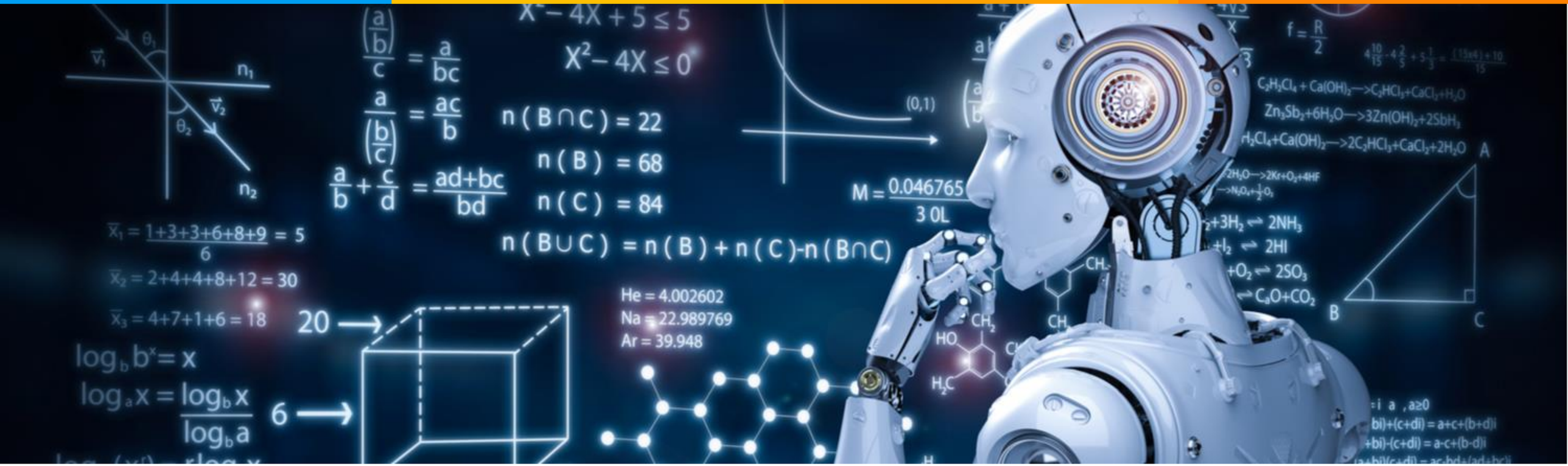


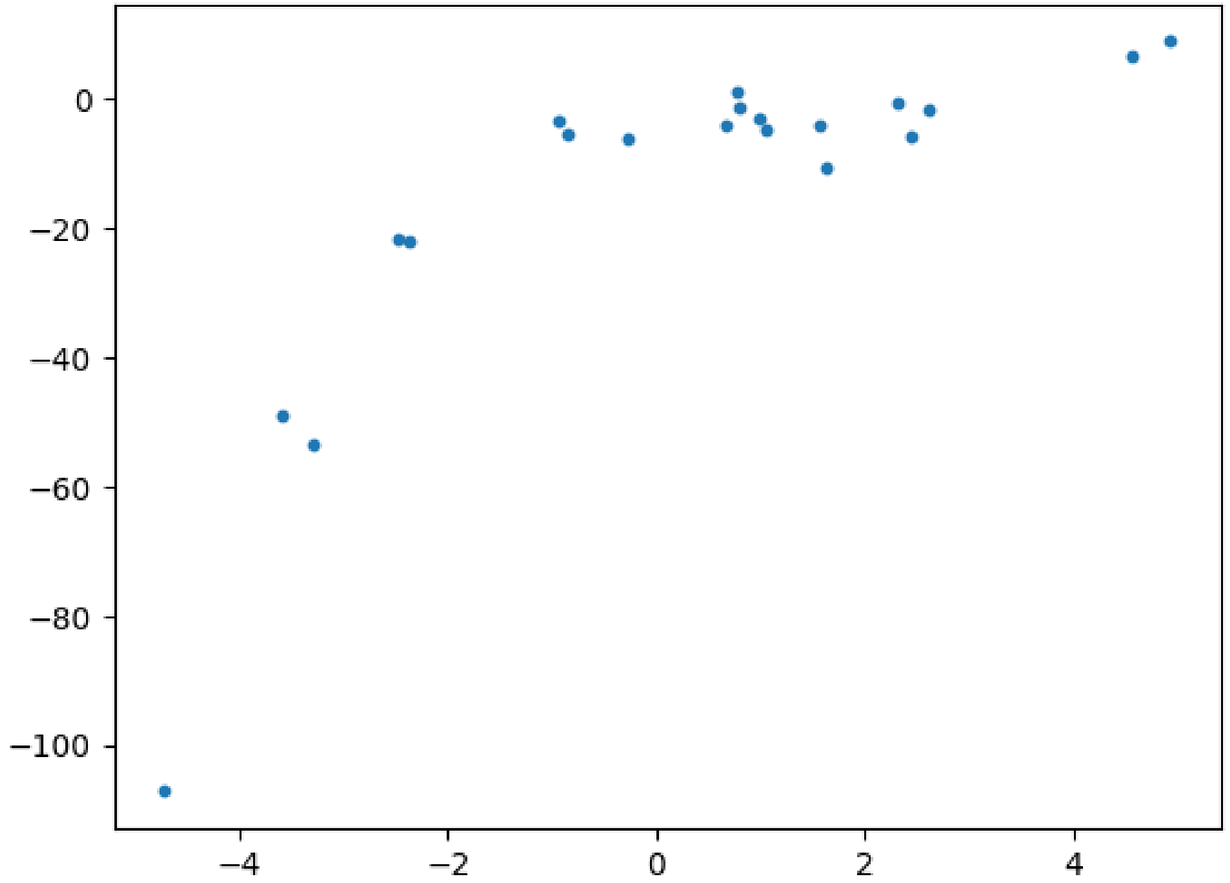
Lecture 19: Ensemble Learning

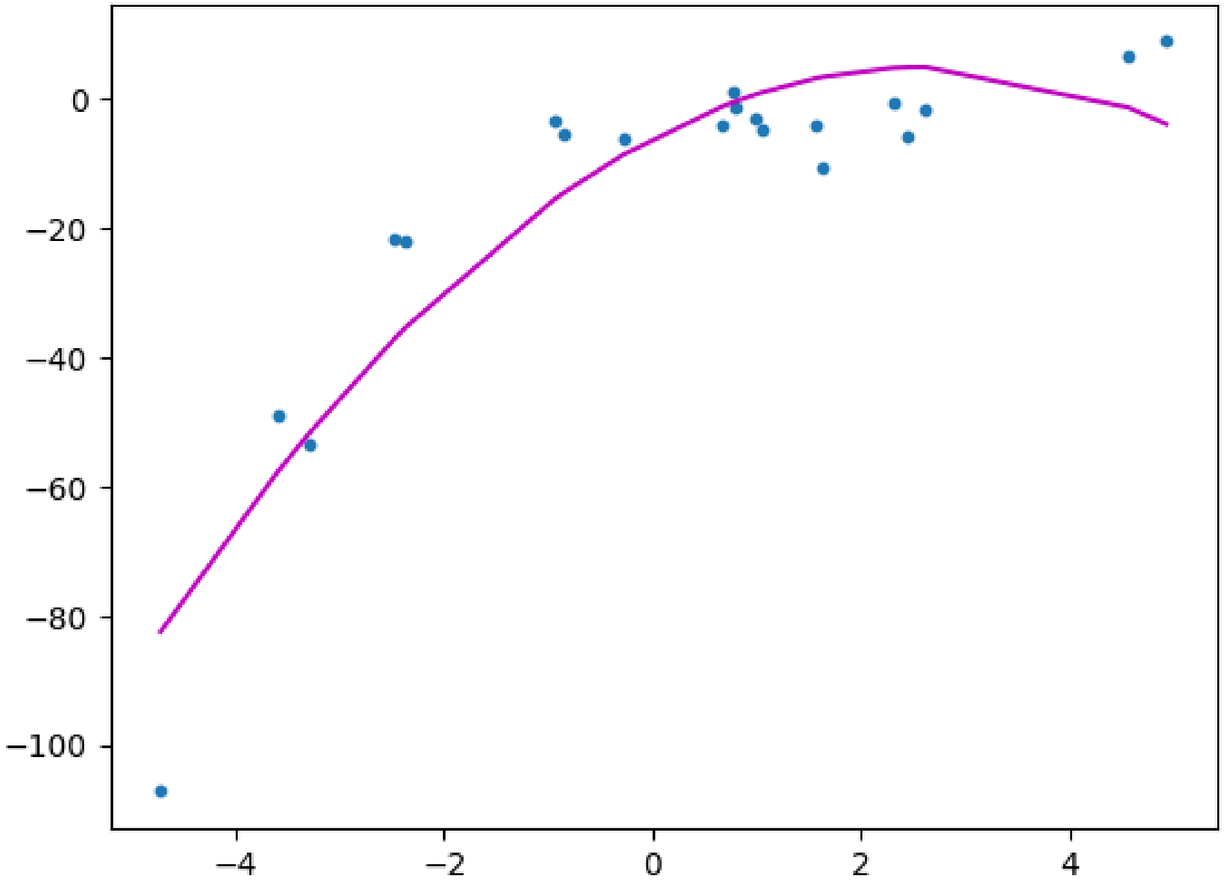
Recap

- Decision Tree
- Gini Impurity
- DT Pruning

