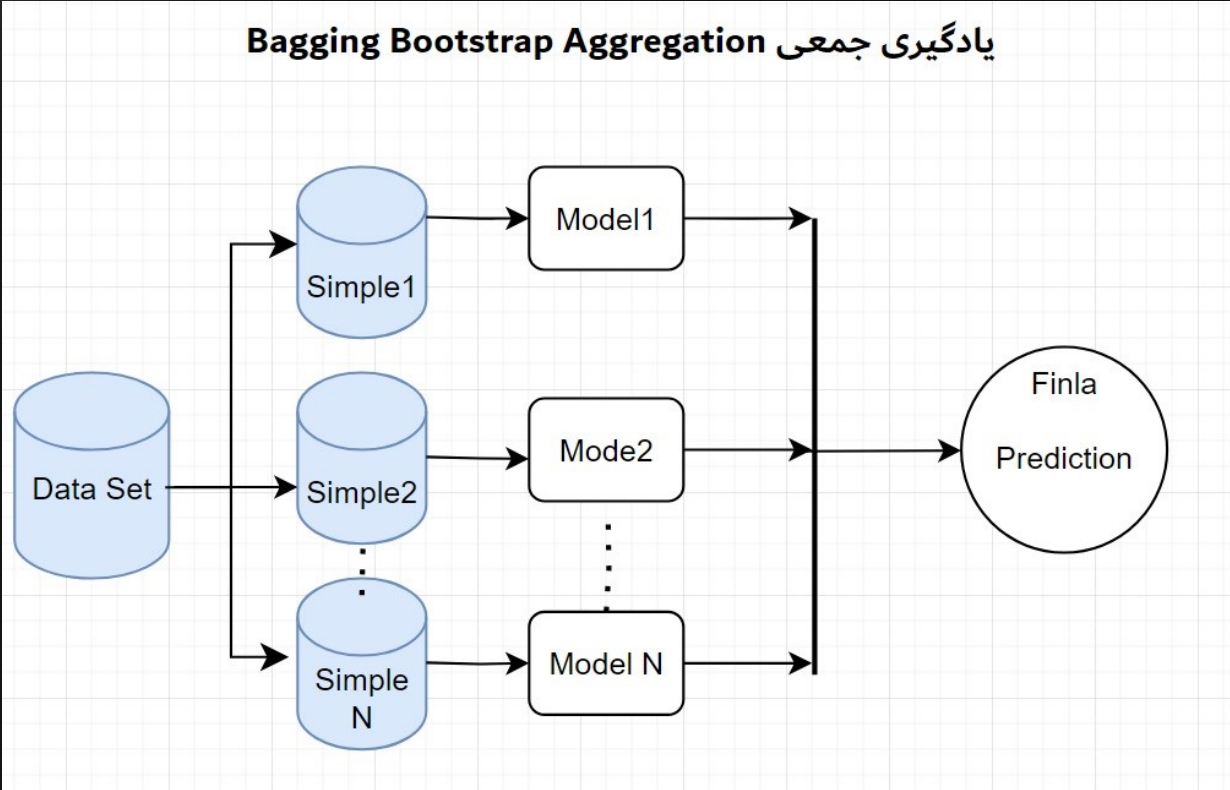
3- Helps in reducing over fitting by combining models that may individually over fit.



Machine Learning using Bagging



Bagging Bootstrap Aggregation یادگیری جمعی





Advantages of Bagging



Reduced Over fitting:

1-By training multiple models on different subsets of data, bagging reduces the risk of over fitting that a single model might have on the original training data.

