Ensemble learning

Lecturer: Dr. Gábor Hullám

Improve prediction performance 1.

Is it possible to improve the accuracy and reliability of the prediction (model)?

A: „We have one model"

- Regularization – apply techniques against overfitting

- Hyperparameter tuning
  - In fact, we have several similar models by hyperparameters
  - From these, we select the "best"" → Model selection

- Cross-validation methods
  - In fact, we have several similar models by data partition usage
  - average performance and the range of performance indicators are of interest → determining reliability

Improve prediction performance 2. – Ensemble of models

B: *„We have several different models”*

- Is it possible to use a combination of predictions (e.g. mean or other aggregate results) of several models?
- Yes!
- **Ensemble learning:** The process of creating several different models to solve a given problem and then aggregating the prediction of these models to produce the final output.

Ensemble of models

Advantage

- If multiple models are taught, the set of models can give better predictions than a model on average

Disadvantage

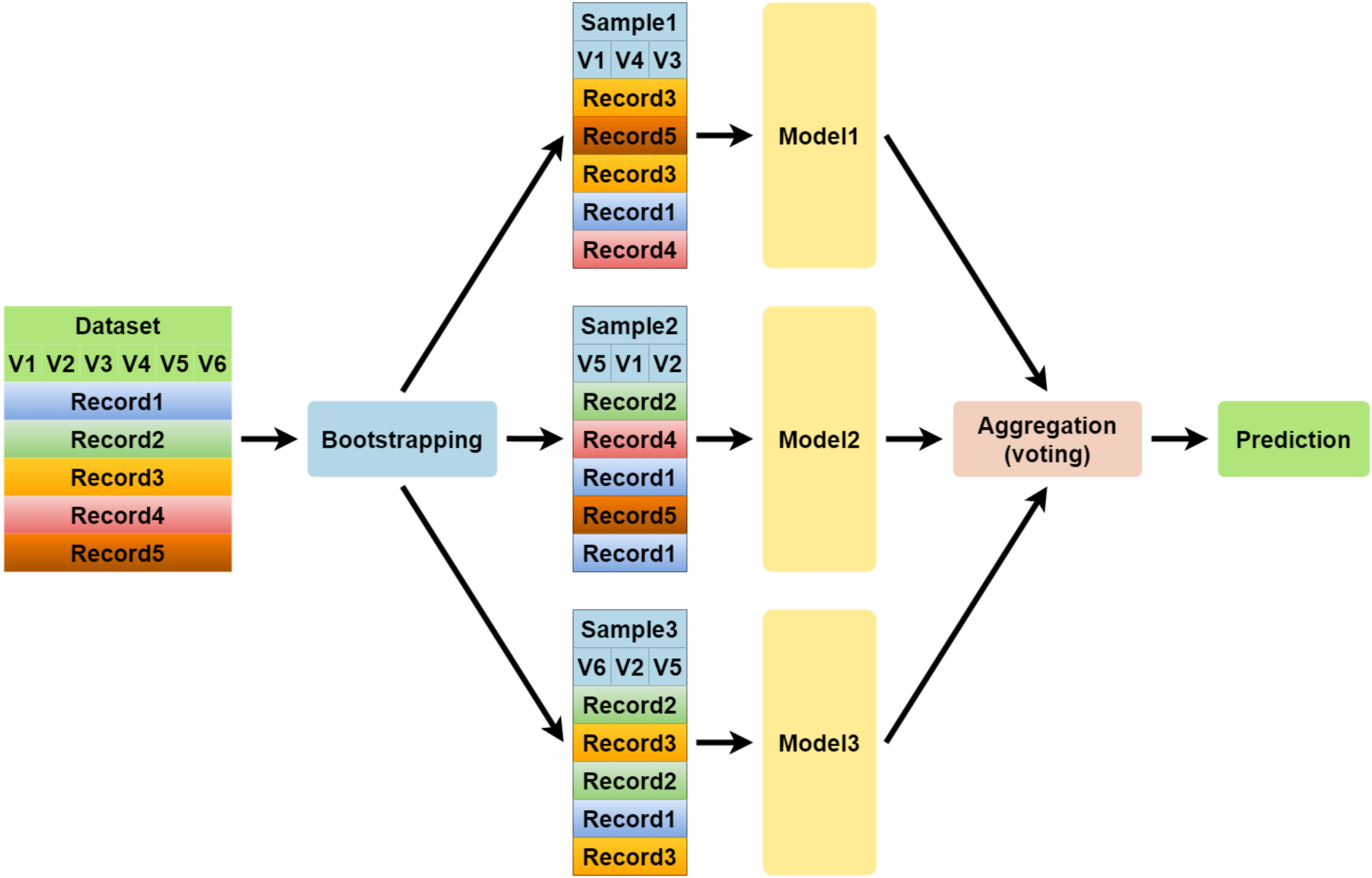
- More models mean more training, more computational resources required