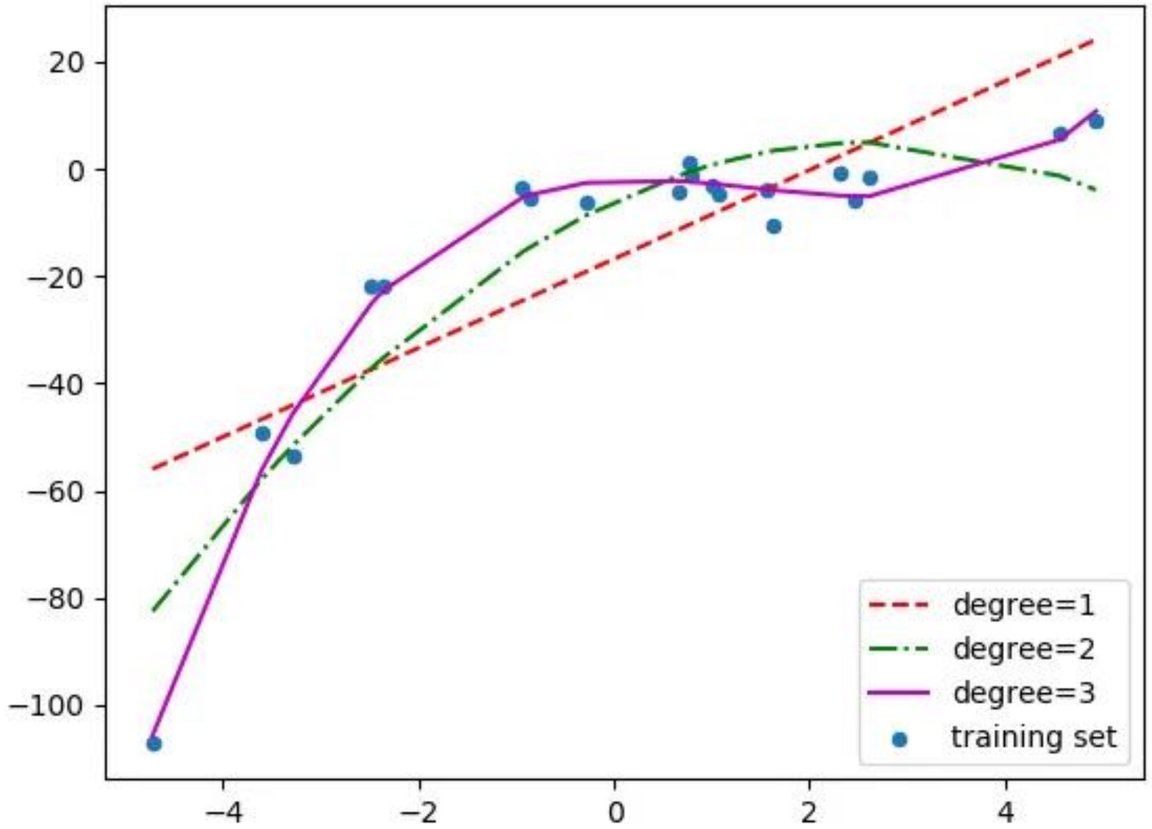
Underfitting, Overfitting, Bias Variance trade off

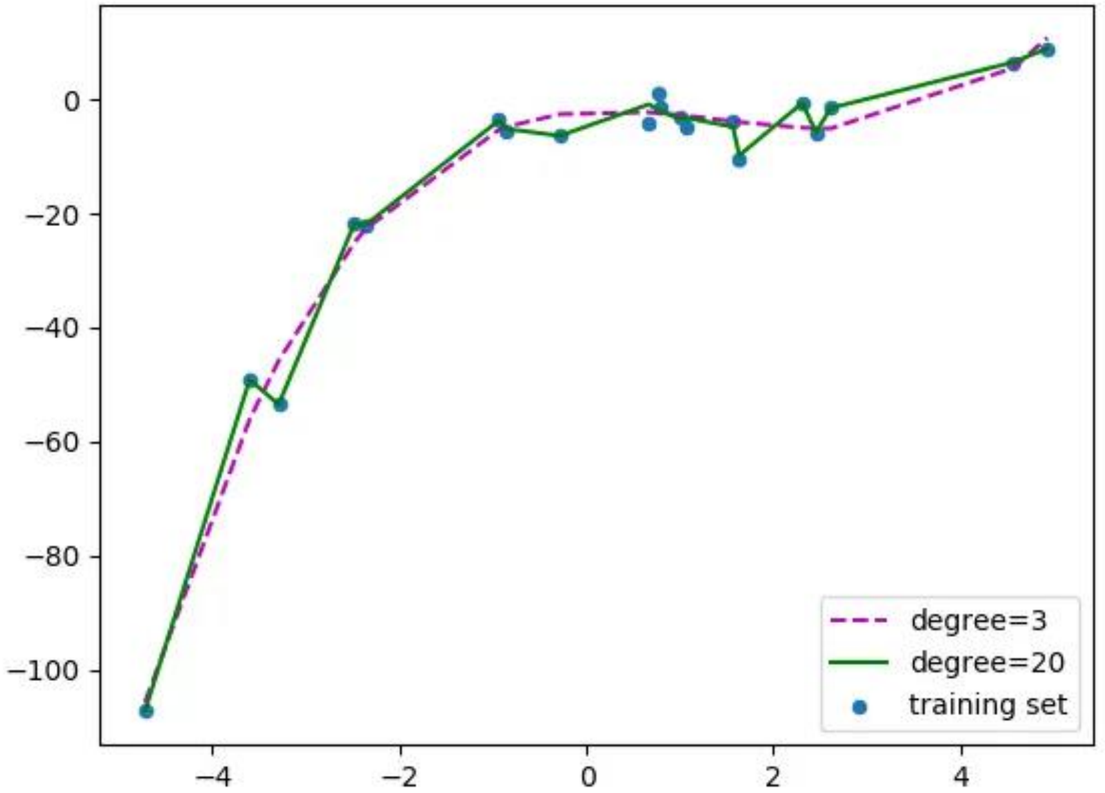
Fitting a curve



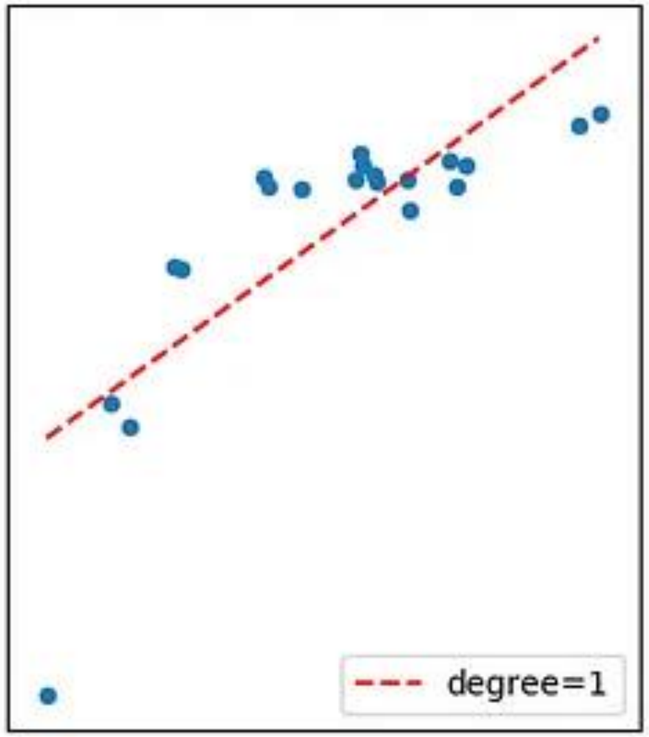


How many degrees is good?

