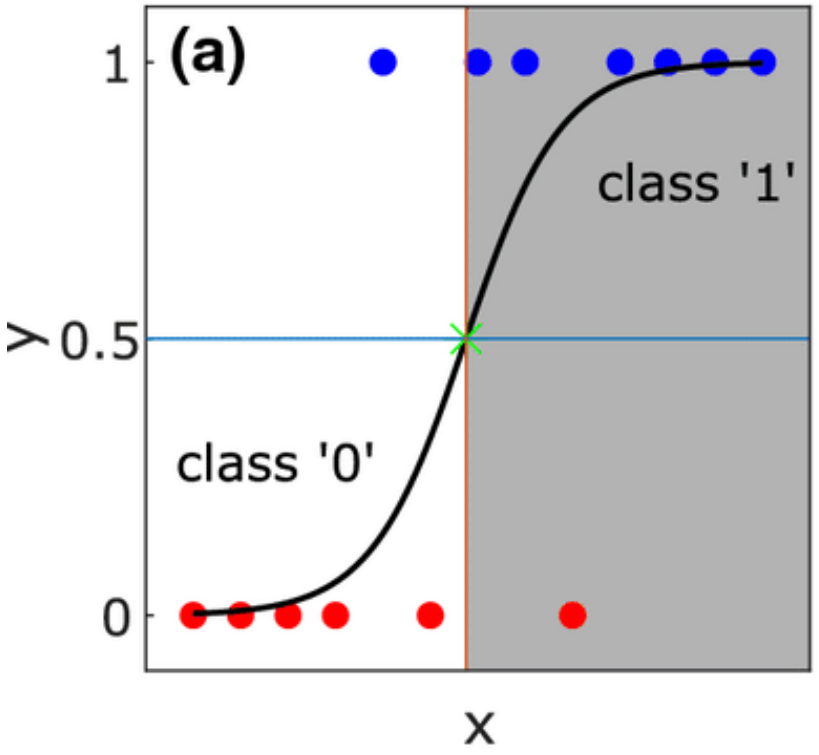
Computer Vision

sec4

Logistic Regression

* Logistic regression is a statistical model used for classification tasks, designed to predict the probability that a given input belongs to a particular class. It can handle both **binary classification** (where there are two possible outcomes) and **multi-class classification** (where there are more than two classes).

* The model estimates probabilities using a logistic (sigmoid) function for binary cases, or a softmax function for multi-class cases



important Parameters:

- C: Inverse of regularization strength (smaller values specify stronger regularization).

- solver: Algorithm to use in the optimization problem
  (e.g., 'lbfgs', 'liblinear', 'newton-cg',’saga’,’sag’).

- max_iter: Maximum number of iterations for the solver.

- Penalty: type of regularization (l1,l2)

Decision Tree

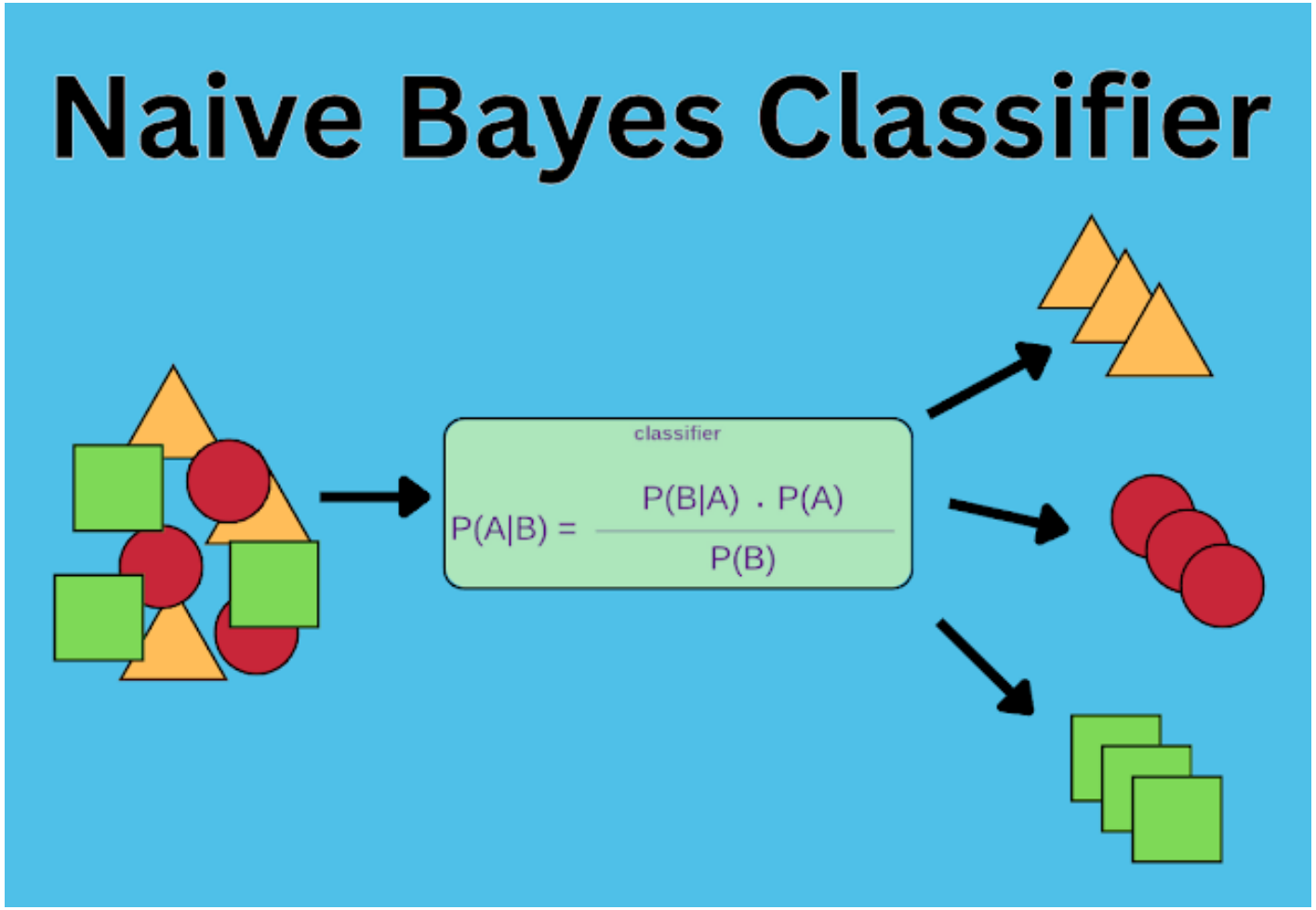
Decision Tree is a supervised learning method used for classification. They split the data into subsets based on the value of input features, creating a tree-like model of decisions.

