Advanced ML Techniques

Unit 4: Session 01

Ensemble Learning Methods

Decision Tree

Random Forest

Bagging

Boosting (AdaBoost, Gradient Boosting)

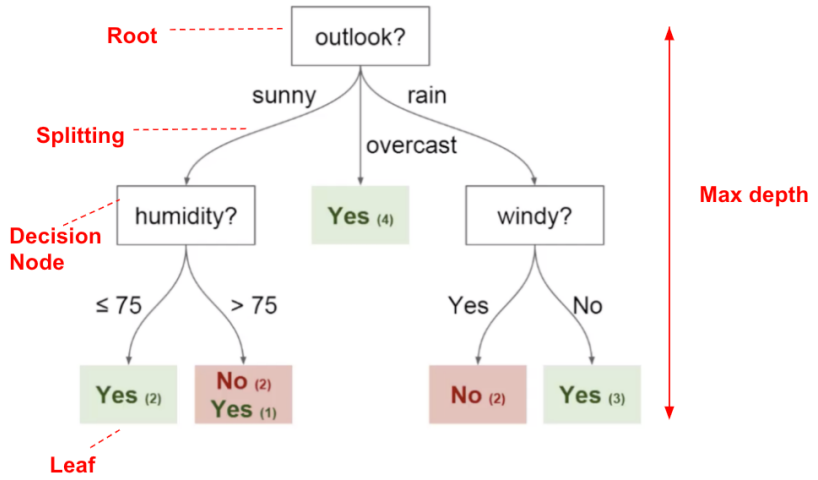
Decision Tree

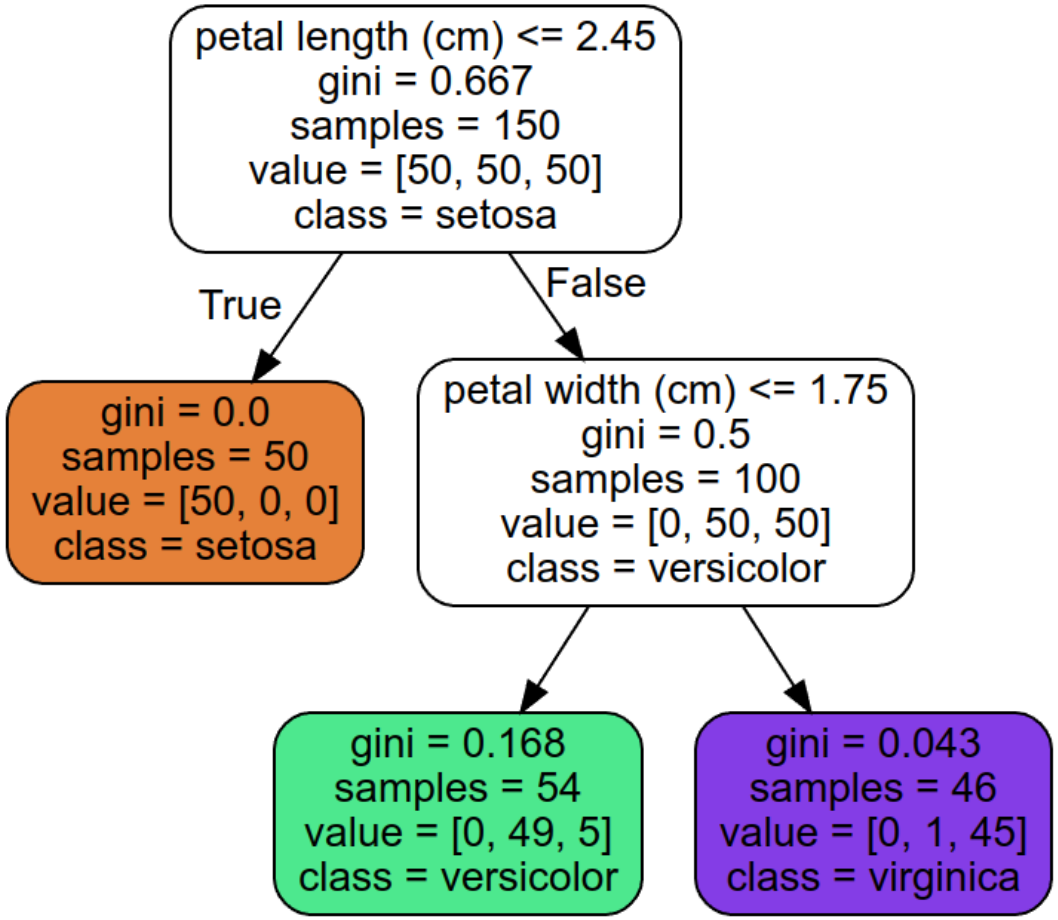
- Non-parametric supervised learning method
- Used for classification and regression.
- Easy to interpret.