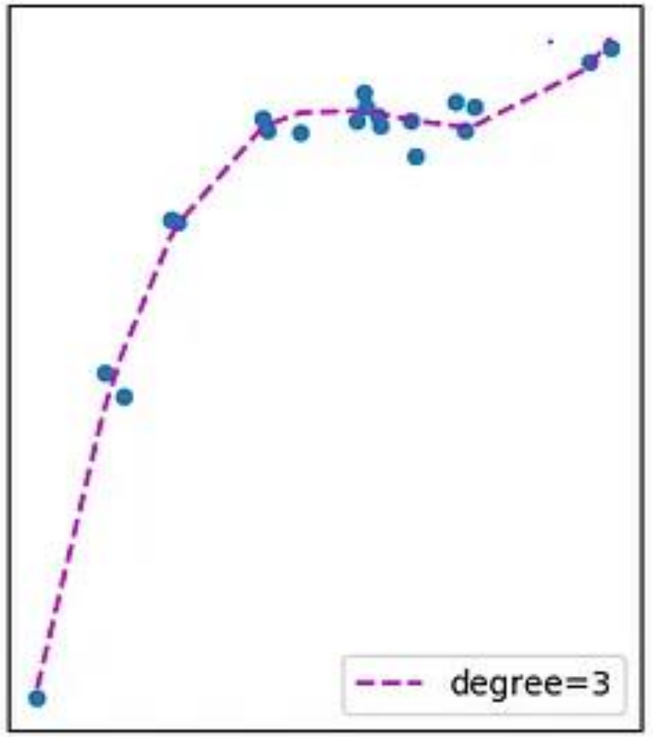




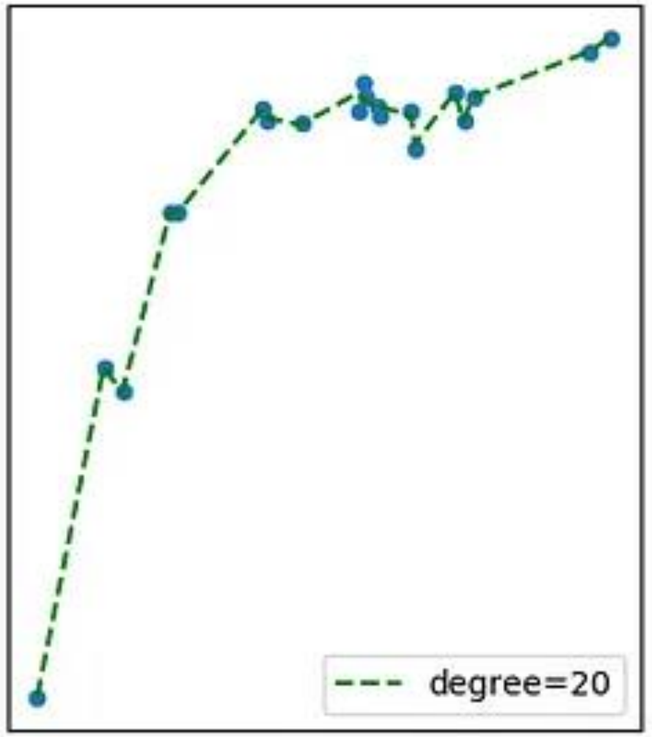
Underfitting & overfitting



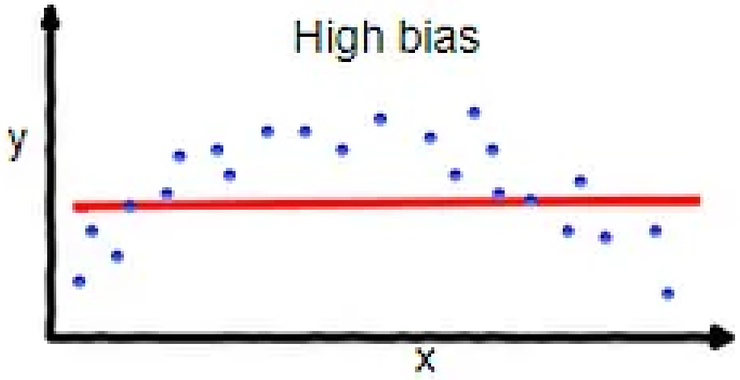
Underfit
High Bias
Low Variance



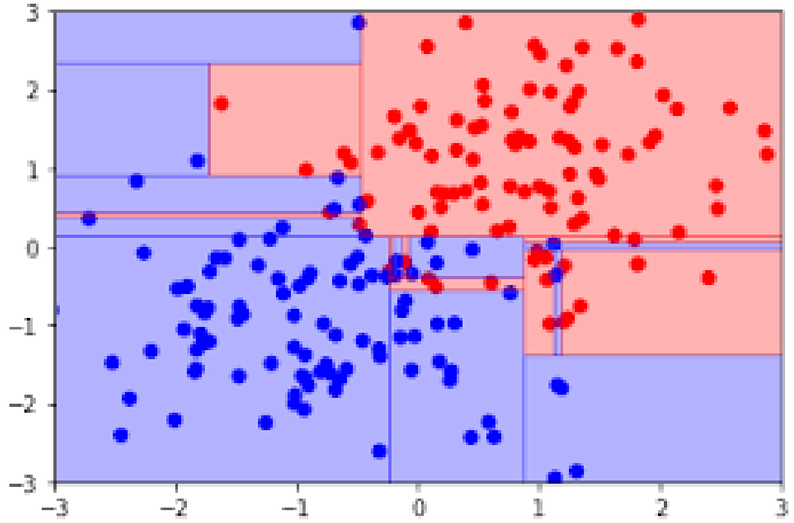
Correct Fit
Low Bias
Low Variance



Overfit
Low Bias
High Variance



Add some new data points, remove some. What happens in each case?



Think Linear Regression

Think Decision Tree

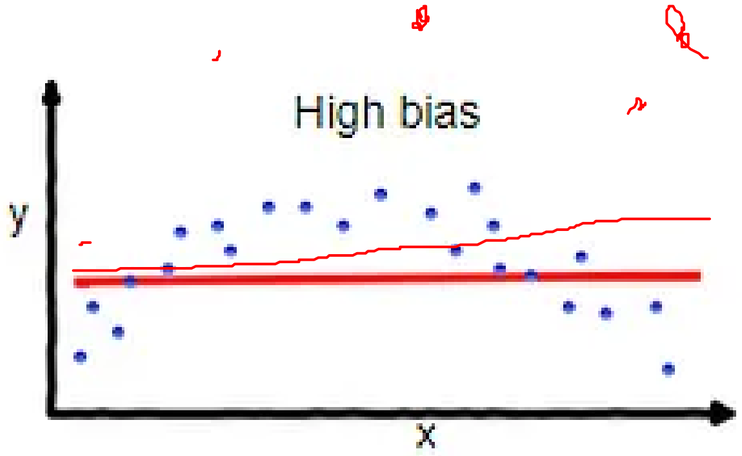
Bias occurs when an algo has *limited flexibility* to learn the true signal from a dataset.

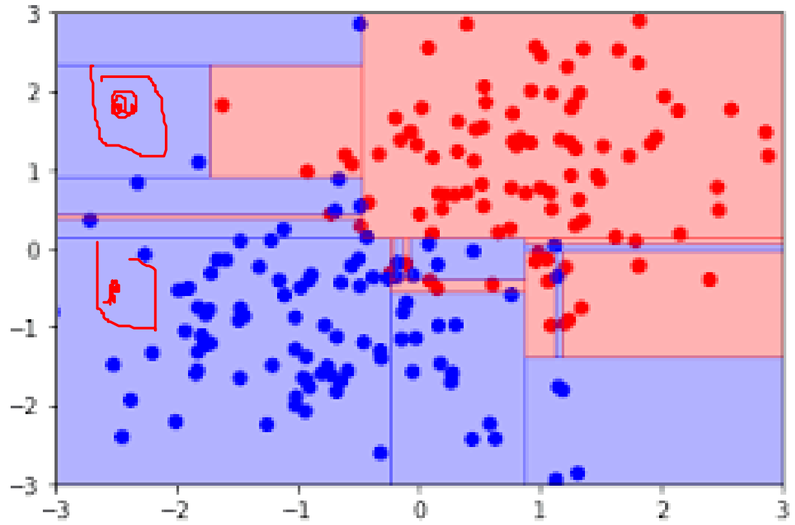
Variance refers to an algo's *sensitivity* to specific sets of training data.

- 5 different training sets (imagine bootstrapping)
- Same algorithm trained on 5 data sets

High bias, low variance algorithms train models that are consistent, but inaccurate *on average*.

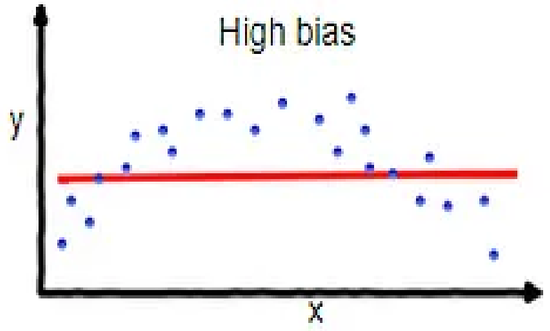
High variance, low bias algorithms train models that are accurate *on average*, but inconsistent.





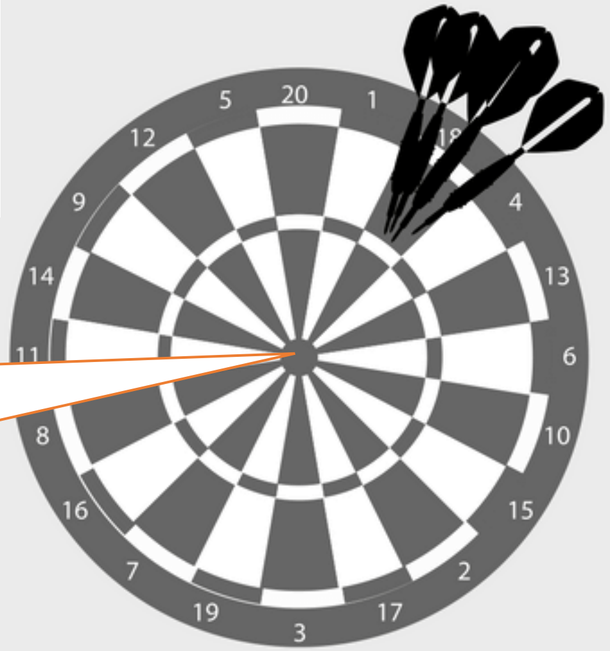
High bias, low variance algorithms train models that are consistent, but inaccurate *on average*.

High variance, low bias algorithms train models that are accurate *on average*, but inconsistent.



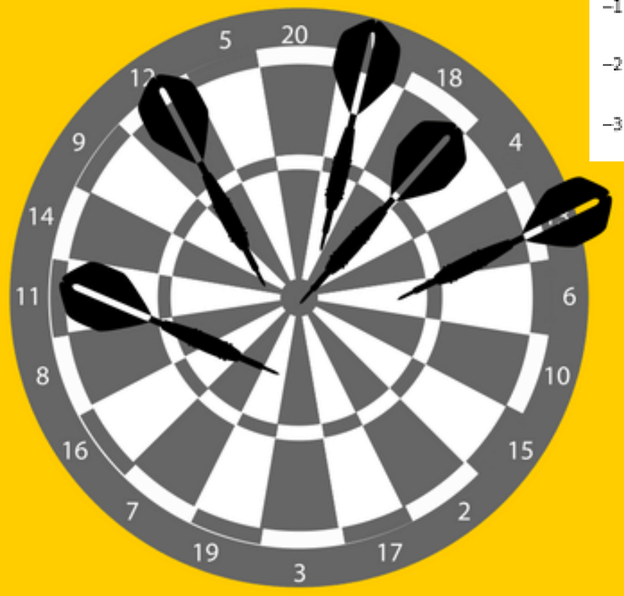
High Bias

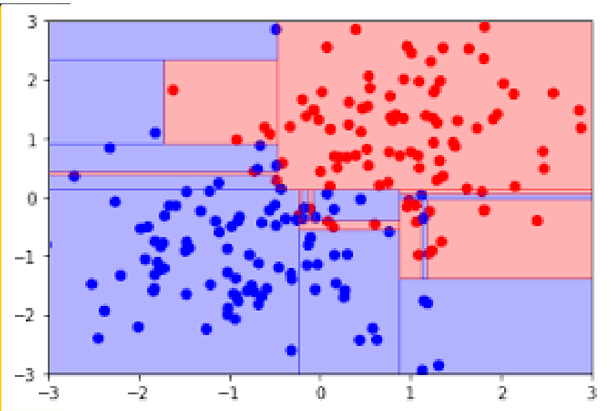
Low Variance



High Variance

Low Bias





Bulls eye means best model & consistent model

But why is there a tradeoff?

Low variance algos tend to be **less complex**, with simple or rigid underlying structure.

- e.g. Regression
- e.g. Naive Bayes
- *Linear algos*
- *Parametric algos*

Low bias algos tend to be **more complex**, with flexible underlying structure.