important Parameters:

- criterion: Function to measure the quality of a split ('gini' , 'entropy').
- max_depth: Maximum depth of the tree (limits the number of splits).

Naïve Bayes

Naive Bayes is a probabilistic algorithms based on Bayes' theorem, which is used for classification tasks.



Important parameters

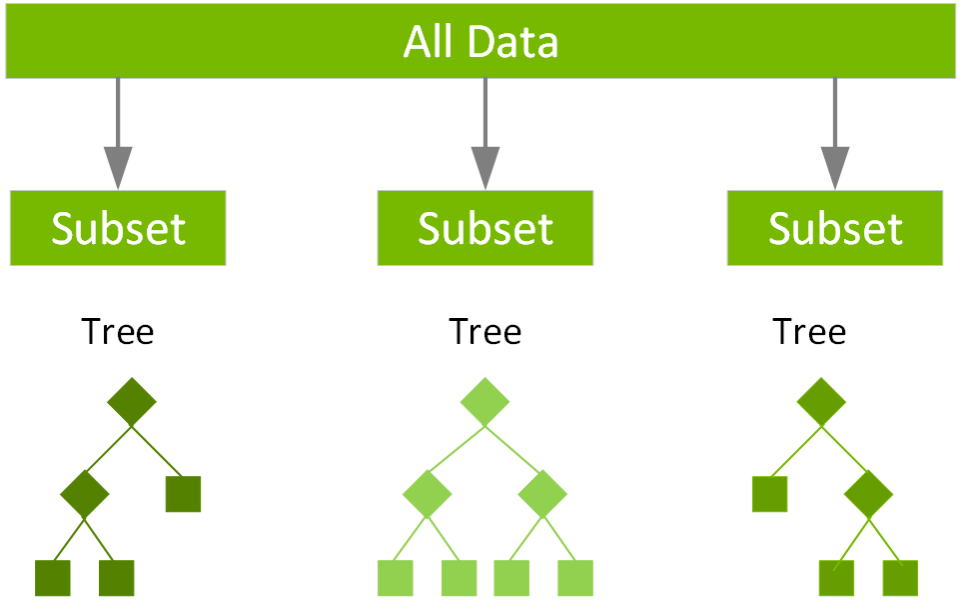
- var_smoothing: A small constant added to feature variances to ensure numerical stability during calculations. Helps avoid issues when variances are too small or zero.

