Slide Report Ensemble

Lê Việt Hưng - AI Intern

Content

1 Introduction to Ensemble

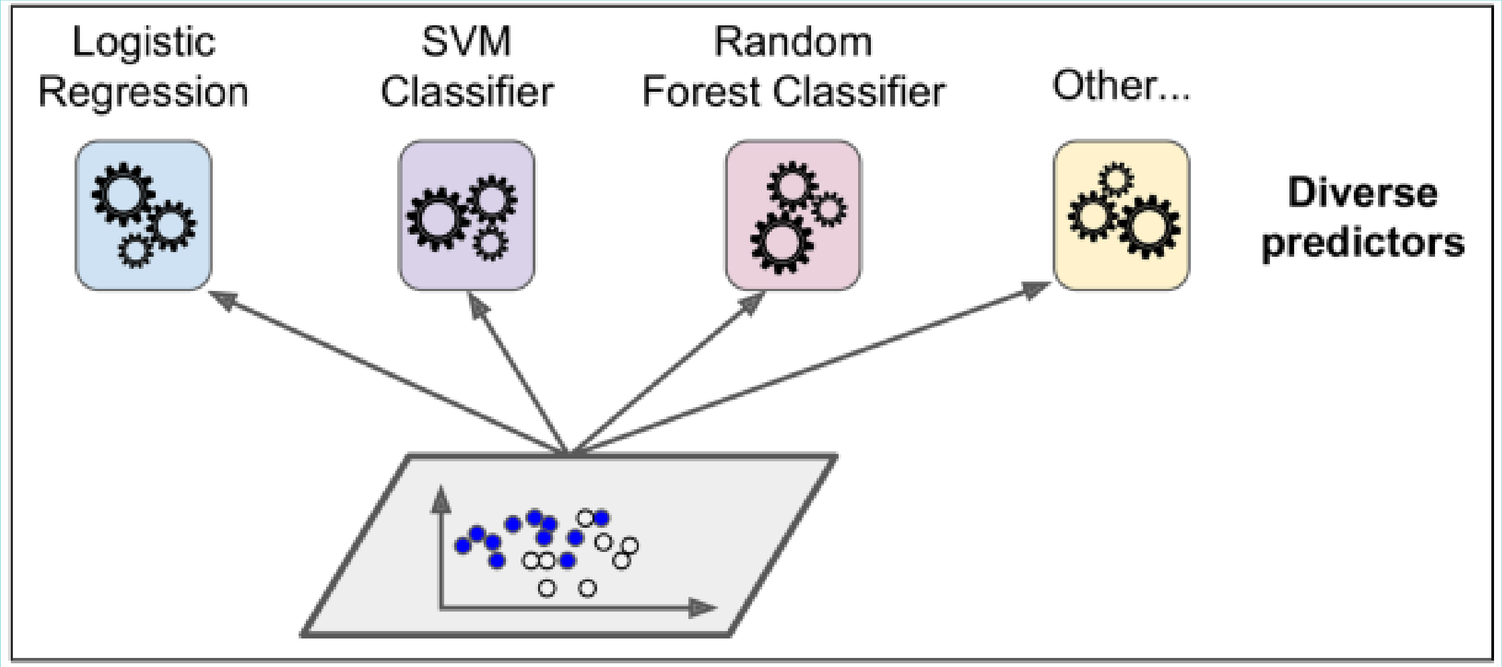
2 Bagging and Pasting

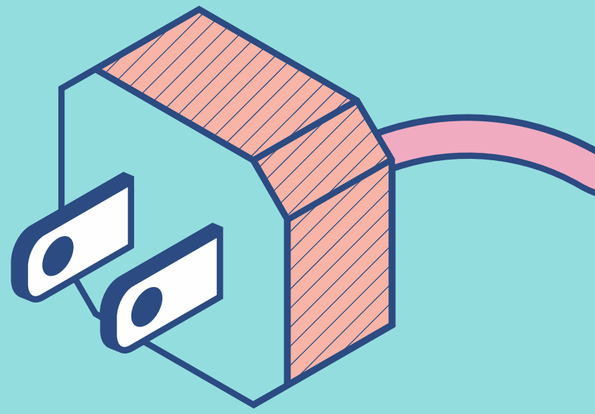
3 Boosting

1.Introduction to Ensemble

In machine learning, when you just use single machine learning model ,we will obtain a lower accuracy or performance.