Bagging

Steps of bagging:

1. Let's take as a starting point a dataset of N records and k variables.
2. Select the true subset k∈K of the variables (these are columns of the data matrix of size K∗N) by uniform distribution, and then apply Bootstrapping and get an N-sized record set for selected variables.
3. Fit a model to the resulting data set.
4. Repeat steps 2 and 3 until a predetermined number of models are created.

Bagging



Dataset
V1 V2 V3 V4 V5 V6
Record1
Record2
Record3
Record4
Record5
Bootstrapping
Sample1
V1 V4 V3
Record3
Record5
Record3
Record1
Record4
Sample2
V5 V1 V2
Record2
Record4
Record1
Record5
Record1
Sample3
V6 V2 V5
Record2
Record3
Record2
Record1
Record3
Model1
Model2
Model3
Aggregation (voting)
Prediction

Bagging

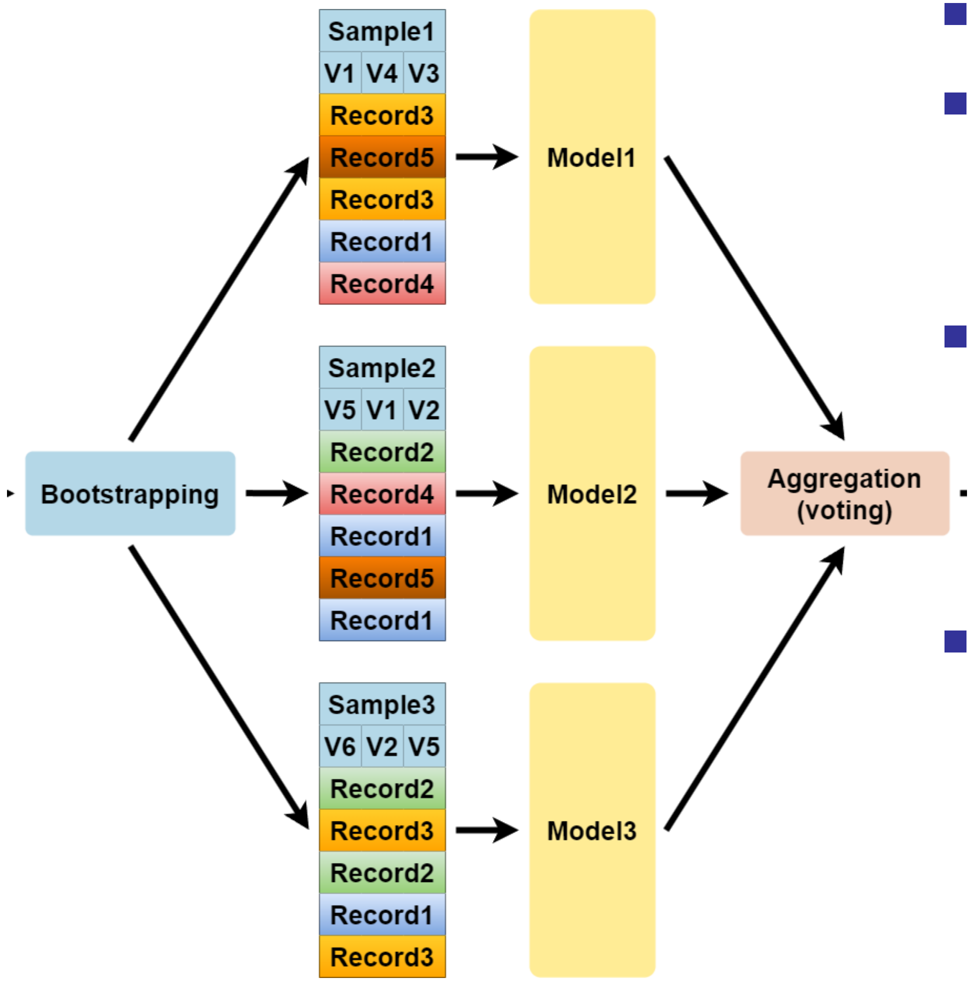
- Repeat steps 2 and 3 until you get a predetermined number of models.