- e.g. Decision trees
- e.g. Nearest neighbors
- *Non-linear algos*
- *Non-parametric algos*

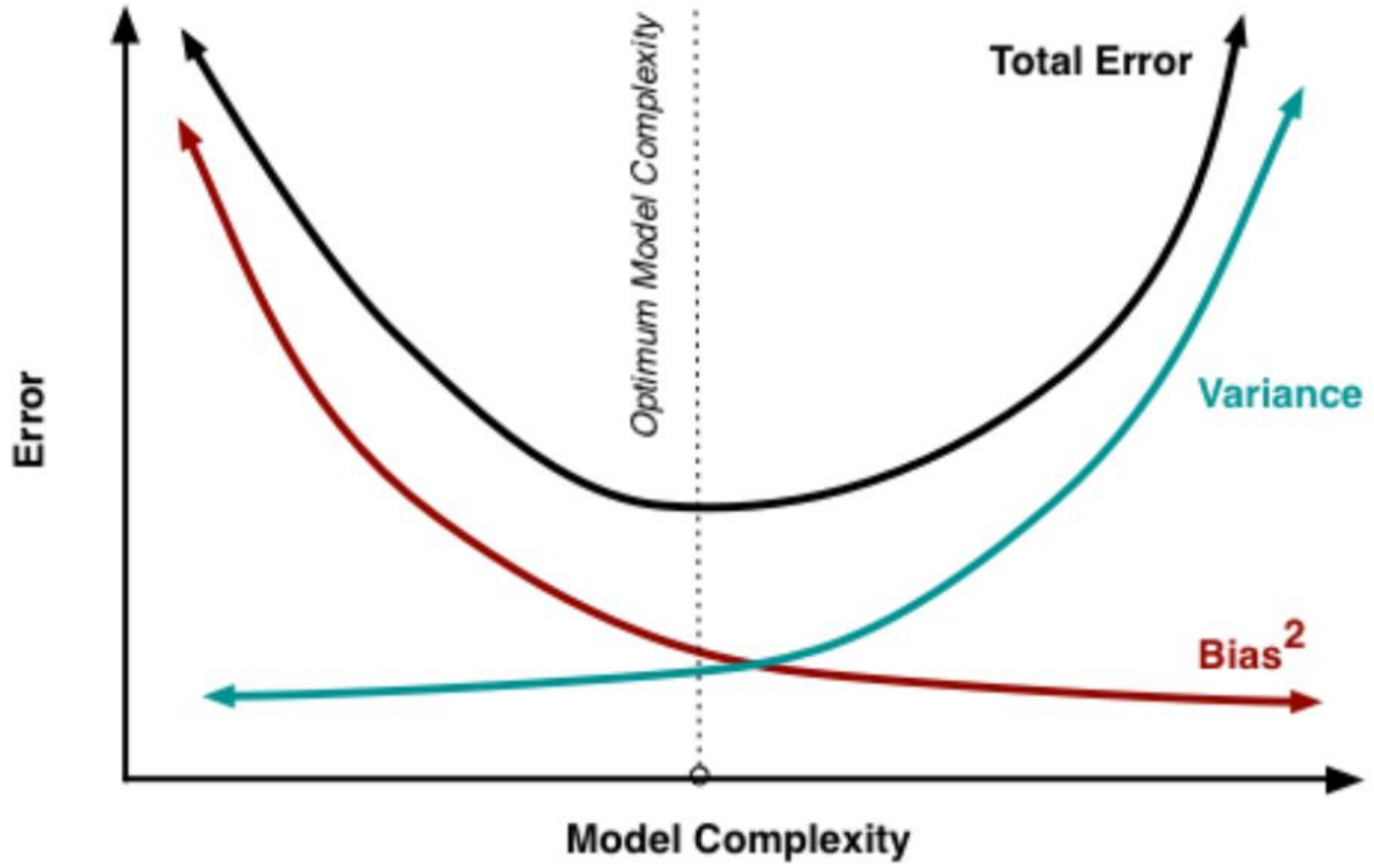
Within each algo family, there's a tradeoff too...

For example, regression can be **regularized** to further reduce complexity.

For example, decision trees can be **pruned** to reduce complexity.

Bias Variance Plot

Total Error = Bias^2 + Variance + Irreducible Error



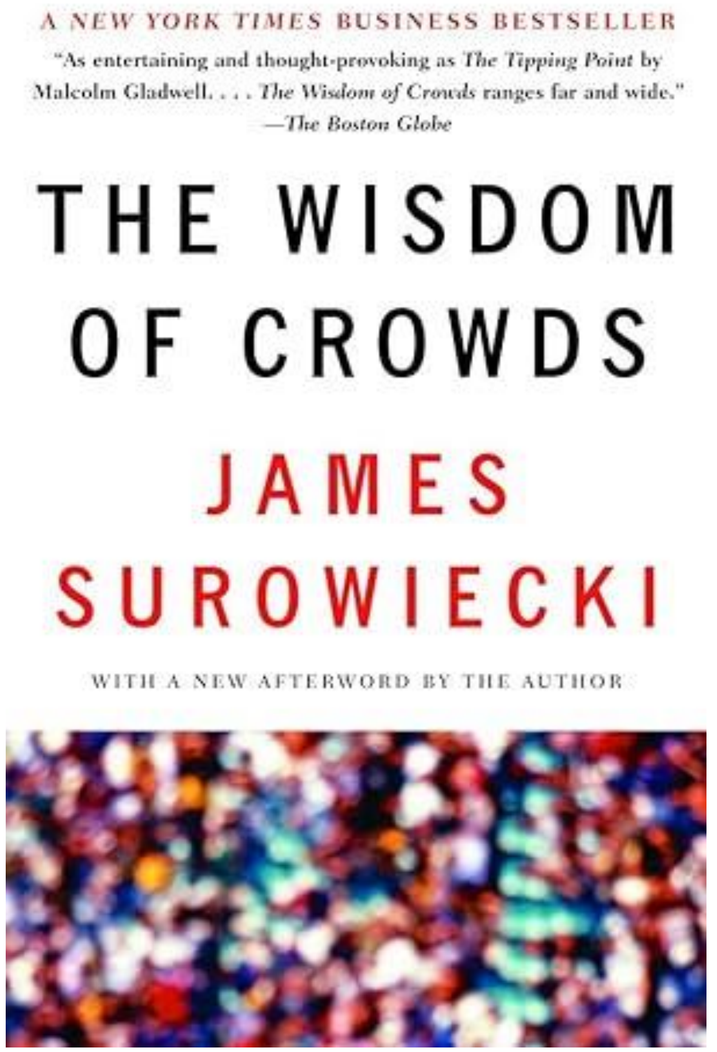
Wisdom of the Crowds

A game of guessing?

- How many candies in the jar?
- How to get the most correct answer?
- No answer is correct, some answers are useful
- Especially if averaged over a group

The wisdom of crowds

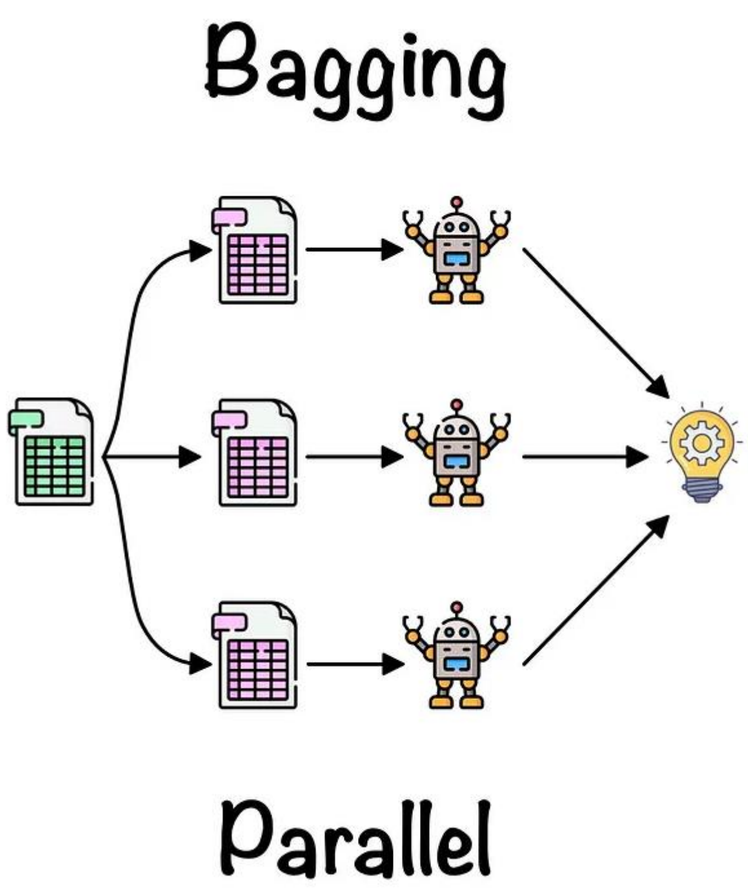
- "Under the right circumstances, groups are remarkably intelligent, and are often smarter than the smartest people in them"
- Using a Group for better decision
  - Diverse background
  - Independent decision by each individual
  - Good method for aggregation