Dataset

Temperature	Outlook	Humidity	Windy	Played?
Mild	Sunny	80	No	Yes
Hot	Sunny	75	Yes	No
Hot	Overcast	77	No	Yes
Cool	Rain	70	No	Yes
Cool	Overcast	72	Yes	Yes
Mild	Sunny	77	No	No
Cool	Sunny	70	No	Yes
Mild	Rain	69	No	Yes
Mild	Sunny	65	Yes	Yes
Mild	Overcast	77	Yes	Yes
Hot	Overcast	74	No	Yes
Mild	Rain	77	Yes	No
Cool	Rain	73	Yes	No
Mild	Rain	78	No	Yes

Decision Tree Diagram

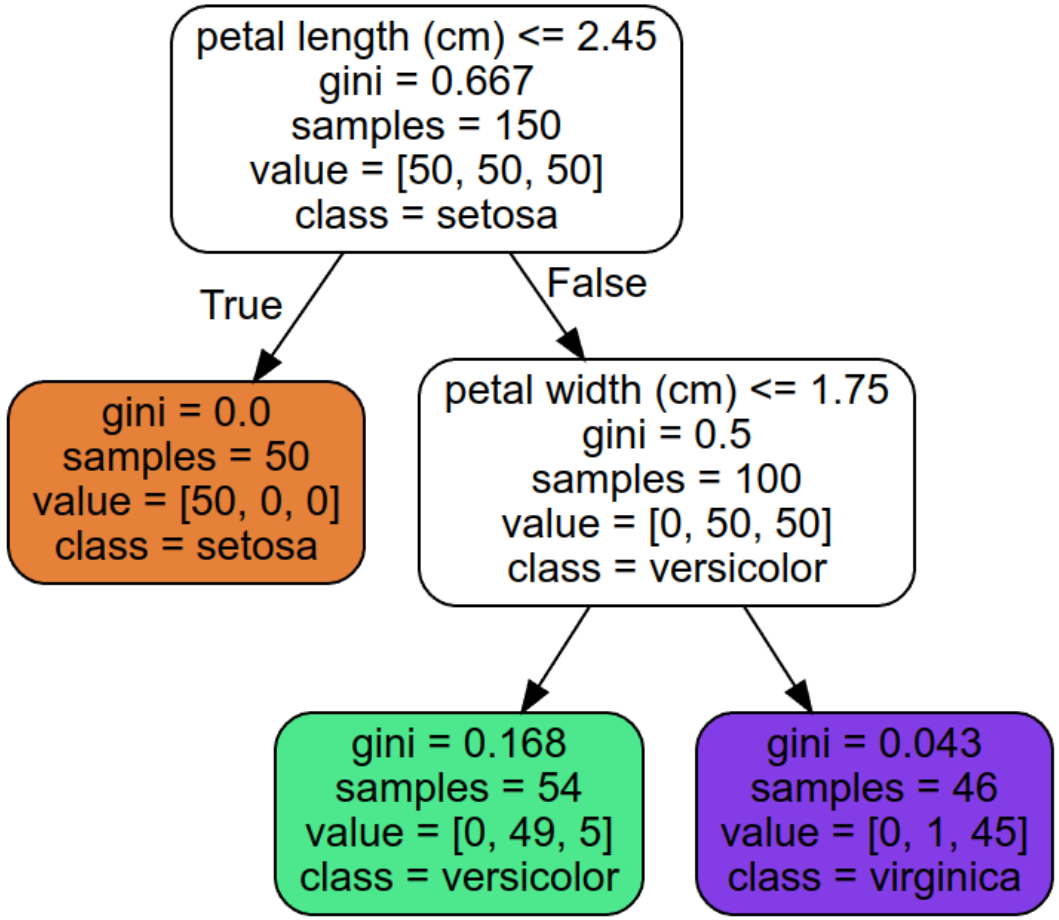




What will be the class for flower with petal length smaller than 2.45cm?

Base on which attribute (feature) to split? What is the best split?

Attribute with the highest **Information Gain** or **Gini Gain**



Problems with Decision Trees

- Overfitting
- Very sensitive to training data.