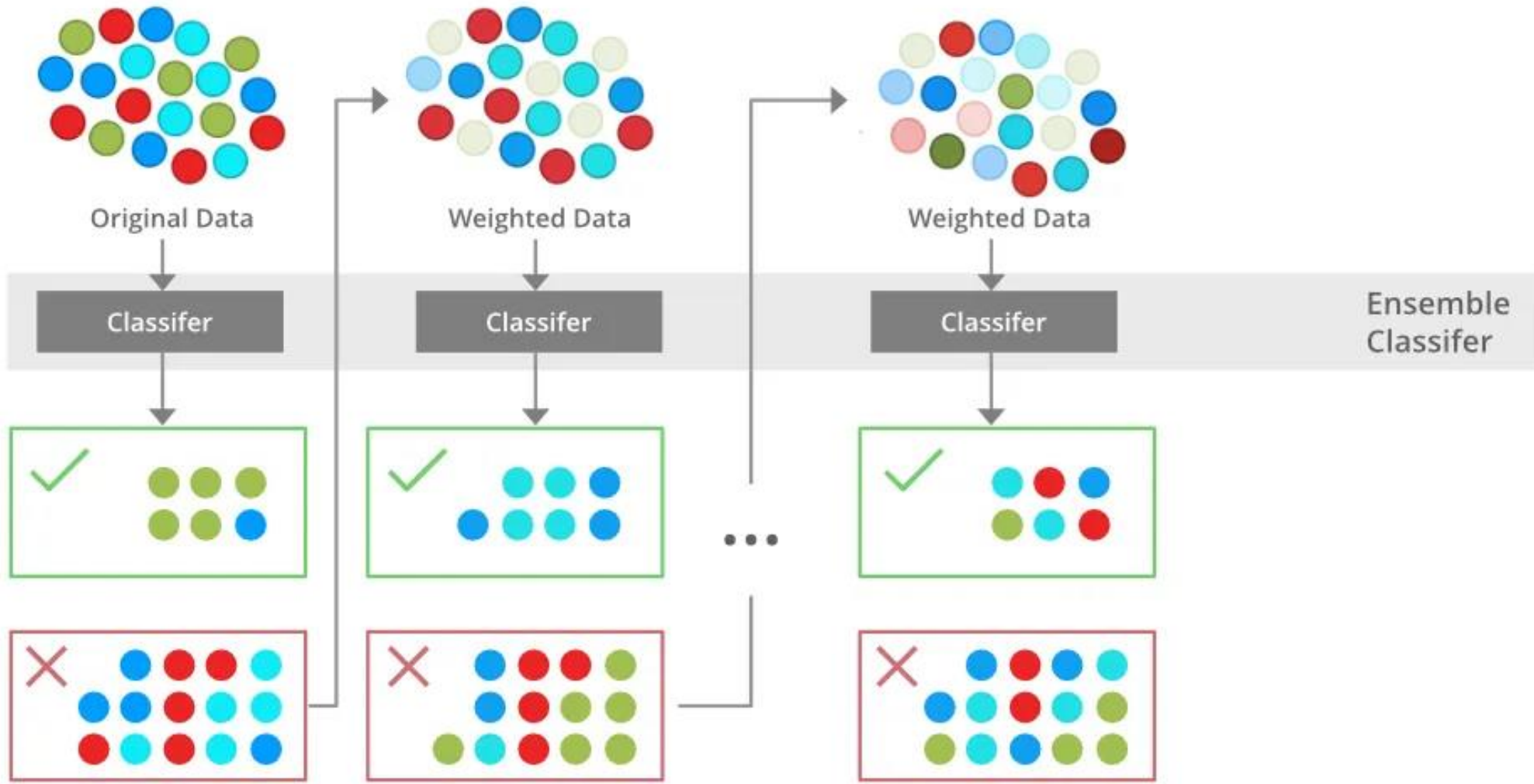
XGBoost

XGBoost (Extreme Gradient Boosting) is a powerful and efficient open-source machine learning library designed for gradient boosting

XGBoost implements the gradient boosting algorithm, which combines the predictions of multiple weak learners (typically decision trees) to create a strong predictive model.

It builds trees sequentially, with each new tree aiming to correct the errors made by the previously built trees.