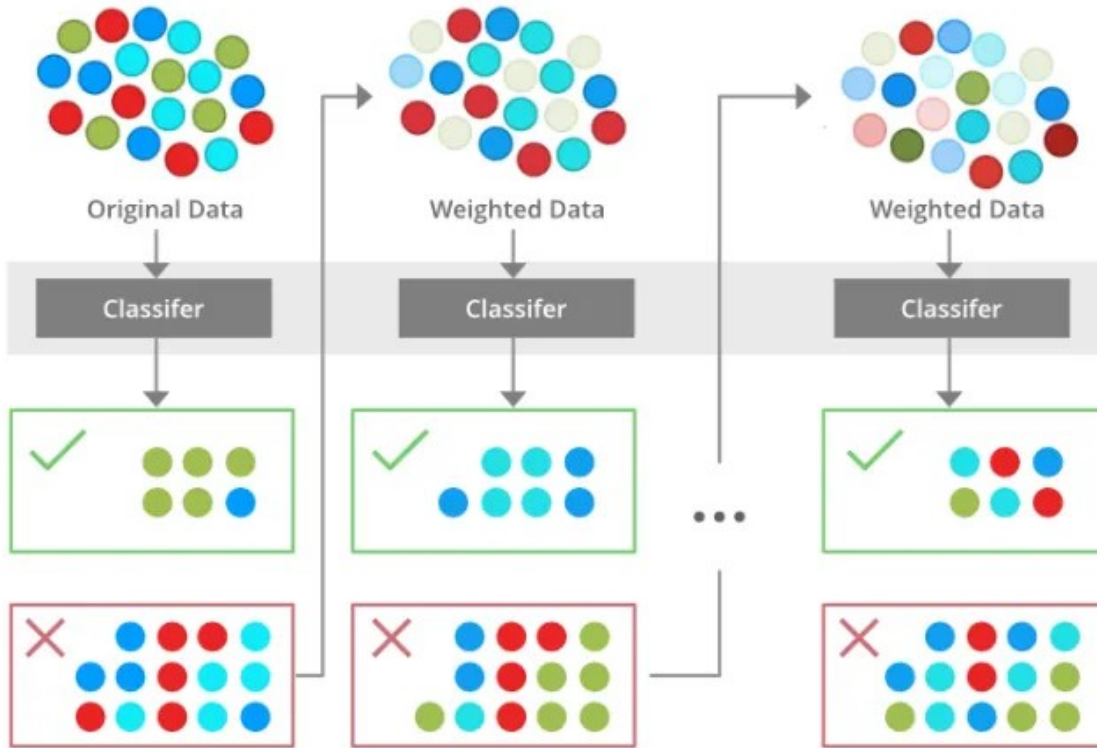
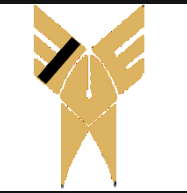
Improved Accuracy:

2-Aggregating predictions from multiple models often results in better performance compared to a single model, as it leverages the strengths of multiple hypotheses.

Model Stability:

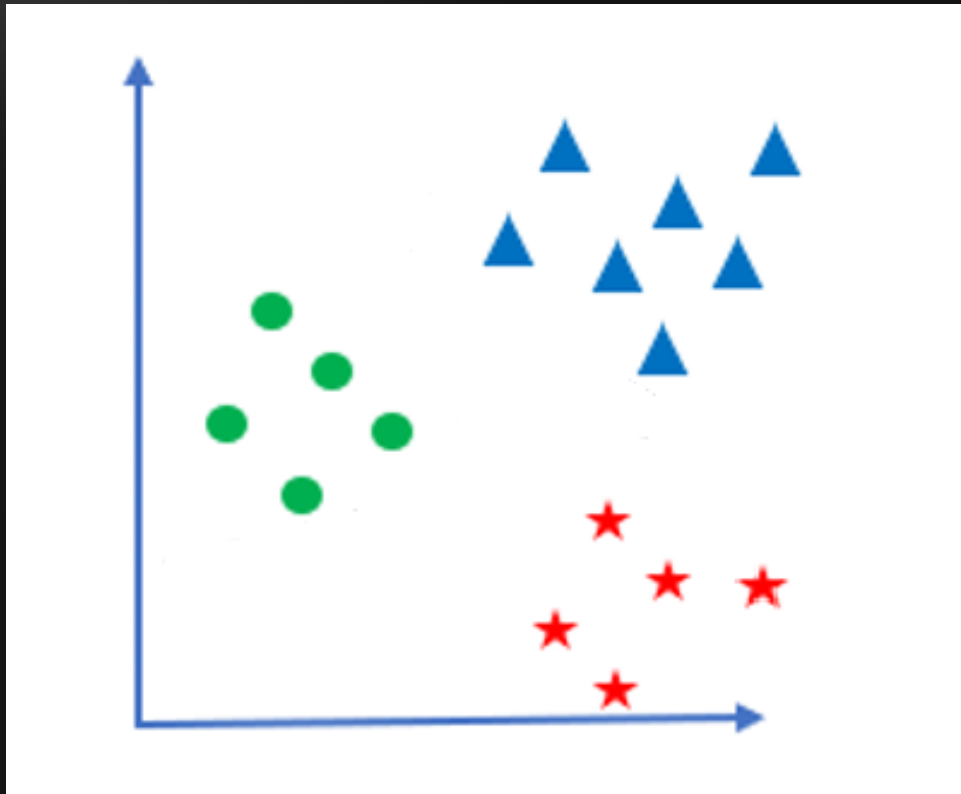
3-Bagging increases the robustness and stability of the model by minimizing the impact of noisy data points and outliers.

Machine Learning using Boosting

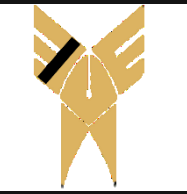




example of Boosting



Advantages of Boosting



Improved Performance:

1- By focusing on the errors of previous models, boosting often significantly improves the overall performance.

Adaptability:

2- Boosting can adapt to the complexity of the data by adjusting the focus on difficult instances.