-

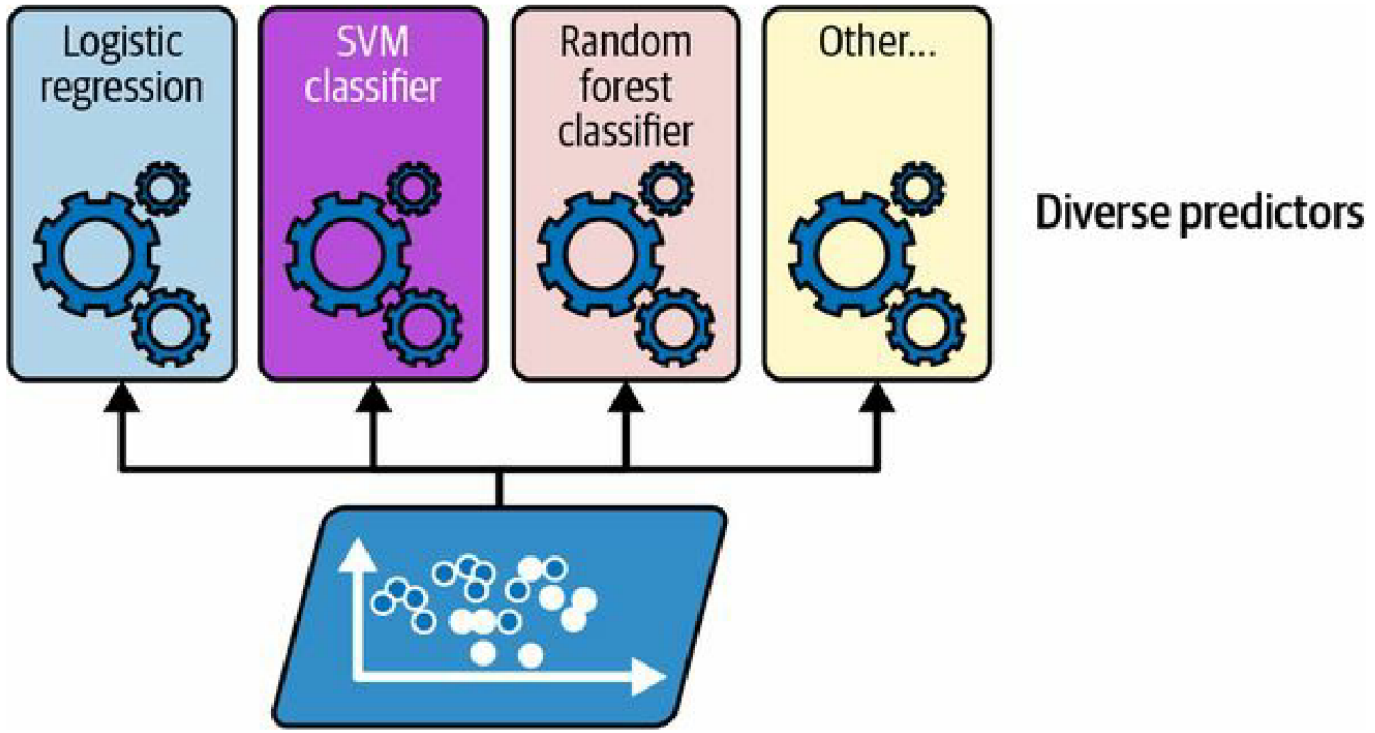
Wisdom of the crowd

-

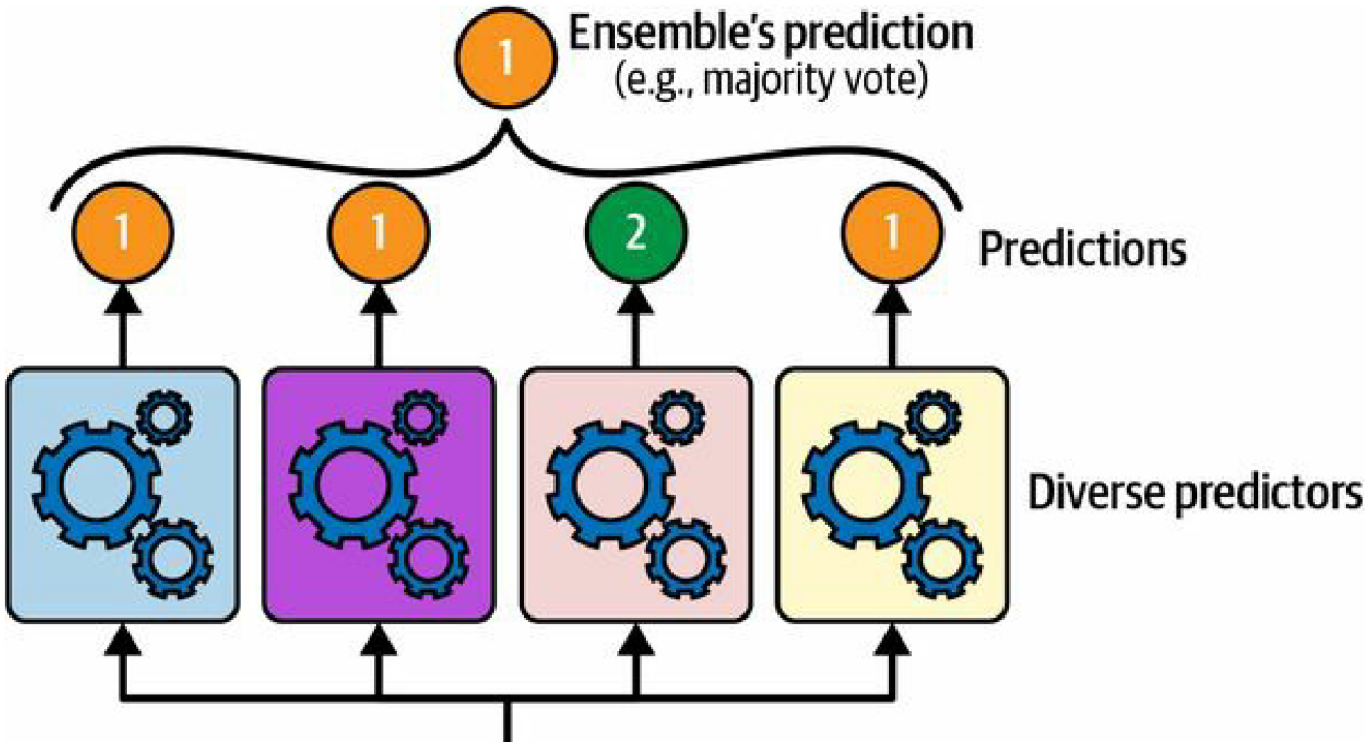
Ensemble learning is a general meta approach to machine learning that seeks better predictive performance by combining the predictions from multiple models.

