Chapter 6

Ensemble Learning & Random Forests

Machine Learning

CONTENTs

- **Voting Classifiers**

- **Bagging**

- **Random Forests**

- **Boosting**

- **Stacking**

- **Ensemble Learning & Random Forests**
  - wisdom of the crowd (wisdom of the crowd)
  - The predictions of a group of predictors (such as classifiers or regressors), it will often get better predictions than with the best individual predictor.
  - A group of predictors is called an **ensemble => Ensemble Learning**
  - Train a group of Decision Tree classifiers, each on a different random subset of the training set. To make predictions, you just obtain the predictions of all individual trees, then predict the class that gets the most votes. **An ensemble of Decision Trees is called a Random Forest**
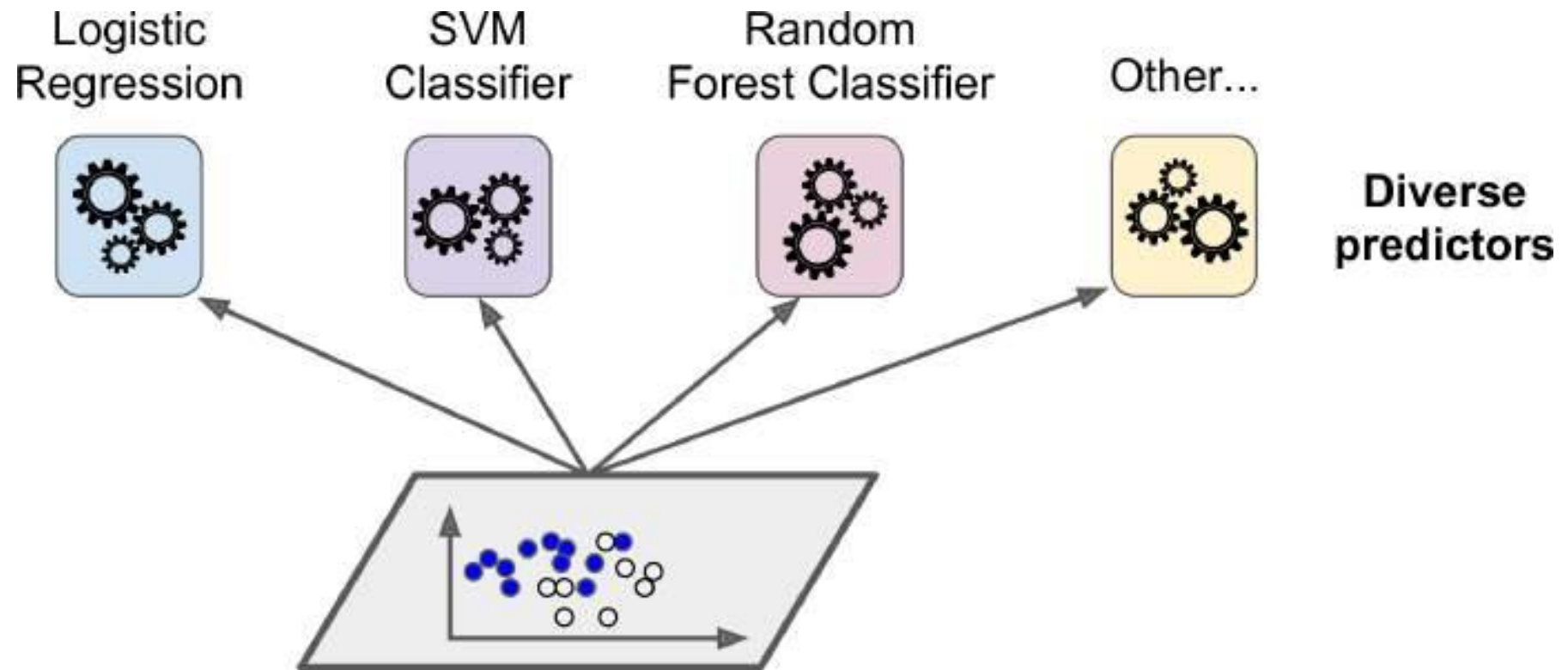
Training diverse classifiers

- to create an even better classifier is to aggregate the predictions of each classifier and predict the class that gets the most votes