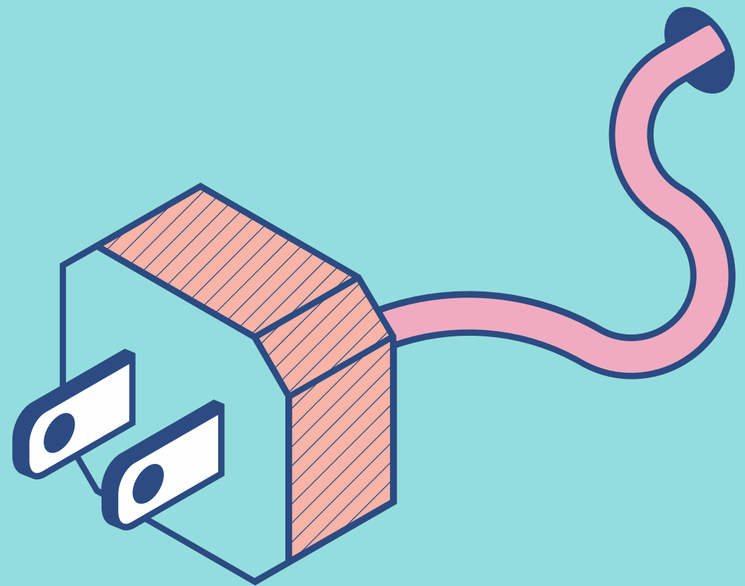
Ensemble refers to a technique that combines the predictions of multiple individual models to create a stronger and more accurate final prediction.



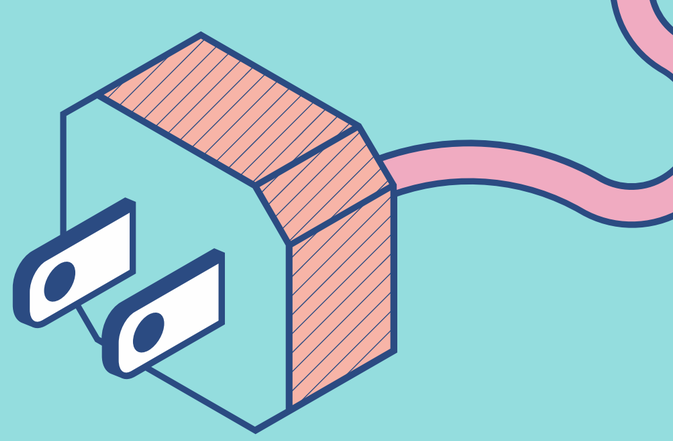


2.Bagging and Pasting

Bagging and Pasting are both ensemble learning techniques that aim to improve the performance and robustness of machine learning models by combining the predictions of multiple base models