- 5. We aggregate the prediction of the resulting set of models for a new (test) sample
  - by majority decision in the case of classification,
  - in the case of regression, it is determined by averaging.

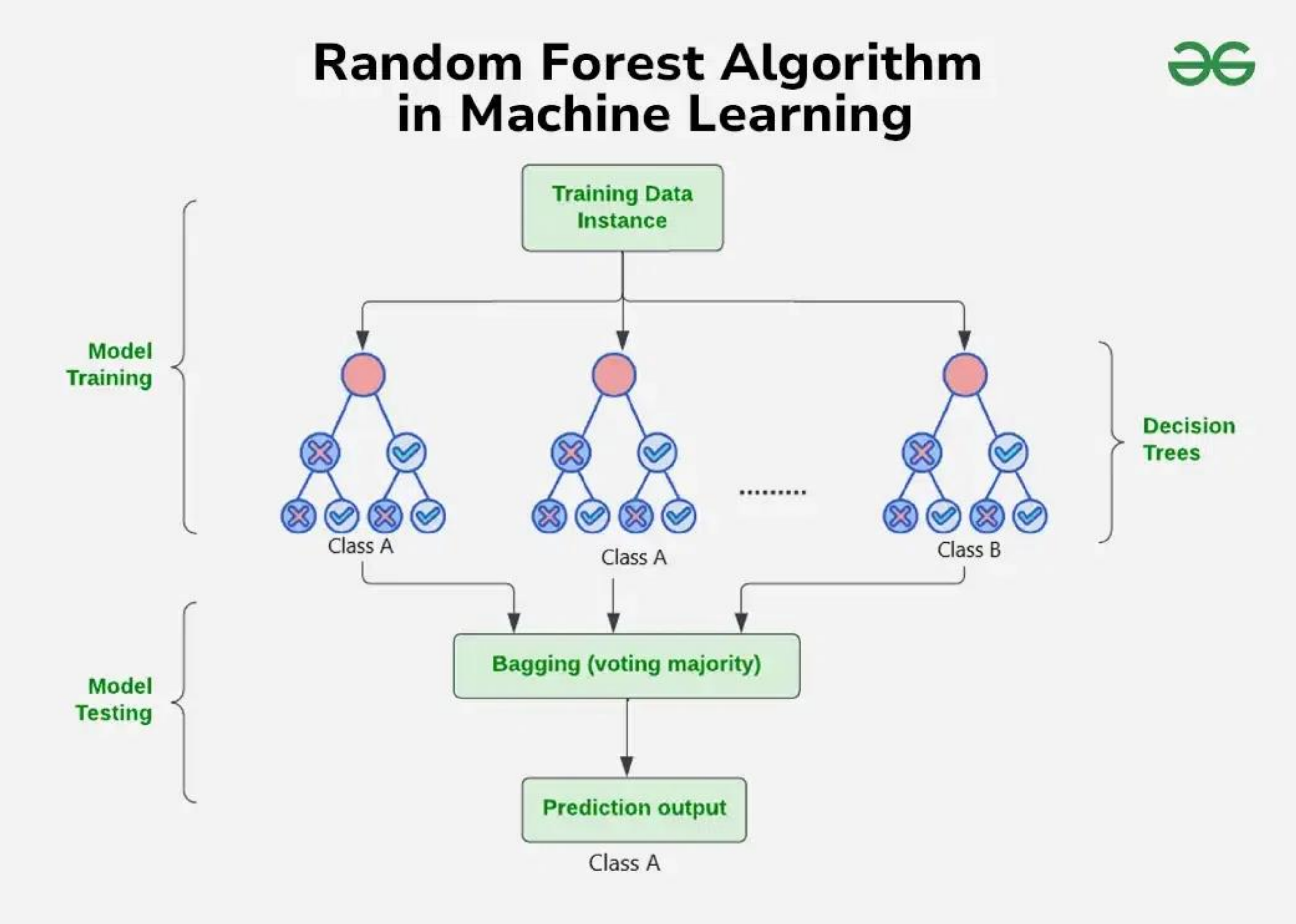
Heterogeneous models

- The data set created by sampling in Step 2 (Bootstrapping) contains only a subset of variables and samples
- Each model learns on different subsets of the original data set
- This ensures that model types with deterministic (e.g. decision tree) and low-variance (e.g. logistic regression) training processes will have a heterogeneous set of models,

Bagging – Random forests



- Decision trees+ bagging =random forest
- the original Bagging algorithm always performed Bootstrap sampling for the entire set of variables.
- In the case of random forests, however, it is preferable that the set of variables is also randomly drawn, since then the individual models are less correlated with each other.
- The common name for this technique is *Feature Bagging*, in case of random forests, it forms part of the Bagging.