- This majority-vote classifier is called a hard voting classifier
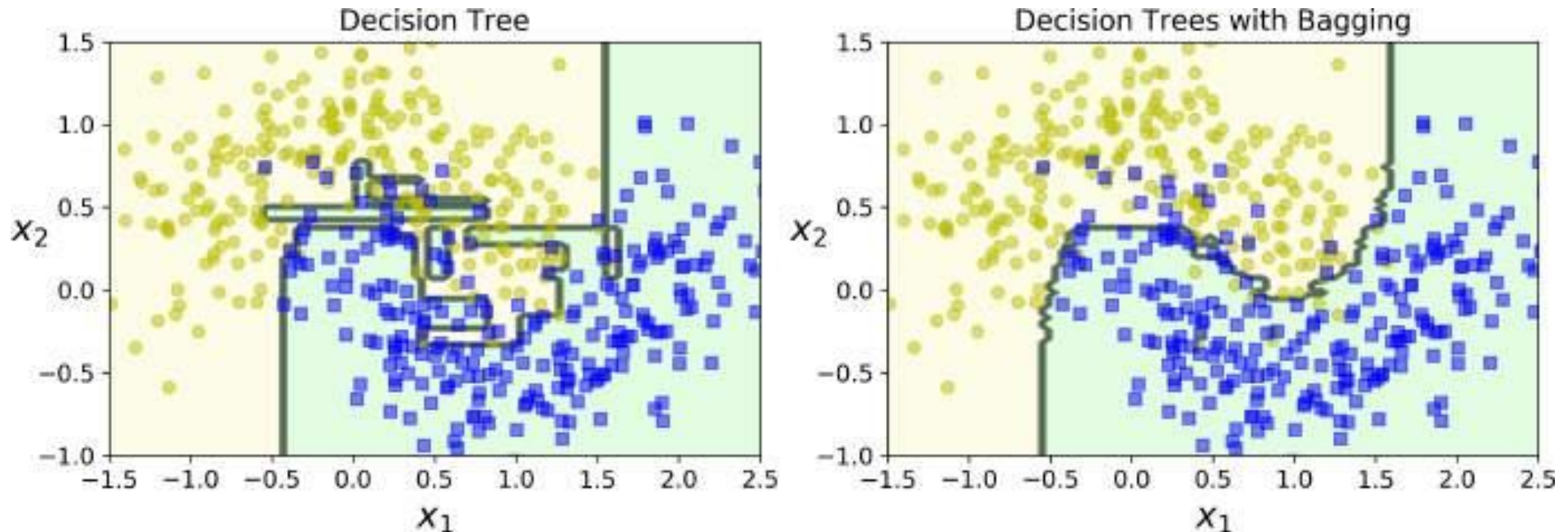