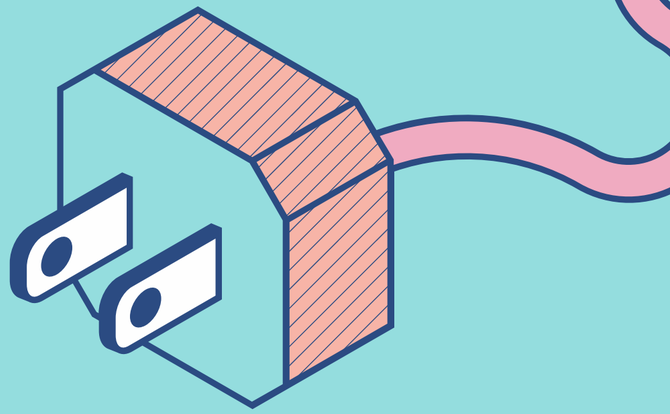
2.Bagging and Pasting

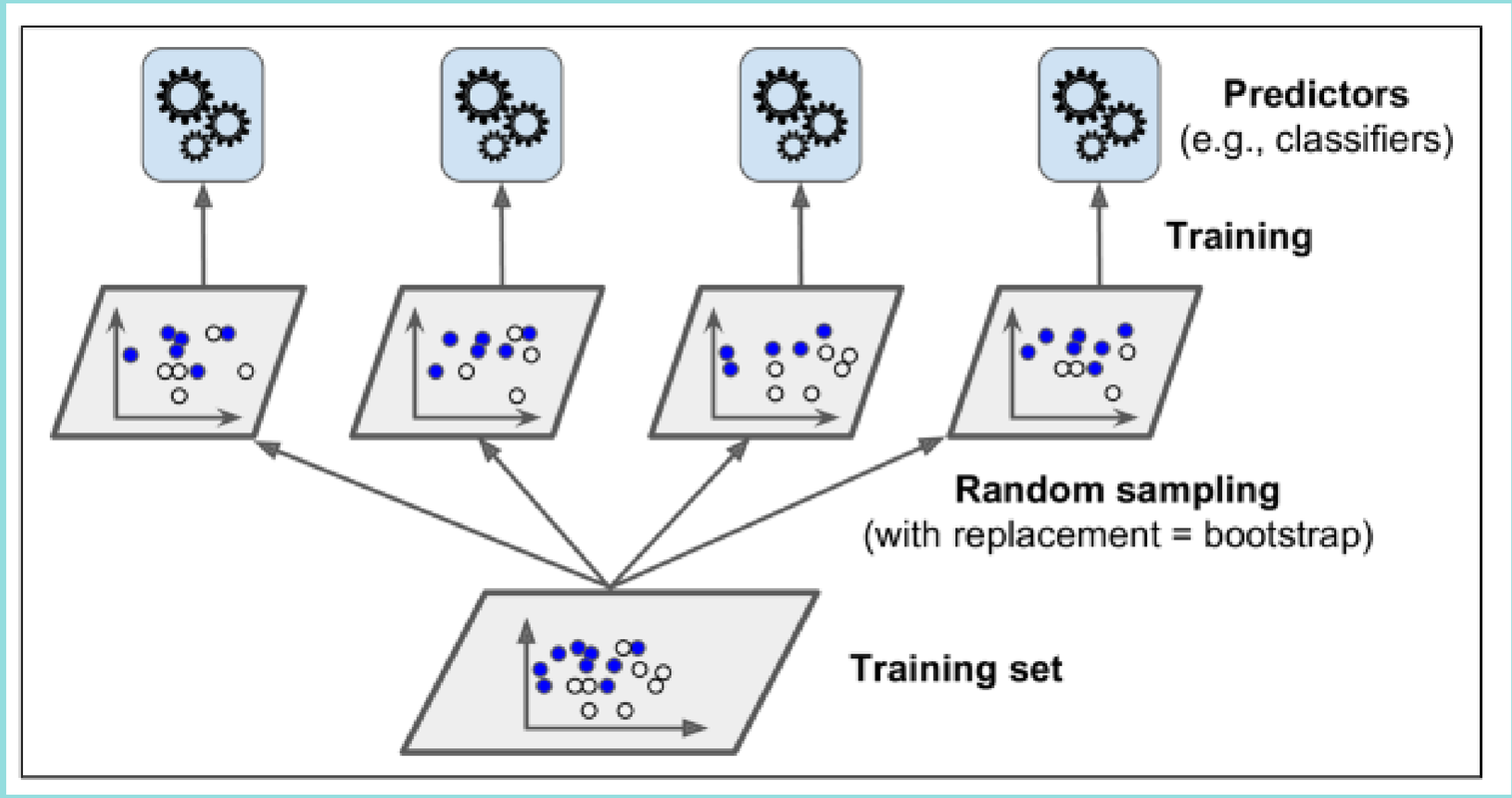


Compare Bagging and Pasting:

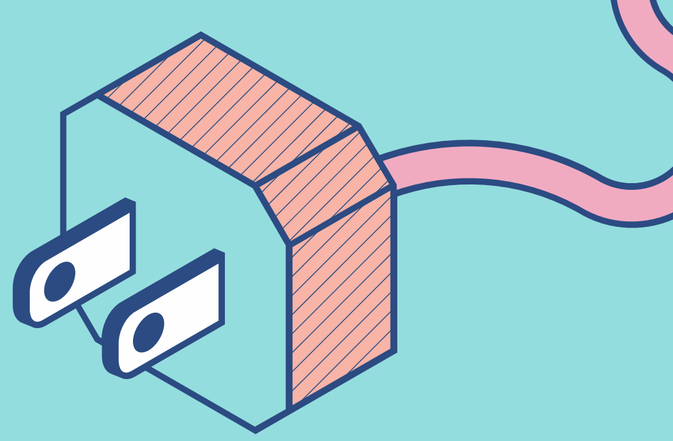
	Bagging	Pasting
Type of sampling	multiple instances of the same data in each subset.	each instance appears in only one subset.
Ensemble Size	larger ensembles due to repeated sampling and training.	smaller ensembles compared to bagging due to unique sampling.
Bias-Variance Tradeoff	reduce variance and minimize overfitting, as base models learn from different subsets.	higher variance compared to bagging since subsets have less diversity, but potentially lower bias.