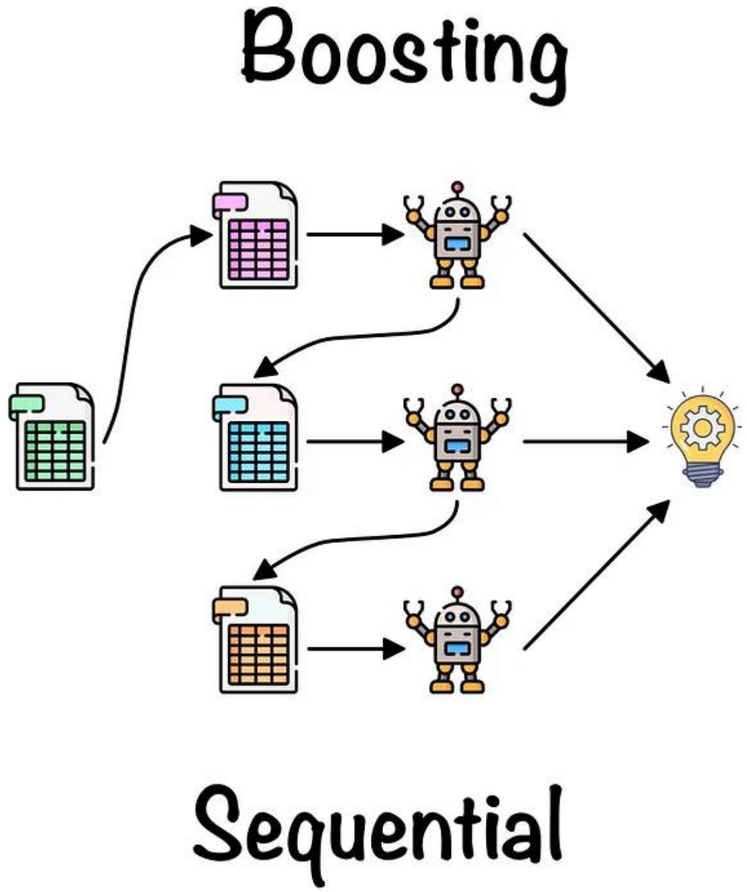


Ensemble Learning

Ensemble Learning

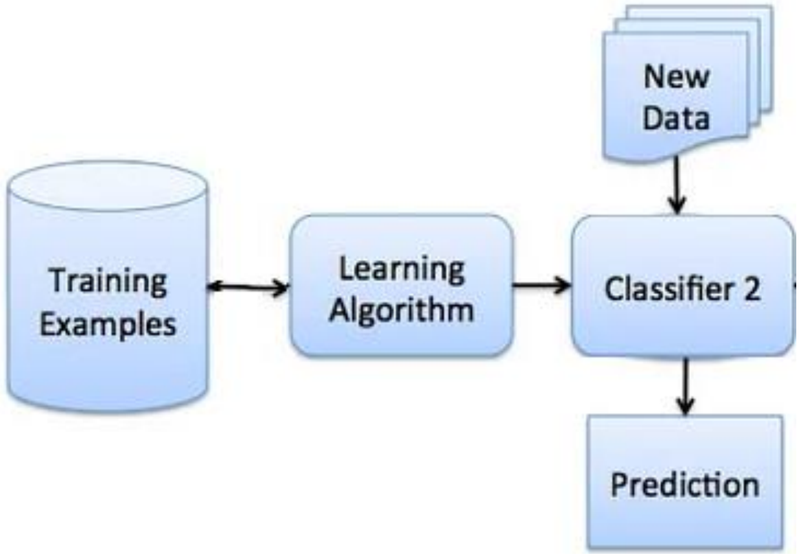
- Multiple ML model used together for prediction

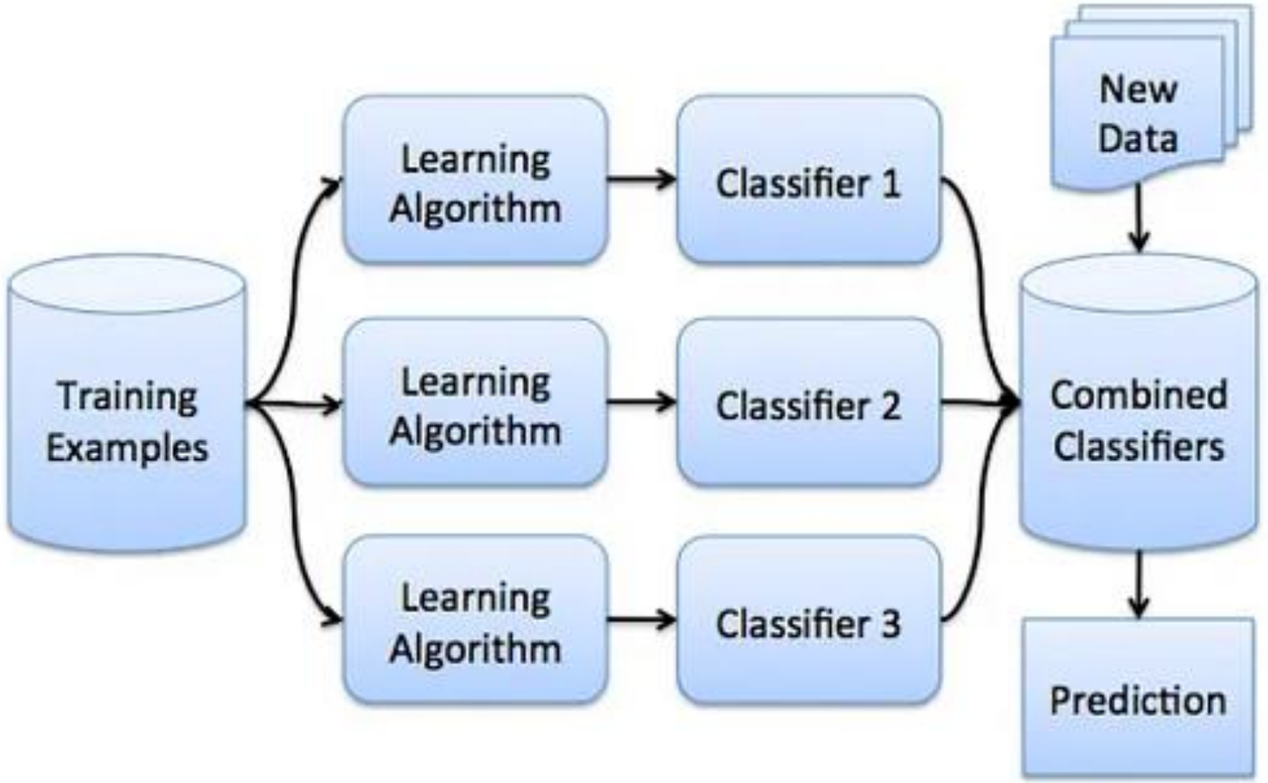




Majority Voting/Averaging models

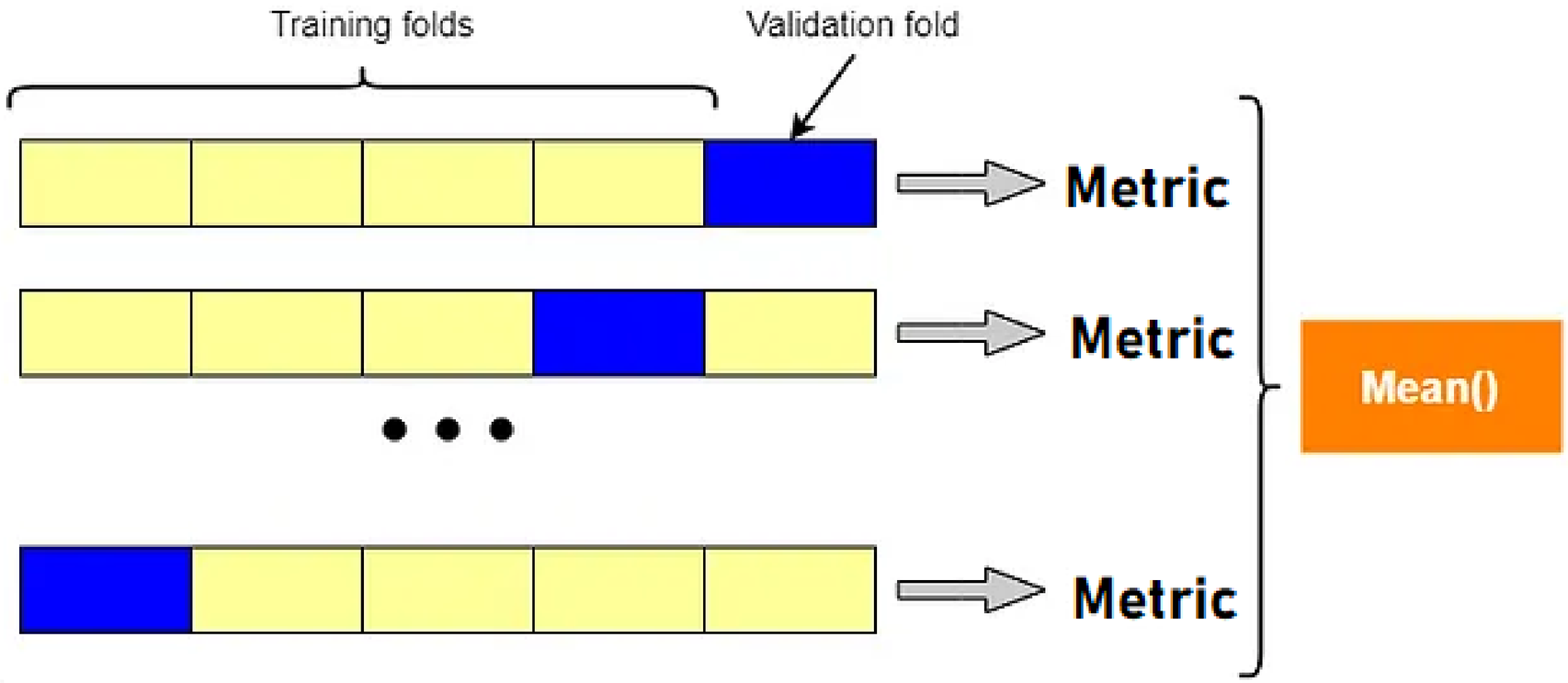
- High variance from single model
- How about using multiple models?
  - Will it reduce variance?
  - Not if data is correlated