Hard voting classifier predictions
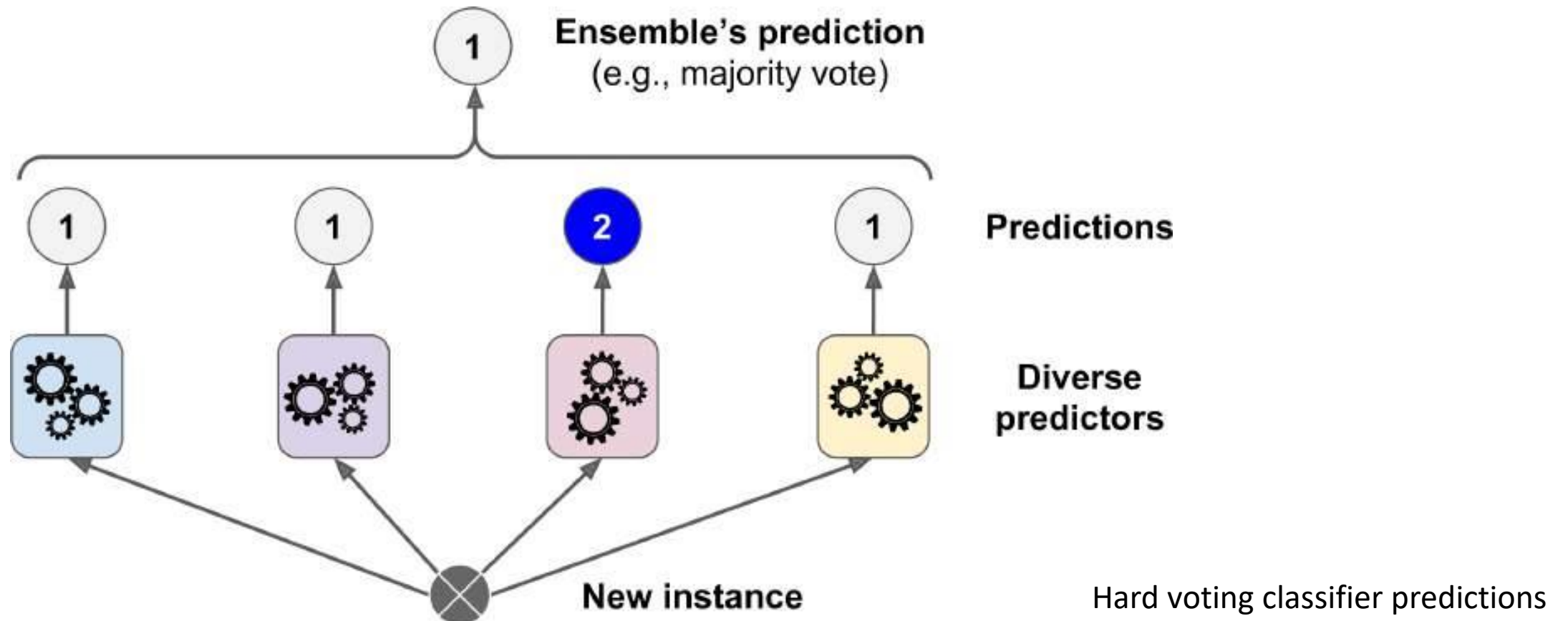
- **Bagging**
  - use the same training algorithm for every predictor, but to train them on different random subsets of the training set
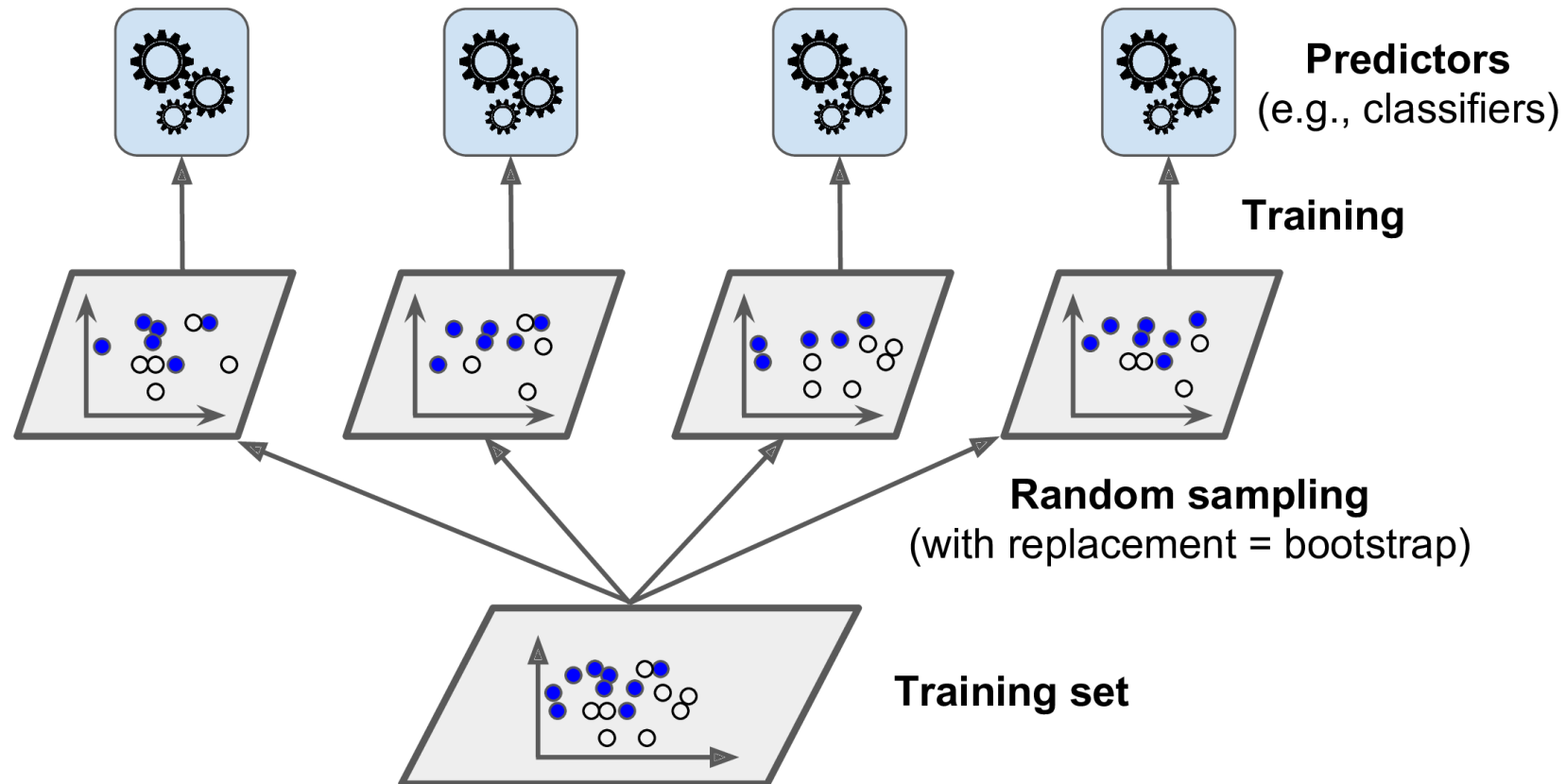
- **Bagging**
  - compares the decision boundary of a single Decision Tree with the decision boundary of a bagging ensemble of 500 trees, both trained on the dataset