2.Bagging and Pasting





2.Bagging and Pasting



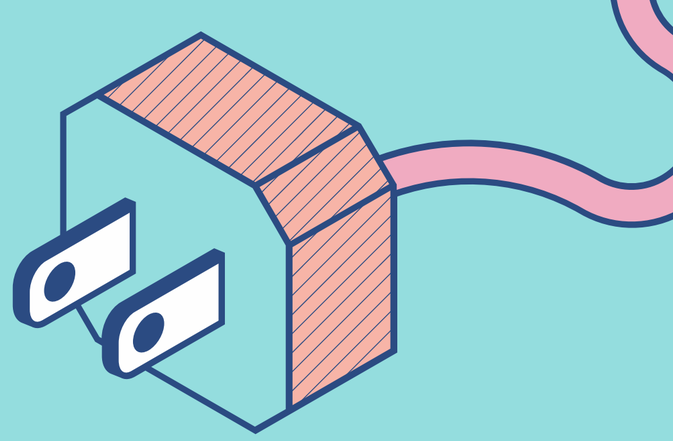
Code example: Bagging example code

```
from sklearn.ensemble import BaggingClassifier
from sklearn.tree import DecisionTreeClassifier

bag_clf = BaggingClassifier(
    DecisionTreeClassifier(), n_estimators=500,
    max_samples=100, bootstrap=True, n_jobs=-1)
bag_clf.fit(X_train, y_train)
y_pred = bag_clf.predict(X_test)
```

If you want to use Pasting ensemble, you can set up bootstrap = False

3.Boosting

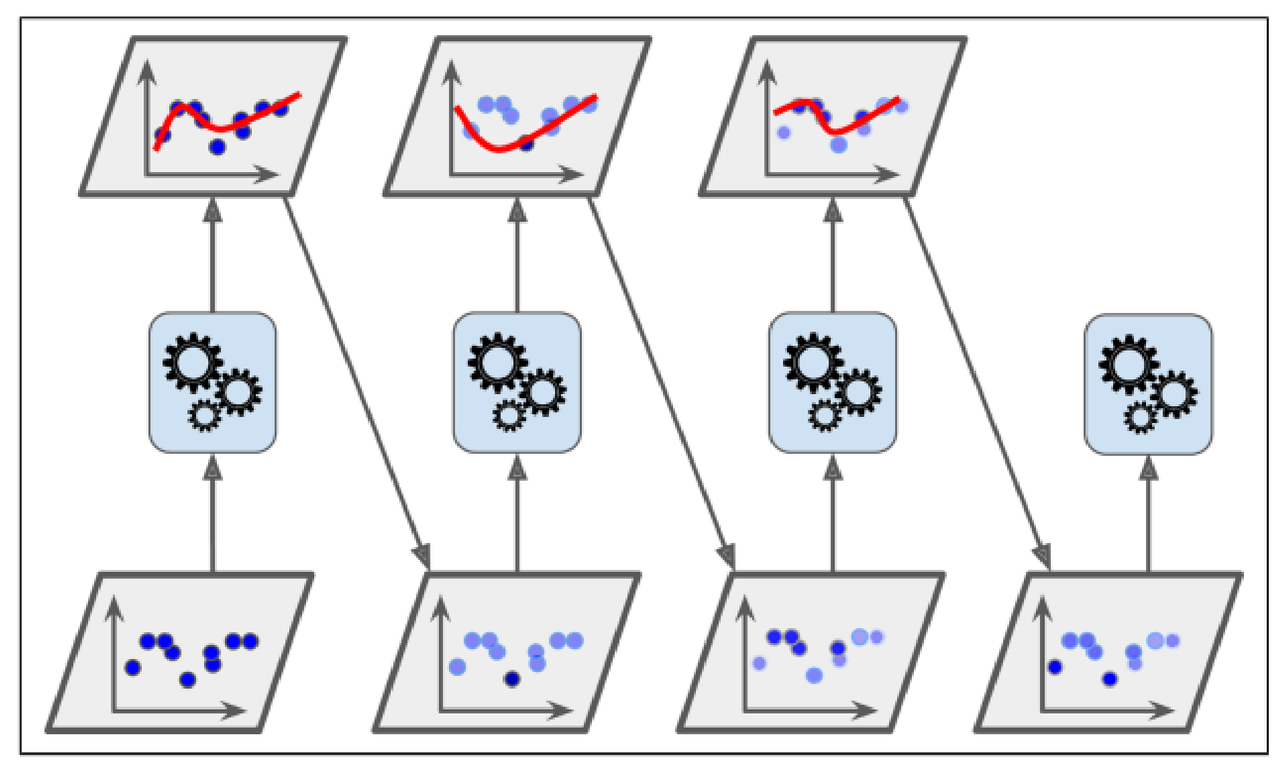


Boosting is a machine learning ensemble technique that aims to improve the predictive accuracy of a model by sequentially training a series of weak models and combining their predictions.