-

Voting Classifier

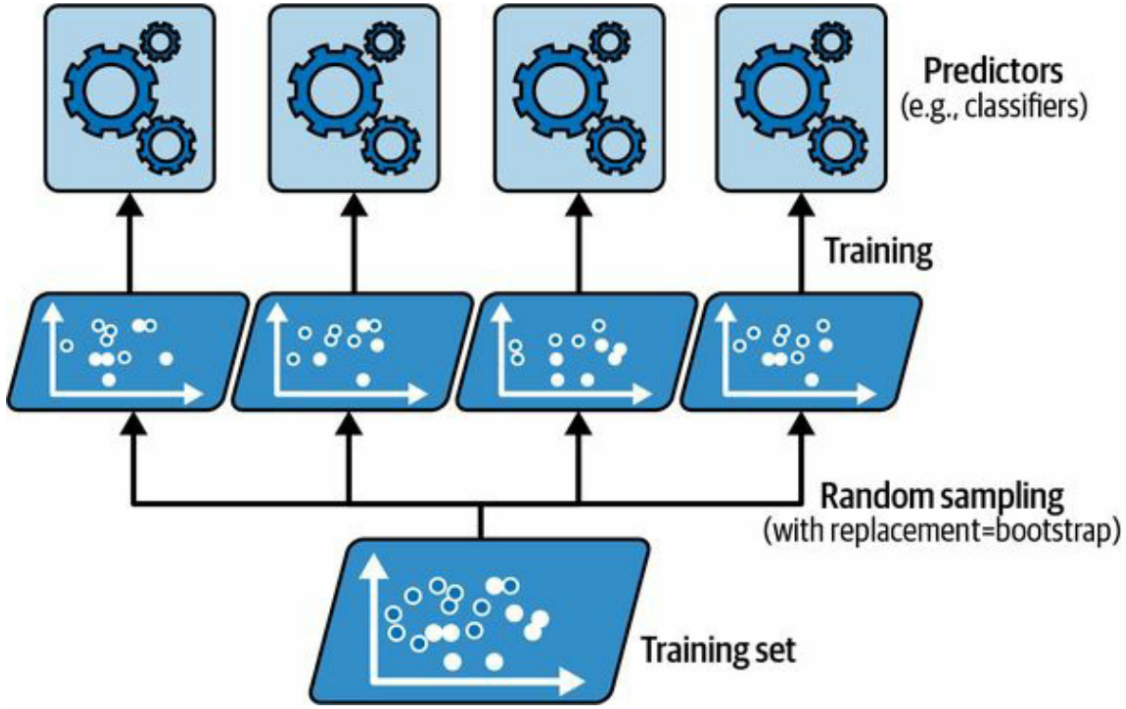


-


1
Ensemble's prediction
(e.g., majority vote)
1
1
2
1
Predictions
Diverse predictors

-

Bagging & Pasting



Random Forest

An ensemble of decision trees.

Random forests solve the problem of overfitting because they combine the output of multiple decision trees to come up with a final prediction.

Parallel Learning

Boosting

Combine several weak learner into a strong learner.