A single Decision Tree versus a bagging ensemble of 500 trees

- **Random Forest is an ensemble of Decision Trees, generally trained via the bagging method**

- **The Random Forest algorithm introduces extra randomness when growing trees:**

  - instead of searching for the very best feature when splitting a node, it searches for the best feature among a random subset of features.

  => This results in a greater tree diversity

- Boosting refers to any Ensemble method that can combine several weak learners into a strong learner.

- The general idea of most boosting methods is to train predictors sequentially, each trying to correct its predecessor.

- There are many boosting methods available, the most popular are AdaBoost (short for Adaptive Boosting) and Gradient Boosting.
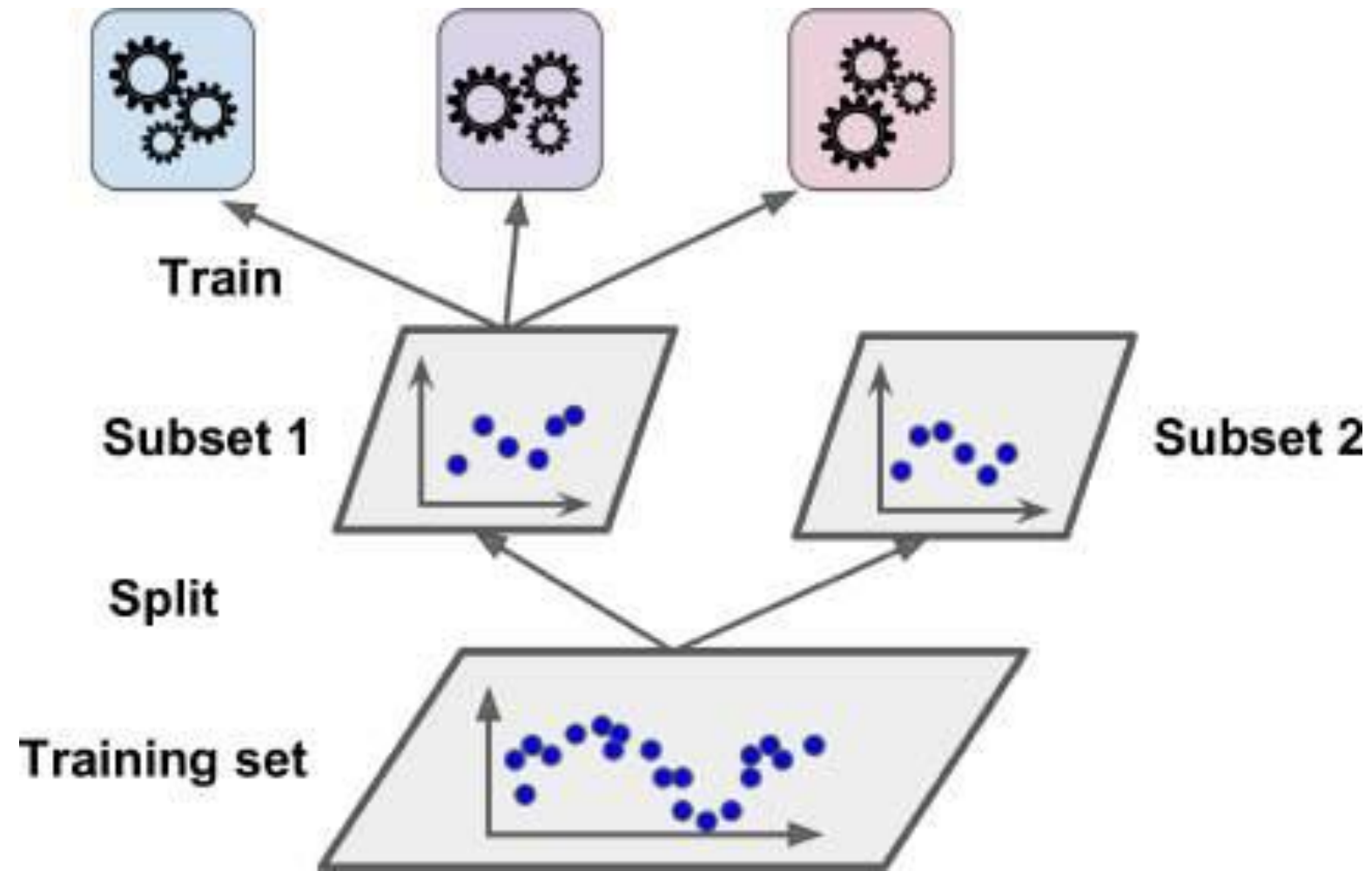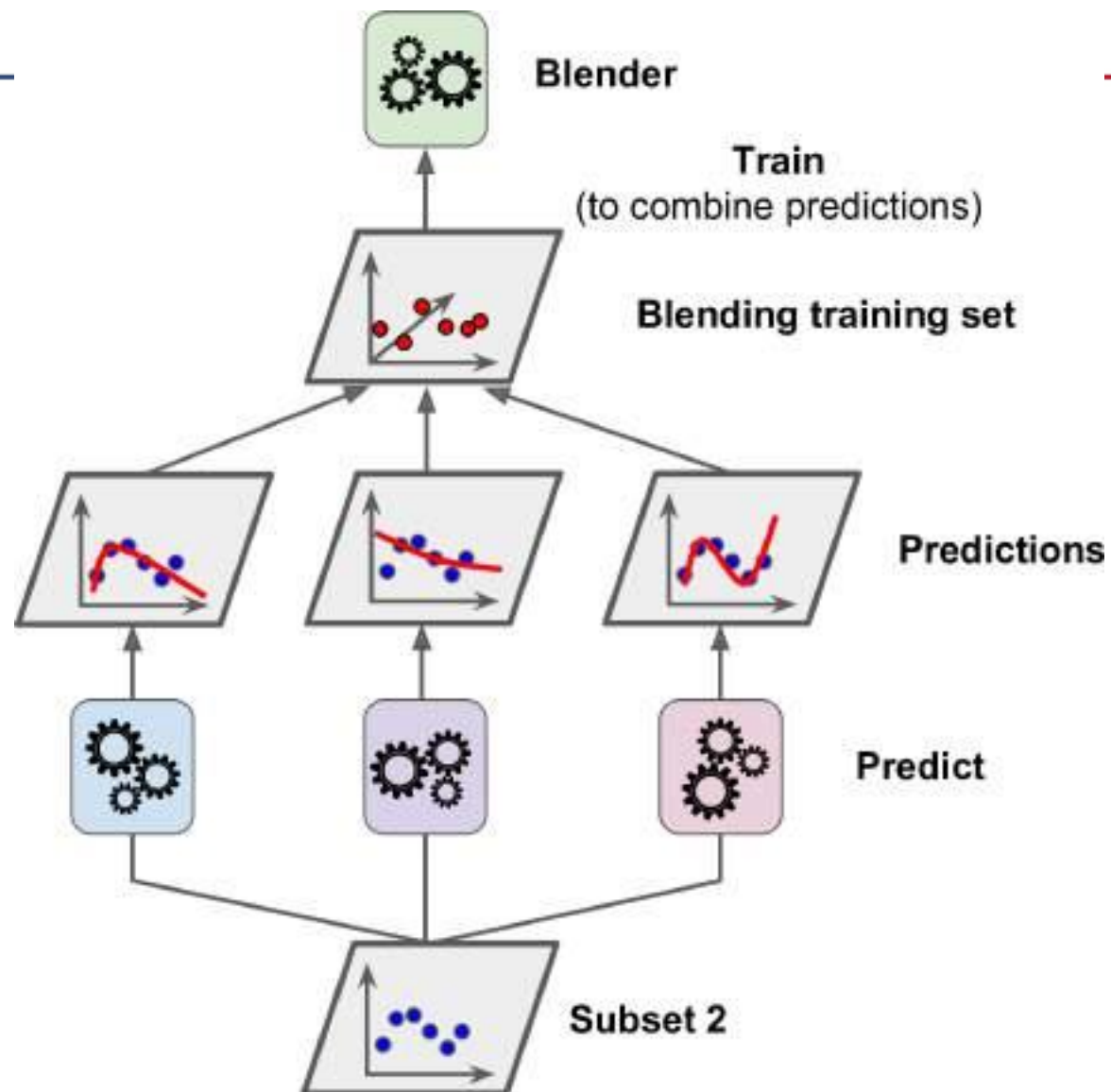