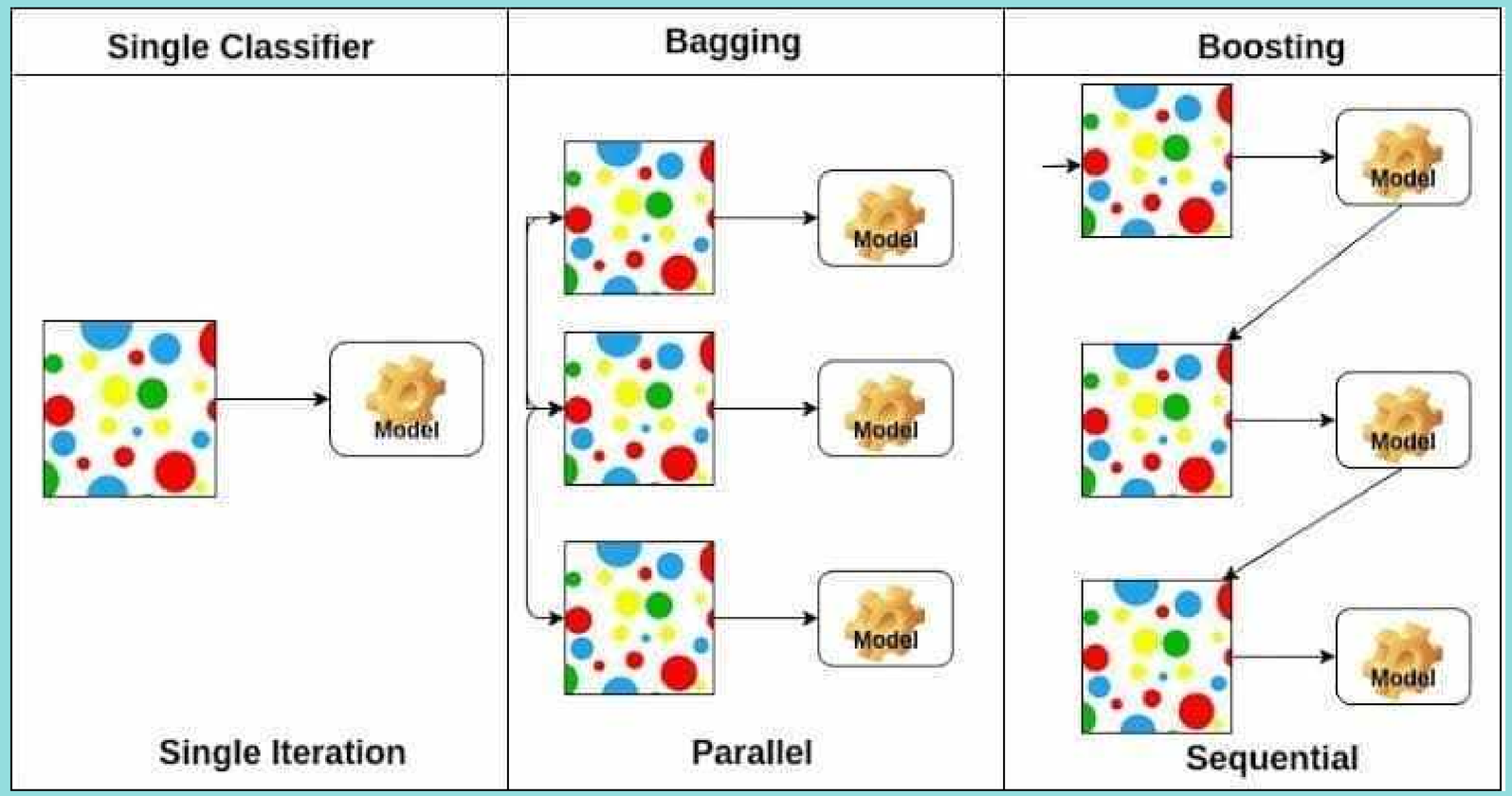
The key idea behind boosting is to focus on the mistakes or misclassifications made by the previous models and give more weight to the examples that were incorrectly predicted.

There are many boosting methods available, but by far the most popular are **AdaBoost** and **Gradient Boosting**

3.Boosting



3.Boosting


Single Classifier
Bagging
Boosting
Model
Model
Model
Model
Model
Model
Model
Model
Single Iteration
Parallel
Sequential

Thank you