https://www.geeksforgeeks.org/random-forest-algorithm-in-machine-learning/

Bagging – Advantages and disadvantages

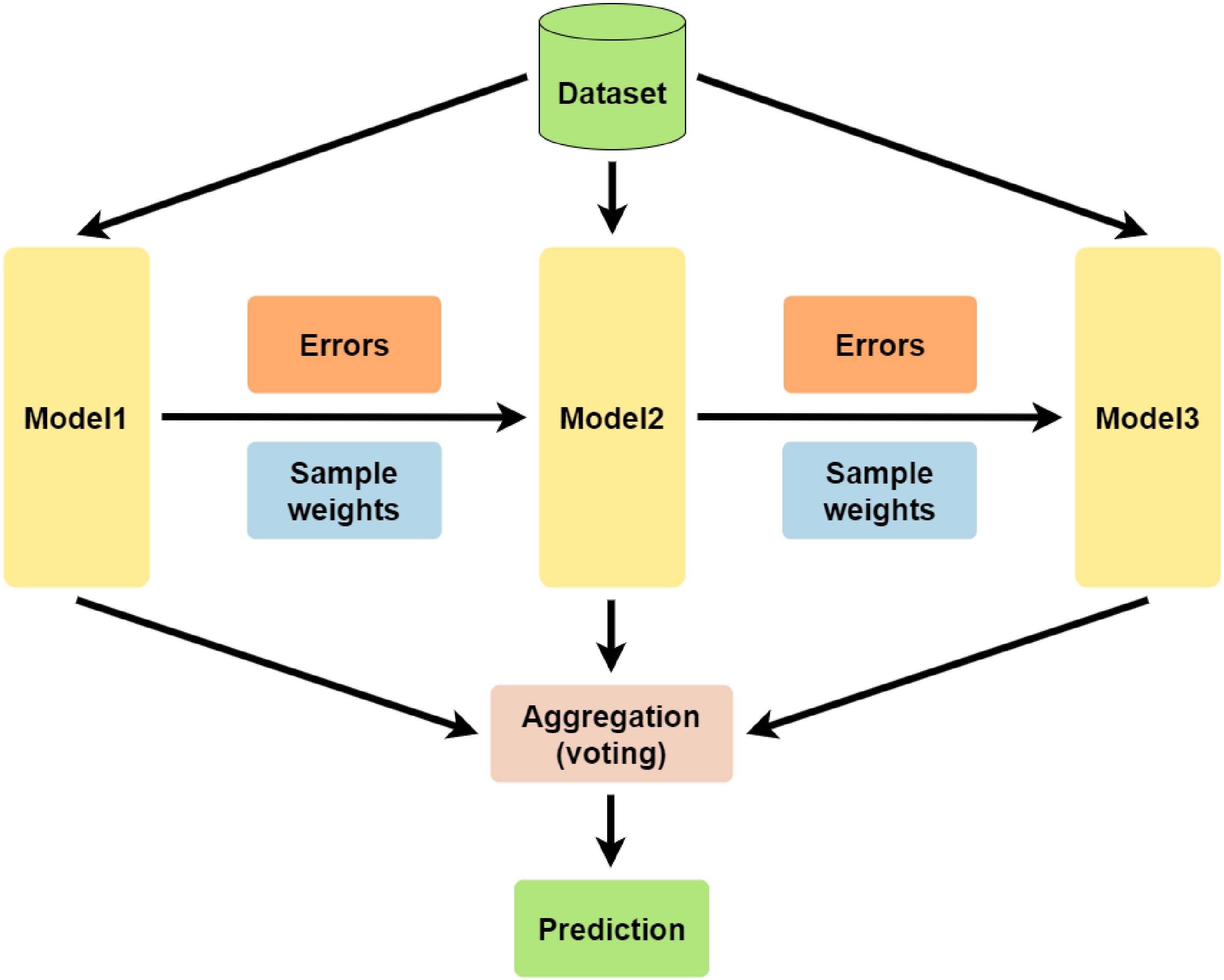
- + The creation of individual models can be parallelized
  - models were trained completely independently of each other, based on a randomly selected sample

- - need a large number of models
  - to ensure that the accuracy of aggregate prediction is significantly better than that of individual models.