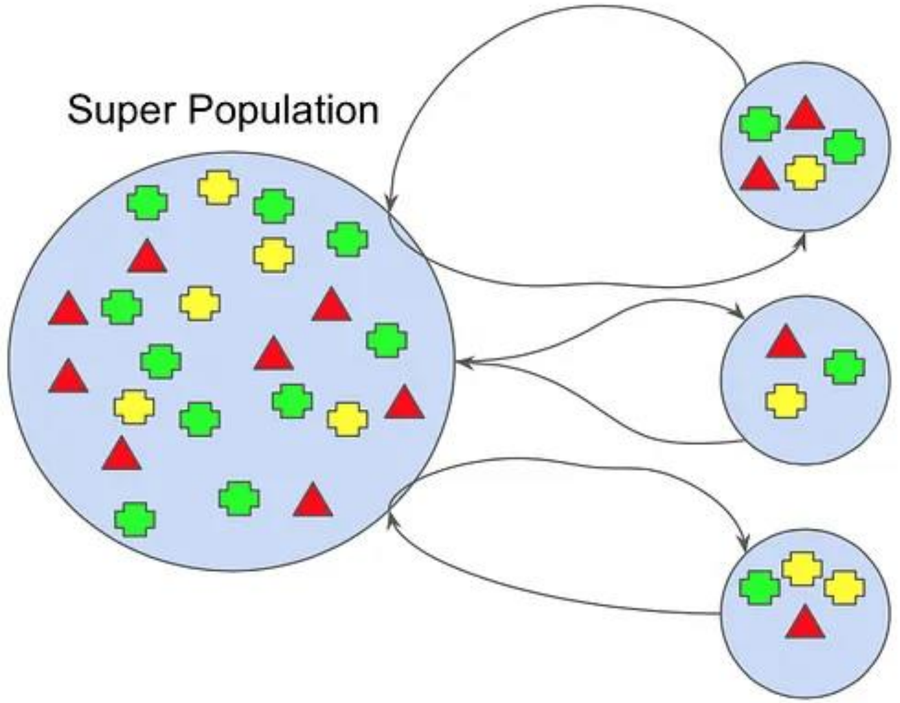
A second look at K-Fold CV

- Most data is repeated across folds
- IID from one record to next.
- Highly correlated from one fold to another



Solution: Bagging

- Bagging = Bootstrapping + Aggregation
- Bootstrapping: Sampling with Replacement



Data becomes uncorrelated

Sample Population 1

Std(a)

Sample Population 2

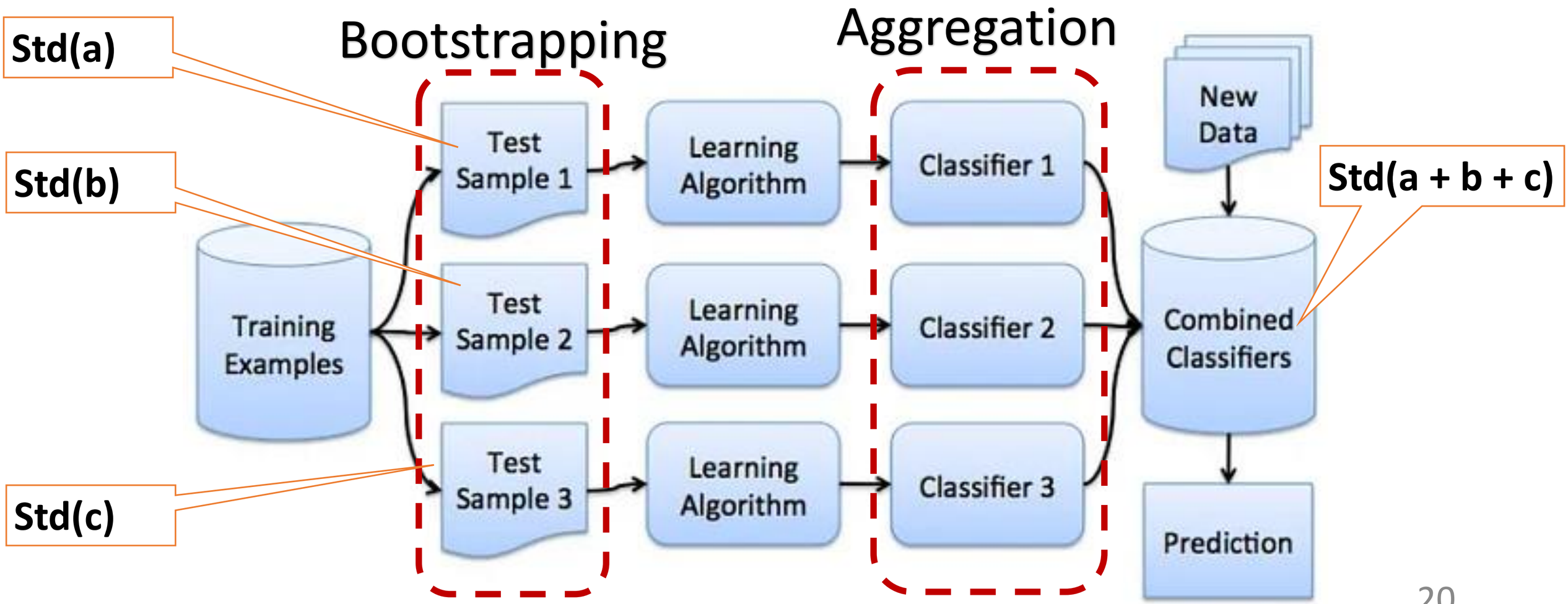
Std(b)

Sample Population 3

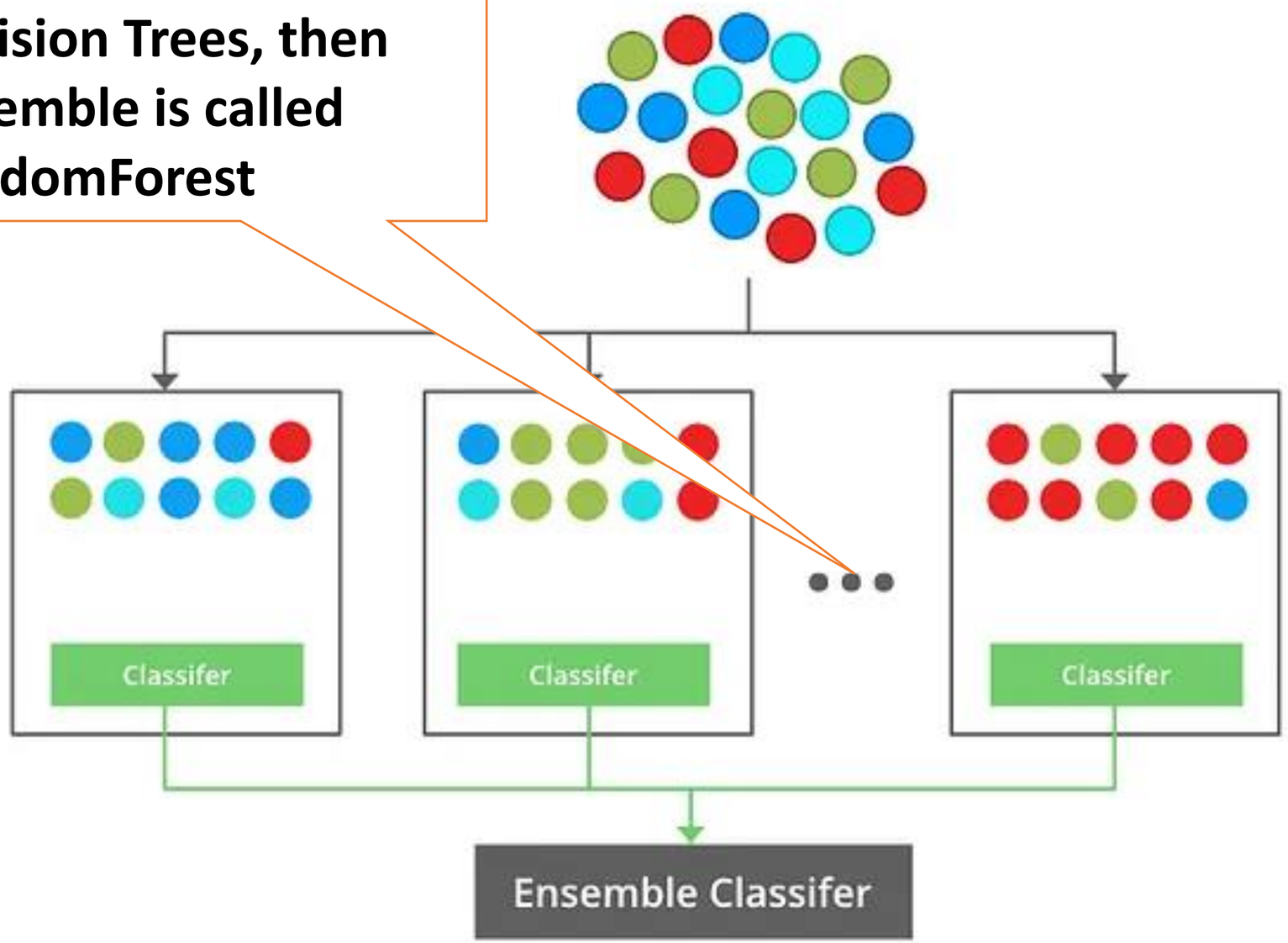
Std(c)