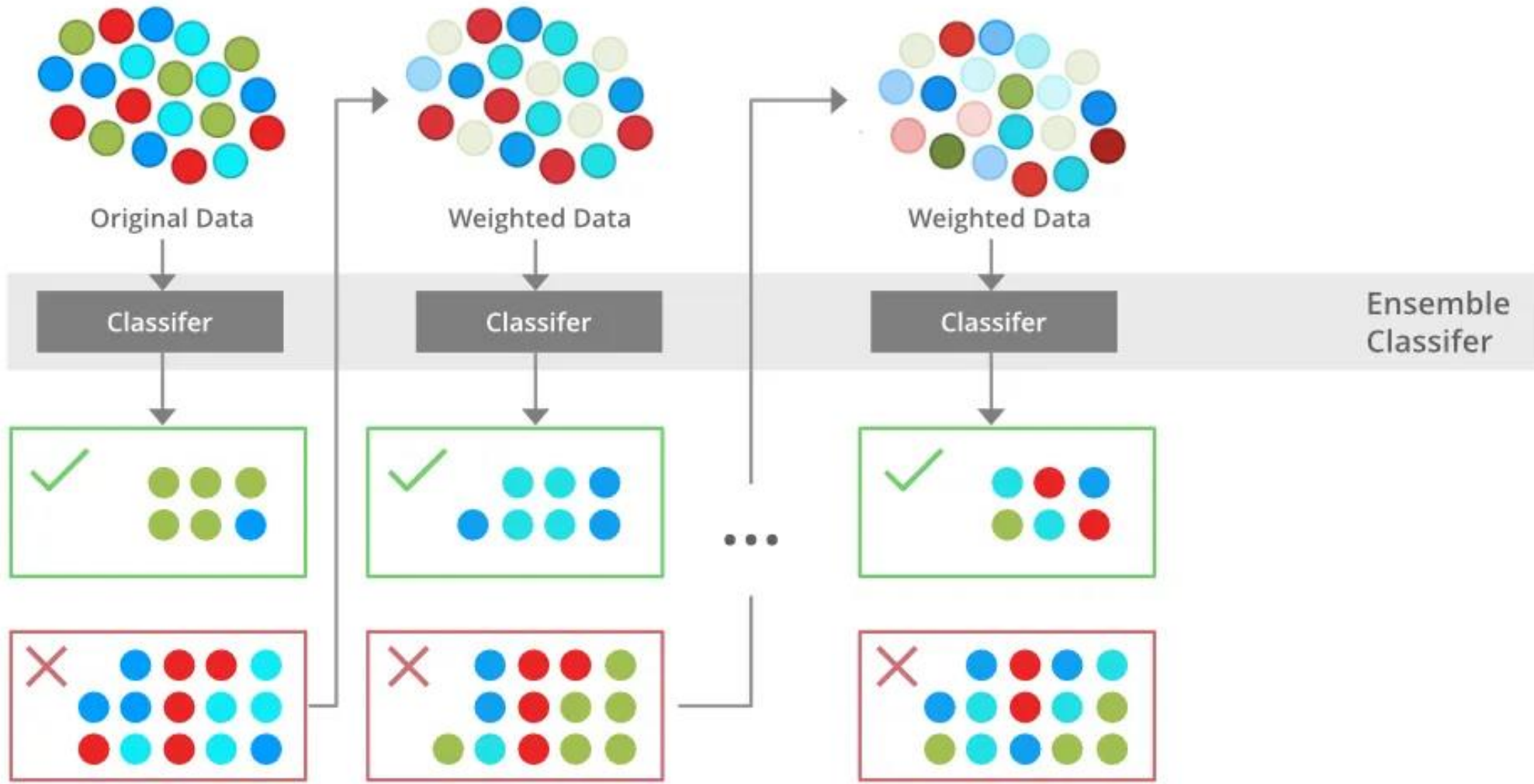

Original Data
Weighted Data
Weighted Data
Classifer
Classifer
Classifer
Ensemble
Classifer

Important parameters

- **n_estimators**: The number of boosting rounds (trees to build)
- **learning_rate**: used to prevent overfitting. A lower value makes the model more robust
- **max_depth**:maximum depth of the individual trees
- **eval_metric**:The default evaluation metric varies based on the task (e.g., logloss for binary classification, mlogloss for multi-class).

CatBoost

- Boosting algorithm
- It builds trees sequentially, with each tree correcting the errors of the previous ones, similar to other boosting algorithms like XGBoost
- It uses a technique called **order-based encoding** that allows it to efficiently incorporate categorical data into the model.
- **Output**: Final prediction is made through a aggregation


Original Data
Weighted Data
Weighted Data
Classifer
Classifer
Classifer
Ensemble
Classifer

Important parameters

- **Iterations**:The total number of trees
- **learning_rate**:Lower values can lead to better generalization
- **Depth:**maximum tree depth

Adaboost

- Boosting algorithm
- Combines multiple weak learners (often decision trees) in a sequential manner. Each weak learner is trained to correct the errors made by the previous ones, focusing on misclassified examples by adjusting their weights.
- **Output**: Final prediction is made through a majority vote

Important parameters

- **Estimator**: base estimator
- **n_estimators**: number of trees
- **learning_rate**:A lower value makes the learning process more robust
- **Algorithm**:SAMME, SAMME.R

LightGBM

- is an open-source, distributed gradient boosting framework developed by Microsoft. It is designed to be efficient and scalable for machine learning tasks, particularly for large datasets.

Important parameters

- **boosting_type**:gbdt, dart, goss
- **num_leaves**:maximum number of leaves in one tree
- **max_depth**:maximum depth of the tree
- **learning_rate**: A lower value makes the learning process more robust
- **n_estimators**:number of trees
- **Objective**:binary, multiclass