Boosting

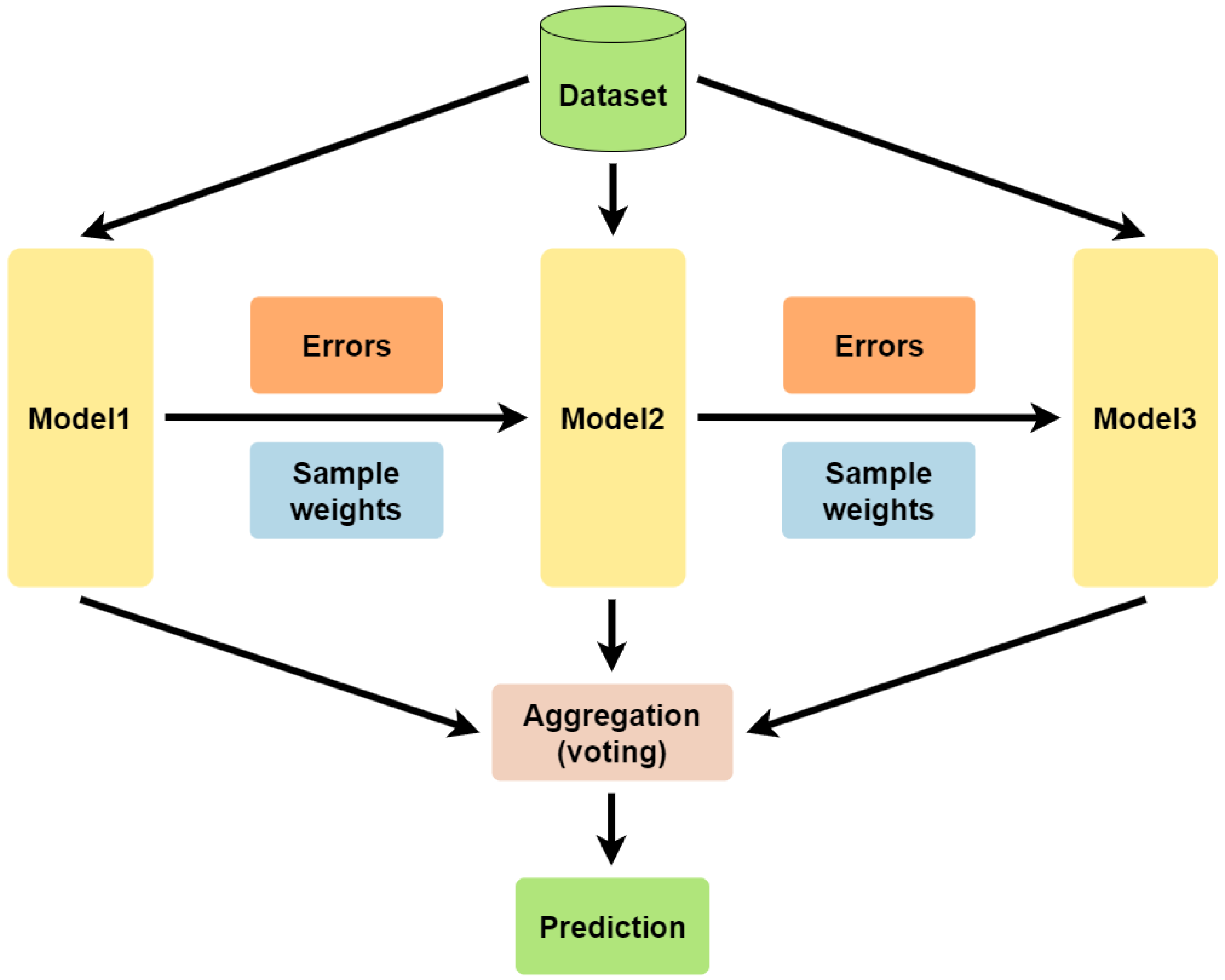
- **Boosting** methods sequentially add new models to the model set, so that each model added somehow "learns" from the weaknesses of already completed models.
- One of the earliest Boosting algorithms is Adaptive Boosting **(AdaBoost)**
  - When each model is trained, more weight is given to the example examples that previous models incorrectly predicted, so that the newly created model will perform better on those samples.