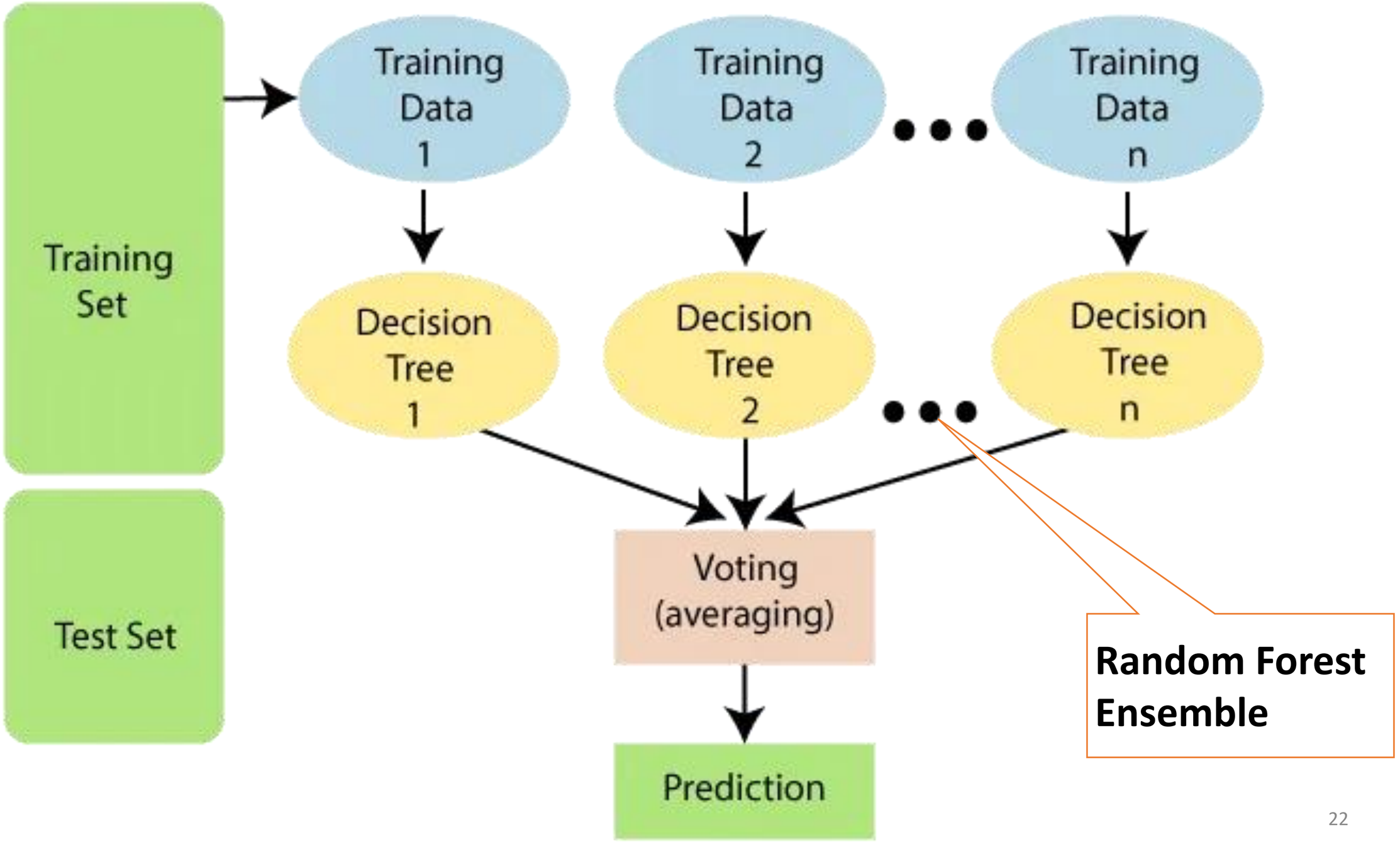
Solution: Bagging (Contd.)

- Aggregation = Combining classifiers



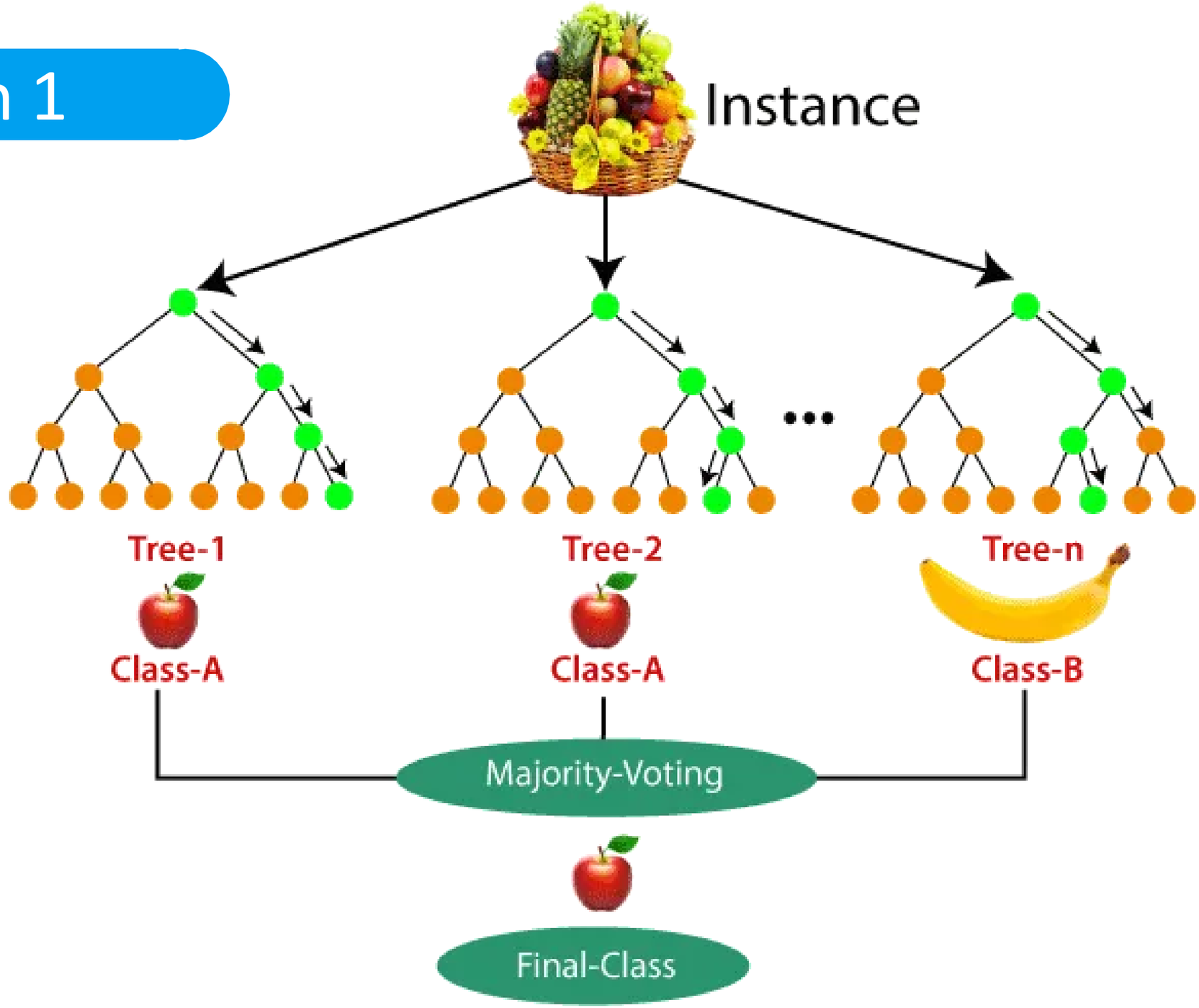
When all classifiers are Decision Trees, then ensemble is called RandomForest



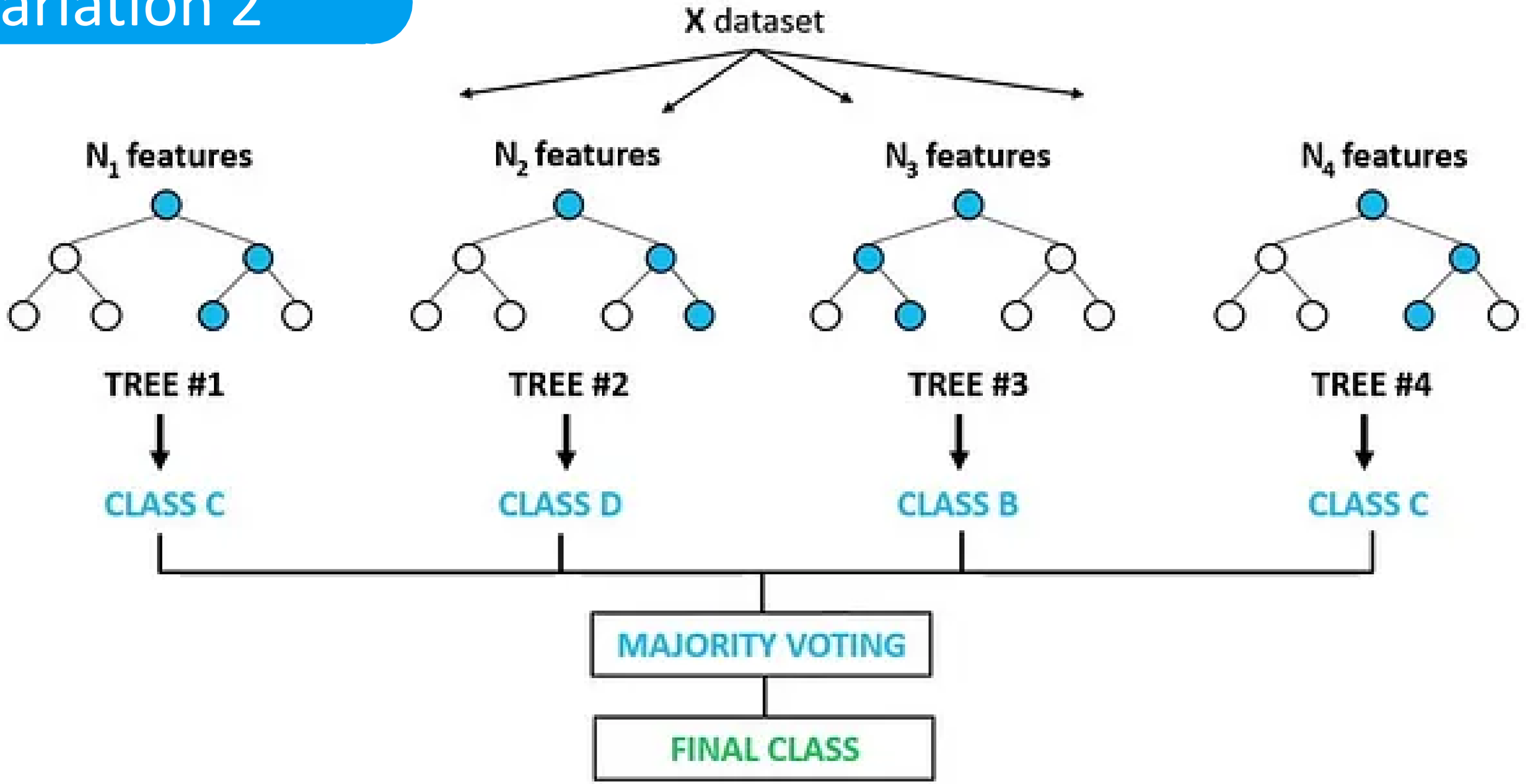

Training Set
Test Set
Training Data 1
Training Data 2
Training Data n
Decision Tree 1
Decision Tree 2
Decision Tree n
Voting (averaging)
Prediction
Random Forest Ensemble

Random Forest

Variation 1


Instance
Tree-1
Tree-2
Tree-n
Class-A
Class-A
Class-B
Majority-Voting
Final-Class