- **AdaBoost**
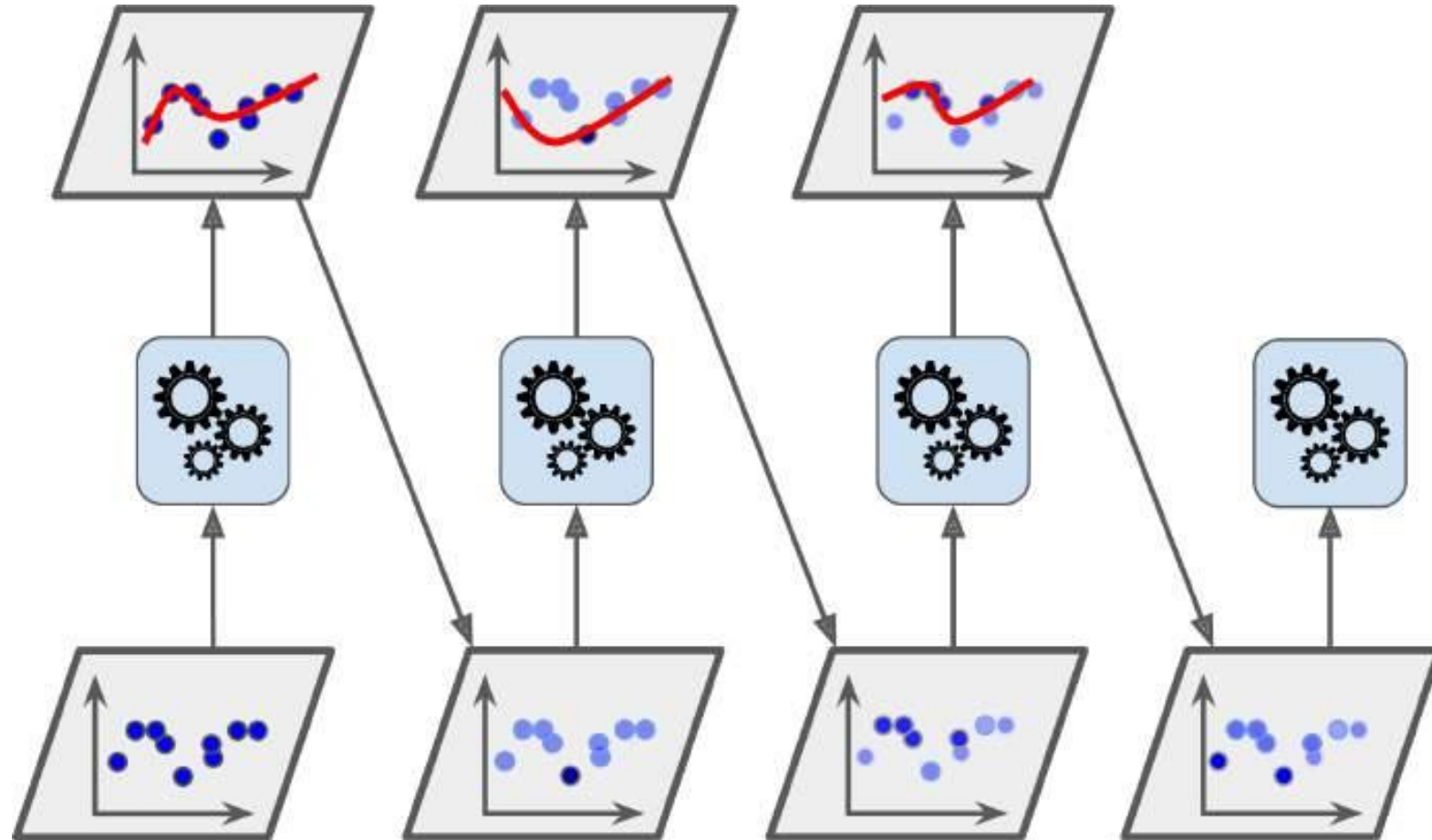
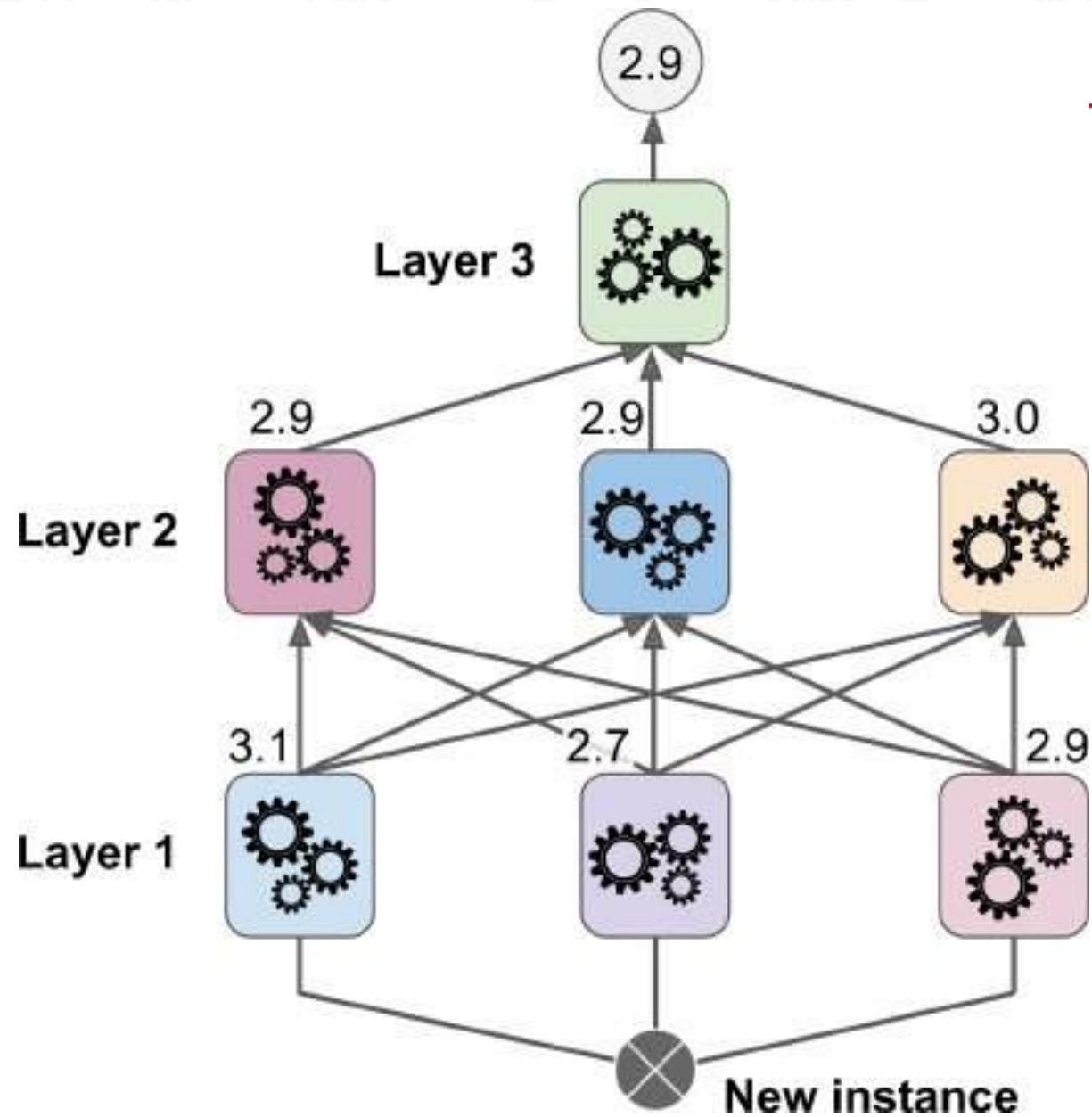- AdaBoost sequential training with instance weight updates
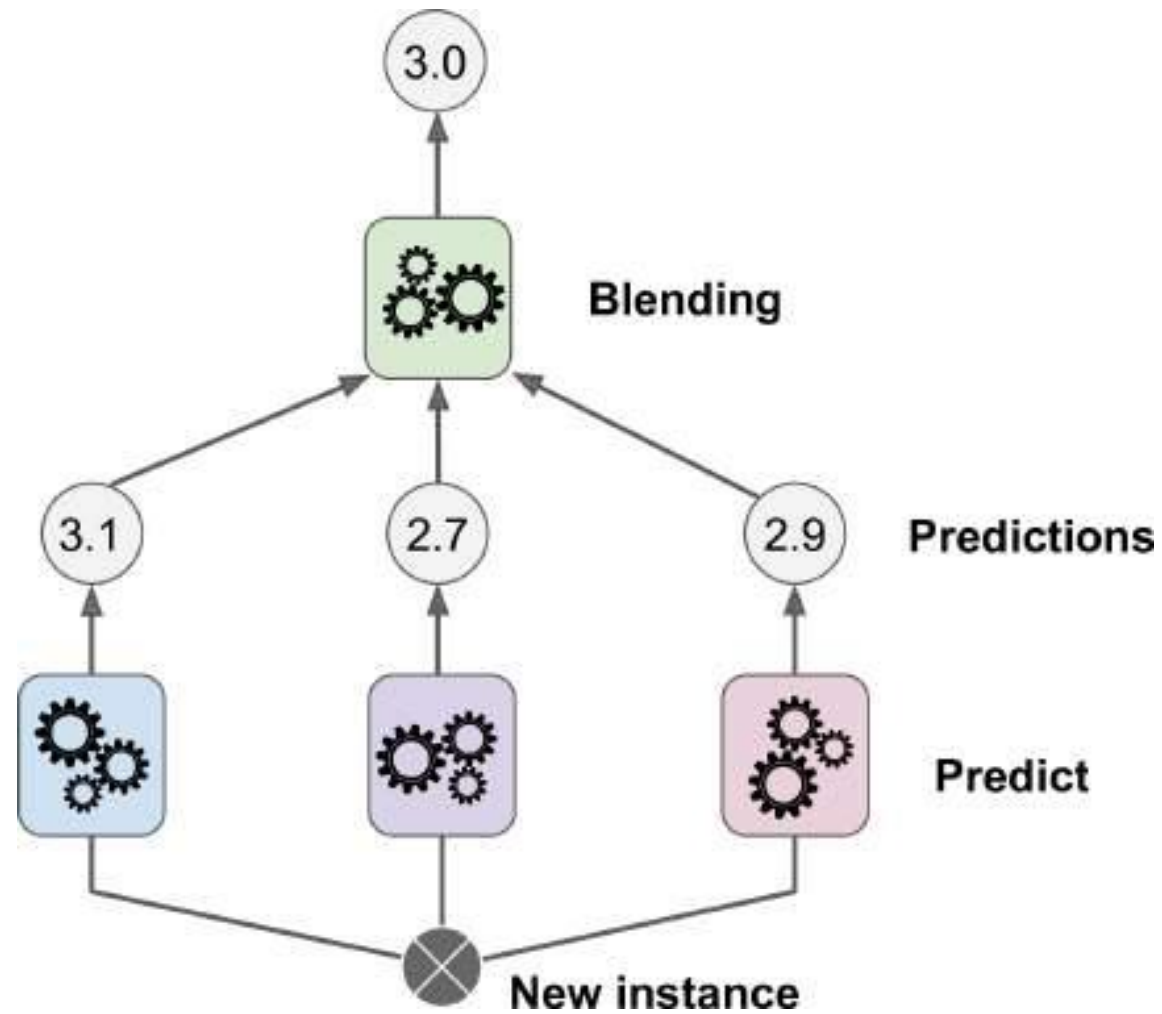
• Aggregating predictions using a blending predictor

Train

Subset 1

Subset 2

Split

Training set

- Training the first layer

- Training the blender

- Predictions in a multilayer stacking ensemble

- **Voting Classifiers**

- **Bagging**

- **Random Forests**

- **Boosting**

- **Stacking**

Enjoy the Course...!