AdaBoost



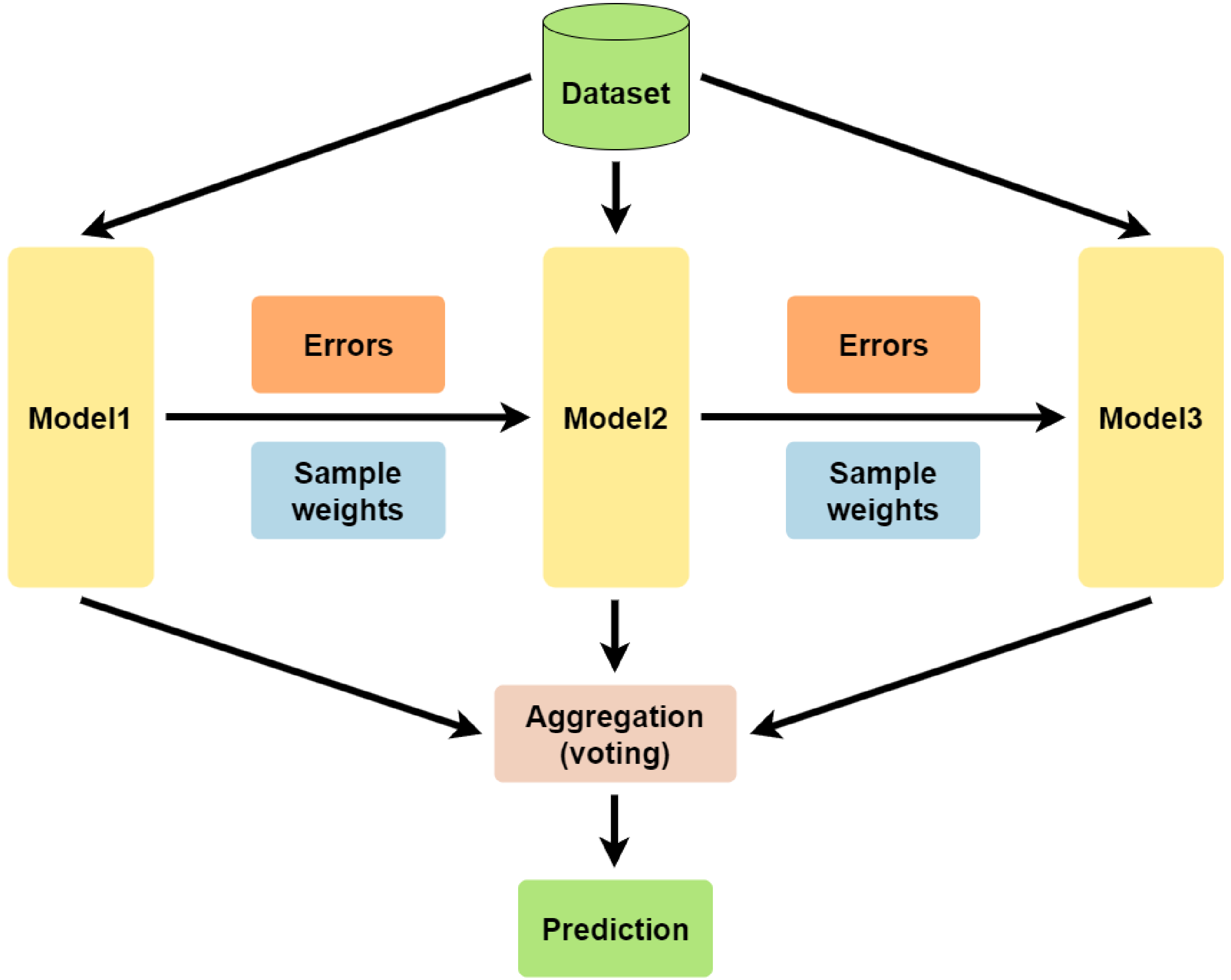
- 1. As a starting point, define a uniform weight value for each record in the data set.