Reduction of Bias:

3- By iteratively focusing on the errors, boosting reduces the bias of the overall model.



Conclusion

In machine learning, reducing error and variance of models is one of the key factors determining the success of the learning process. Various studies have proven that combining the outputs of different classification algorithms can reduce generalization error without increasing the model's variance. This essential point is the core of a concept known as ensemble learning. Numerous research efforts have preferred ensemble learning over single-model learning in various domains. The main advantage of ensemble learning is combining multiple independent models to improve predictive performance and create a stronger model that outperforms each individual model. In the literature, various techniques exist to enhance classification algorithms. The main difference between these ensemble methods lies in training the base models and how they are combined.

Several research efforts have introduced ensemble learning to deep learning models to address issues that arise during the learning process of deep learning models. The primary challenge for deep learning models is typically the need for substantial knowledge and experience to optimally tune hyperparameters to achieve minimal global error. However, finding the optimal parameters requires a tedious technique in the search space, which in turn becomes a time-consuming and costly task. Therefore, several research efforts have applied ensemble deep learning in many domains, with most of these efforts focusing on simple ensemble methods.

This paper provided a comprehensive review of various ensemble learning strategies, especially concerning deep learning.



End Thank you for your attention

A comprehensive review on ensemble deep
learning: Opportunities and
challenges

Weighted Average in regression



$$\frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i} = \text{Weighted Average}$$

میانگین وزنی در رگرسیون

فرض کنید سه مدل رگرسیون داریم که پیش‌بینی‌های آنها برای یک نمونه داده به ترتیب ۲، ۴ و ۵ است. دقت این مدل‌ها به گونه‌ای است که وزن‌های آنها به ترتیب ۰.۲، ۰.۳ و ۰.۵ است.

برای محاسبه میانگین وزنی این پیش‌بینی‌ها، از فرمول استفاده می‌کنیم:

$$\text{Weighted Average} = \frac{(0.2 \times 2) + (0.3 \times 4) + (0.5 \times 5)}{0.2 + 0.3 + 0.5}$$

ابتدا صورت کسر را محاسبه می‌کنیم:

$$(0.2 \times 2) + (0.3 \times 4) + (0.5 \times 5) = 0.4 + 1.2 + 2.5 = 4.1$$

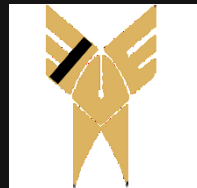
سپس مخرج کسر را محاسبه می‌کنیم:

$$0.2 + 0.3 + 0.5 = 1.0$$

حالا میانگین وزنی را محاسبه می‌کنیم:

$$\text{Weighted Average} = \frac{4.1}{1.04} = 4.1$$

Weighted Average in classification



میانگین وزنی در طبقه‌بندی

روش ساده برای میانگین وزنی در طبقه‌بندی فرض کنید سه مدل داریم که می‌خواهیم پیش‌بینی‌های آنها را برای یک نمونه ترکیب کنیم. برچسب‌های پیش‌بینی‌شده توسط این مدل‌ها و وزن‌های آنها به شرح زیر است:

- مدل ۱: پیش‌بینی (A) با وزن ۰.۲
- مدل ۲: پیش‌بینی (B) با وزن ۰.۳
- مدل ۳: پیش‌بینی (A) با وزن ۰.۵

مجموع وزن‌ها برای کلاس (A)

$$A: 0.2+0.5=0.7$$

مجموع وزن‌ها برای کلاس (B)

$$B= 0.3$$

چون کلاس (A) دارای مجموع وزن بیشتری است، پیش‌بینی نهایی (A) خواهد بود.



Weighted Averaging Voting



رای گیری میانگین گیری وزنی

مثال: فرض کنید سه مدل داریم که نمرات احتمال برای یک نمونه را پیش بینی می کنند و وزن های آنها به ترتیب ۰.۲، ۰.۳ و ۰.۵ است

مدل ۱: [گربه: ۰.۲، سگ: ۰.۸] (وزن: ۰.۲)

مدل ۲: [گربه: ۰.۶، سگ: ۰.۴] (وزن: ۰.۳)

مدل ۳: [گربه: ۰.۵، سگ: ۰.۵] (وزن: ۰.۵)

برای محاسبه مجموع وزنی: گربه:
= گربه

$$\text{گربه} \quad (0.2 * 0.2) + (0.6 * 0.3) + (0.5 * 0.5) = 0.04 + 0.18 + 0.25 = 0.47$$

$$\text{سگ} \quad (0.8 * 0.2) + (0.4 * 0.3) + (0.5 * 0.5) = 0.16 + 0.12 + 0.25 = 0.53$$

پیش بینی نهایی: "سگ" (چون ۰.۵۳ بالاتر از ۰.۴۷ است)