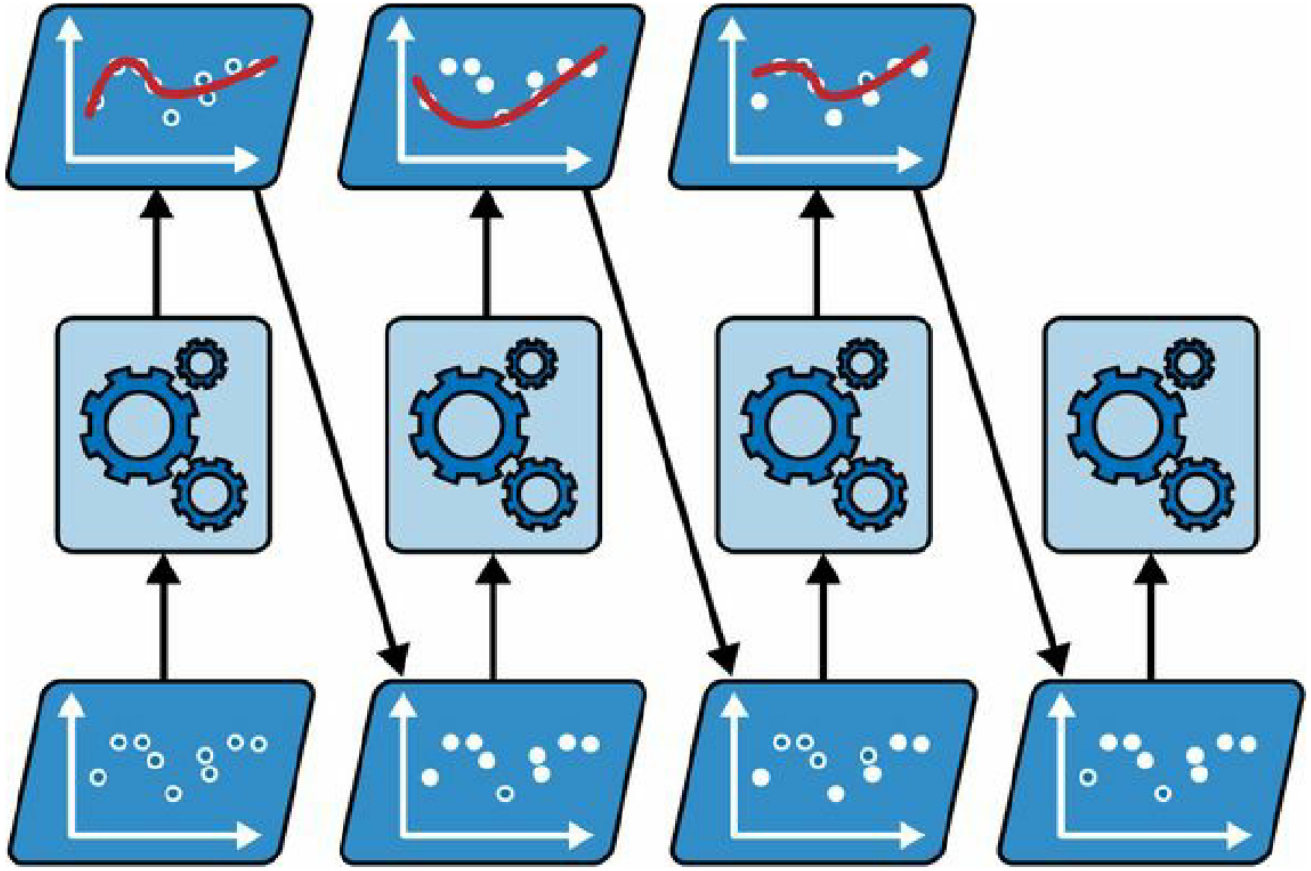
Sequential Learning.

Each predictor is trying to correct it's predecessor.

AdaBoost (Adaptive Boosting)

- Step 01: Train a base classifier. Use it to predict training data.
- Step 02: Increases the relative weight of misclassified training instances.
- Step 03: Train second classifier using the updated weight.
- Repeat ...