AdaBoost



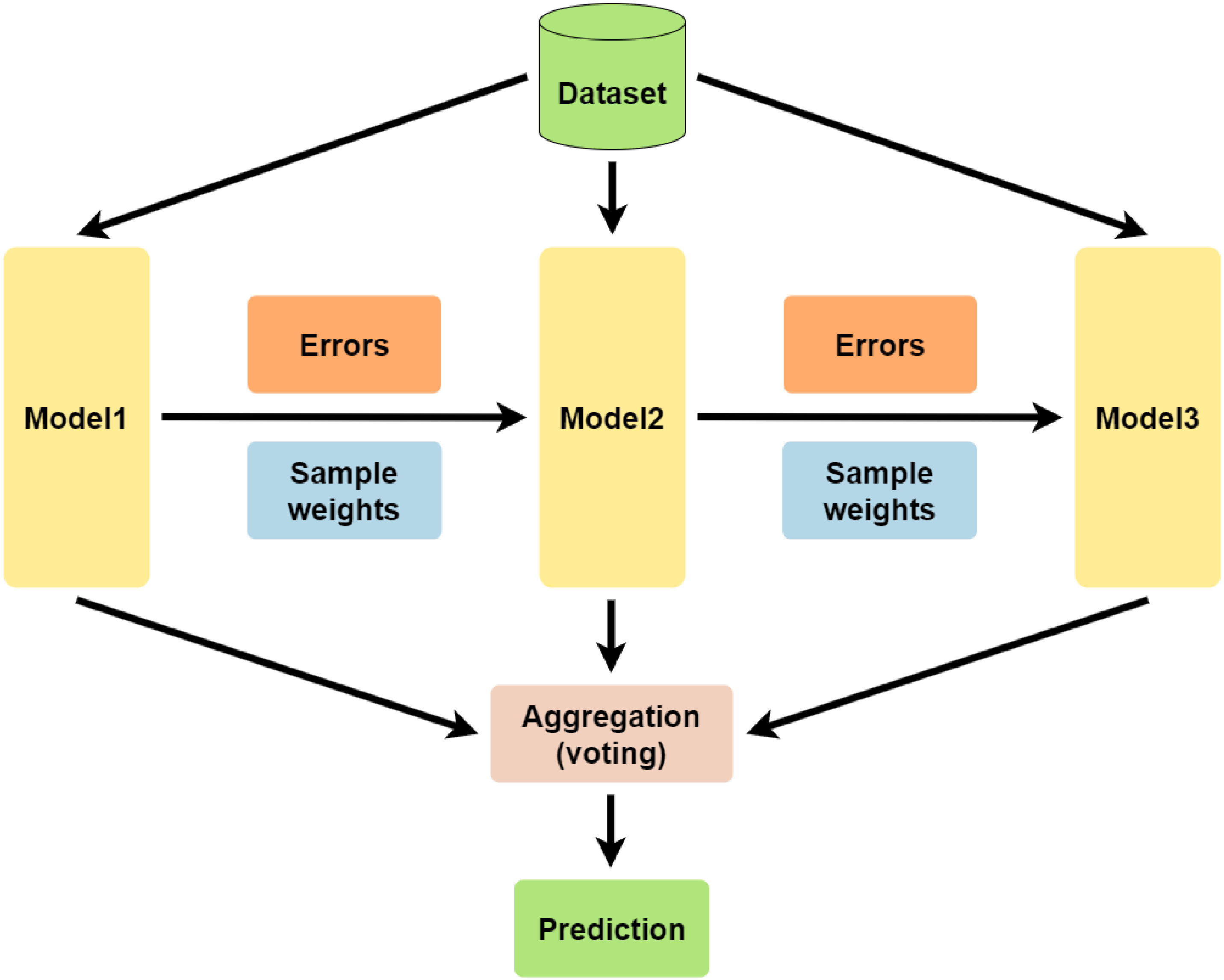
- 2. Train a model with instantaneous weights.
- 3. Increase the weight of samples for which the latest model gave an incorrect prediction, and then normalize the modified weight vector.

AdaBoost



- 4. Repeat from step 2 until a predetermined number of models is obtained.
- 5. The prediction of the model ensemble for new samples is obtained by aggregating the output of each model in a weighted manner.

AdaBoost



- During aggregation, the weight of each model increases the better predictive performance is given on the training set.

Boosting – Advantages and disadvantages

- Boosting methods iteratively fix errors in an existing set of models
- + accelerating process convergence
- + To achieve similarly good predictive performance, fewer models are sufficient than compared to the Bagging algorithm
- - The training of each model depends on the error of the previous model
- - Models cannot be created in parallel, but can only be done sequentially.

Stacking

- Assumption so far: all members of the model ensemble come from the same model family (e.g. decision tree, SVM, etc.).
- In general, a model ensemble can consist of classifiers (or regressors) of any type
  - Condition: they offer a solution to the same classification (or regression) problem.
- The common case where a model set consists of estimators from different model families is called **stacking**.

Stacking

- Model diversity can be an advantage
- Predictions of models with different approaches can complement each other well
- Each model in the model ensemble typically errs on different inputs, so a simple majority decision they make is more reliable than prediction for individual models.
- This is not necessarily true for all classification (or even regression) problems

Stacking

- Individual consideration of different approaches remains important
- In case most models perform particularly poorly on the dataset, then the performance of the model ensemble may be worse than some individual models.