AdaBoost Sequential Learning

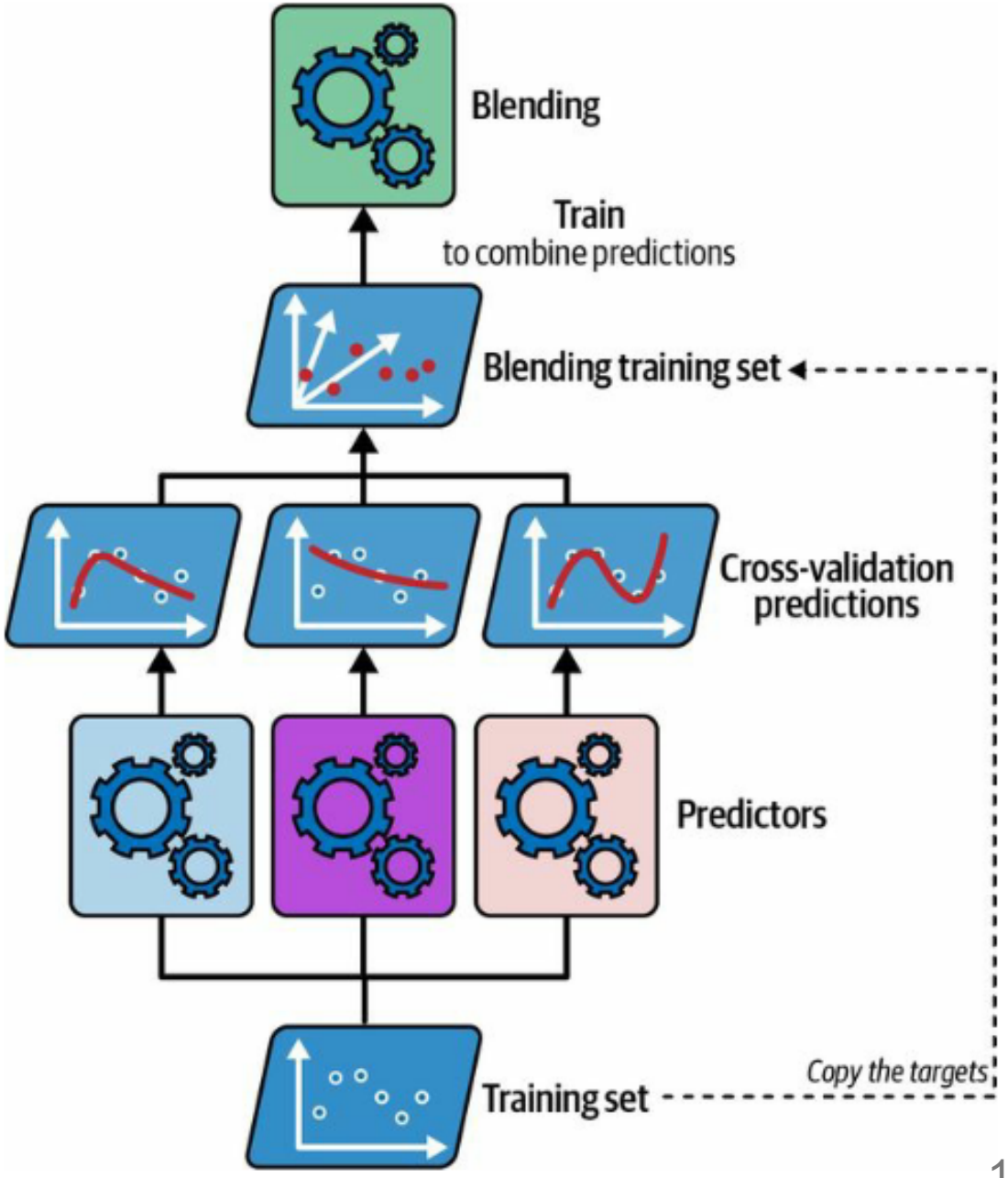


Gradient Boosting

- Tries to fit the new predictor to the **residual errors** made by the previous predictor.

Stacking

Instead of aggregating the predictions of other predictors, **a new model is trained** to perform this aggregation.



Thank You

Happy Learning :)