Variation 2


X dataset
N1 features
N2 features
N3 features
N4 features
TREE #1
TREE #2
TREE #3
TREE #4
CLASS C
CLASS D
CLASS B
CLASS C
MAJORITY VOTING
FINAL CLASS

Random Forest hyperparameters

- Decision Tree hyper params
  - Criteria = gini/entropy
  - Max tree depth, Max leaf nodes
  - Min samples in split, Min samples in leaf
- Bagging hyperparams: Bootstrap Y/N
- Random Forest hyper params
  - Number of trees (n_estimators)
  - Max samples per tree
  - Max features per tree

Random Forest advantages

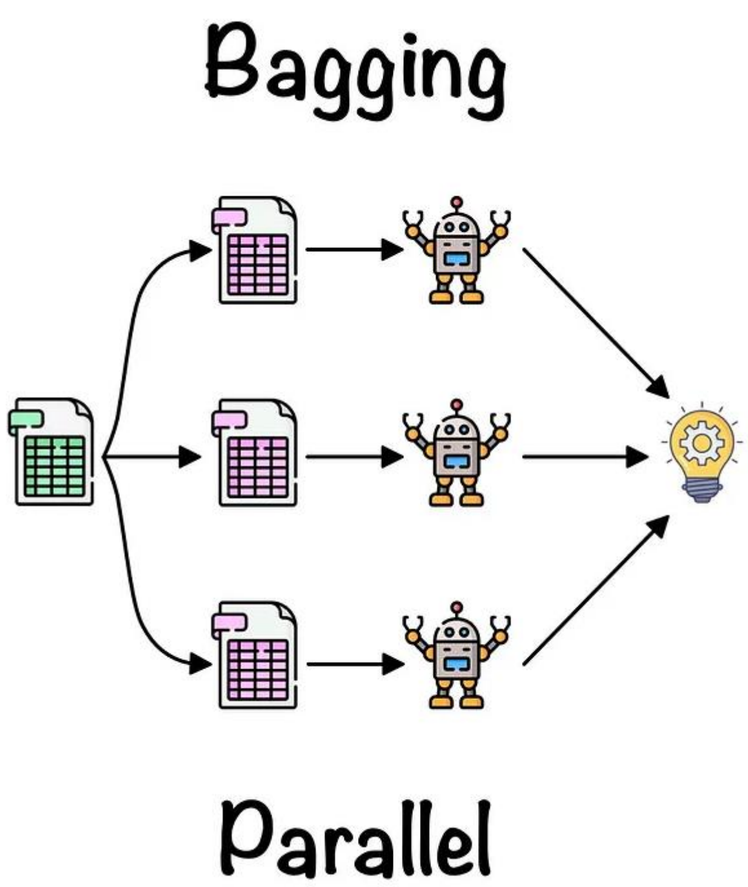
- Feature correlation does not matter
- Feature distribution does not matter
- No need to scale data
- Works well even when data is missing
- Overcomes the problem of overfitting
- Not very expensive
- Flexible and high accuracy

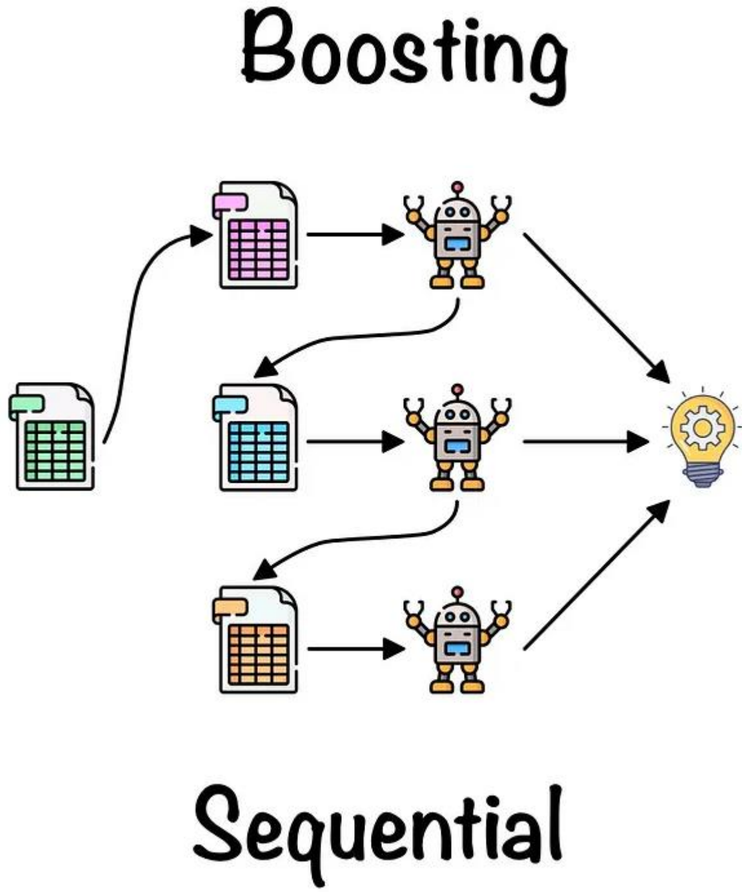

$X^2-4X+5\le 5$
$X^2-4X\le 0$
$n(B\cap C)=22$
$n(B)=68$
$n(C)=84$
$n(B\cup C)=n(B)+n(C)-n(B\cap C)$
$\frac{a}{b}+\frac{c}{d}=\frac{ad+bc}{bd}$
$\log_b b^x = x$
$\log_a x = \frac{\log_b x}{\log_b a}$
$\bar{x}_1 = \frac{1+3+3+6+8+9}{6} = 5$
$\bar{x}_2 = 2+4+4+8+12 = 30$
$\bar{x}_3 = 4+7+1+6 = 18$
20
6
(0,1)
$M = \frac{0.046765}{3\,0L}$
He = 4.002602
Na = 22.989769
Ar = 39.948
$f = \frac{R}{2}$
$C_2H_3Cl_4 + Ca(OH)_2 \rightarrow C_2HCl_3 + CaCl_2 + H_2O$
$Zn_3Sb_2 + 6H_2O \rightarrow 3Zn(OH)_2 + 2SbH_3$
$N_2 + 3H_2 \rightleftharpoons 2NH_3$
$H_2 + I_2 \rightleftharpoons 2HI$
$2SO_2 + O_2 \rightleftharpoons 2SO_3$
A
B
C

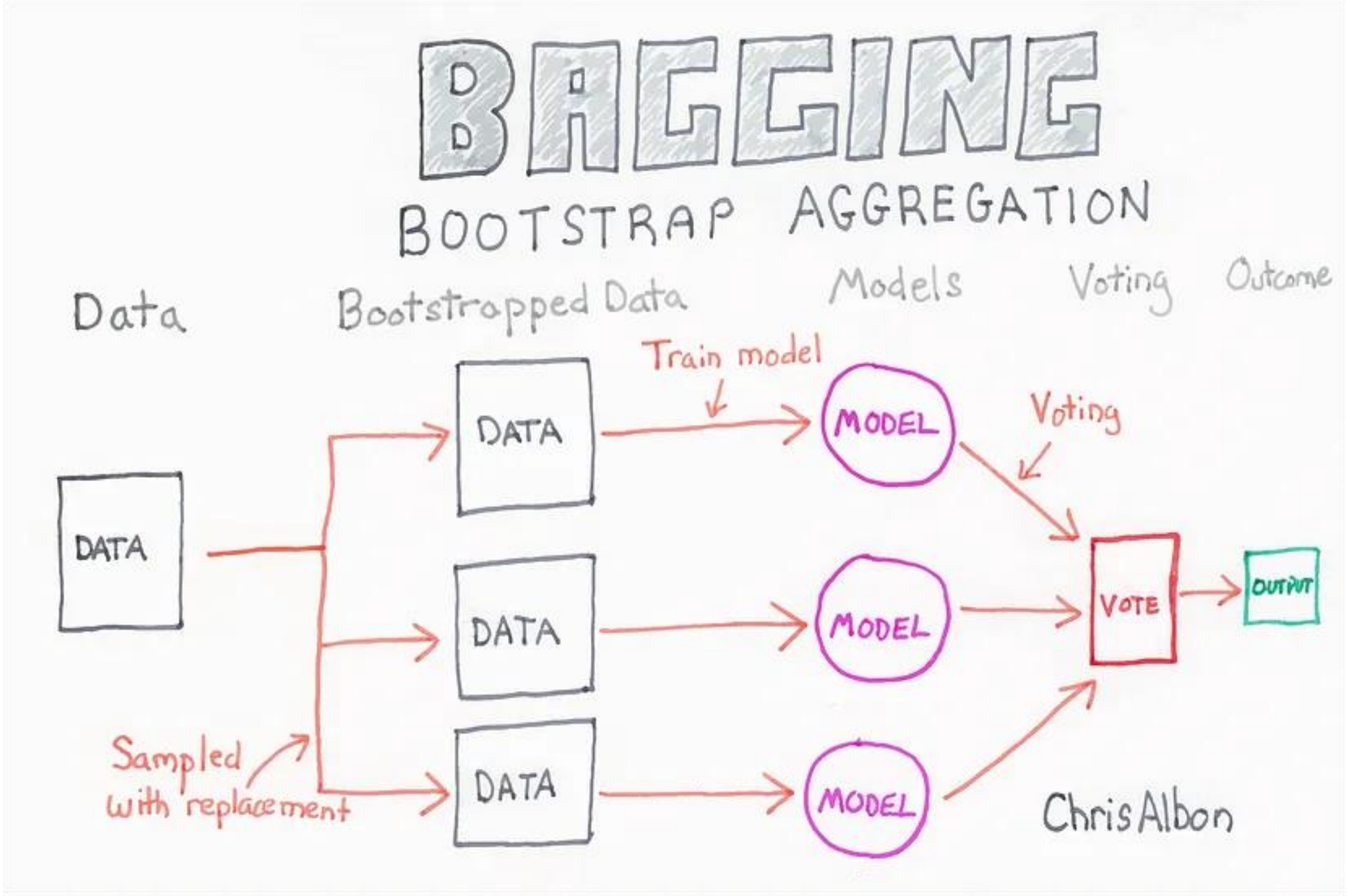
Boosting

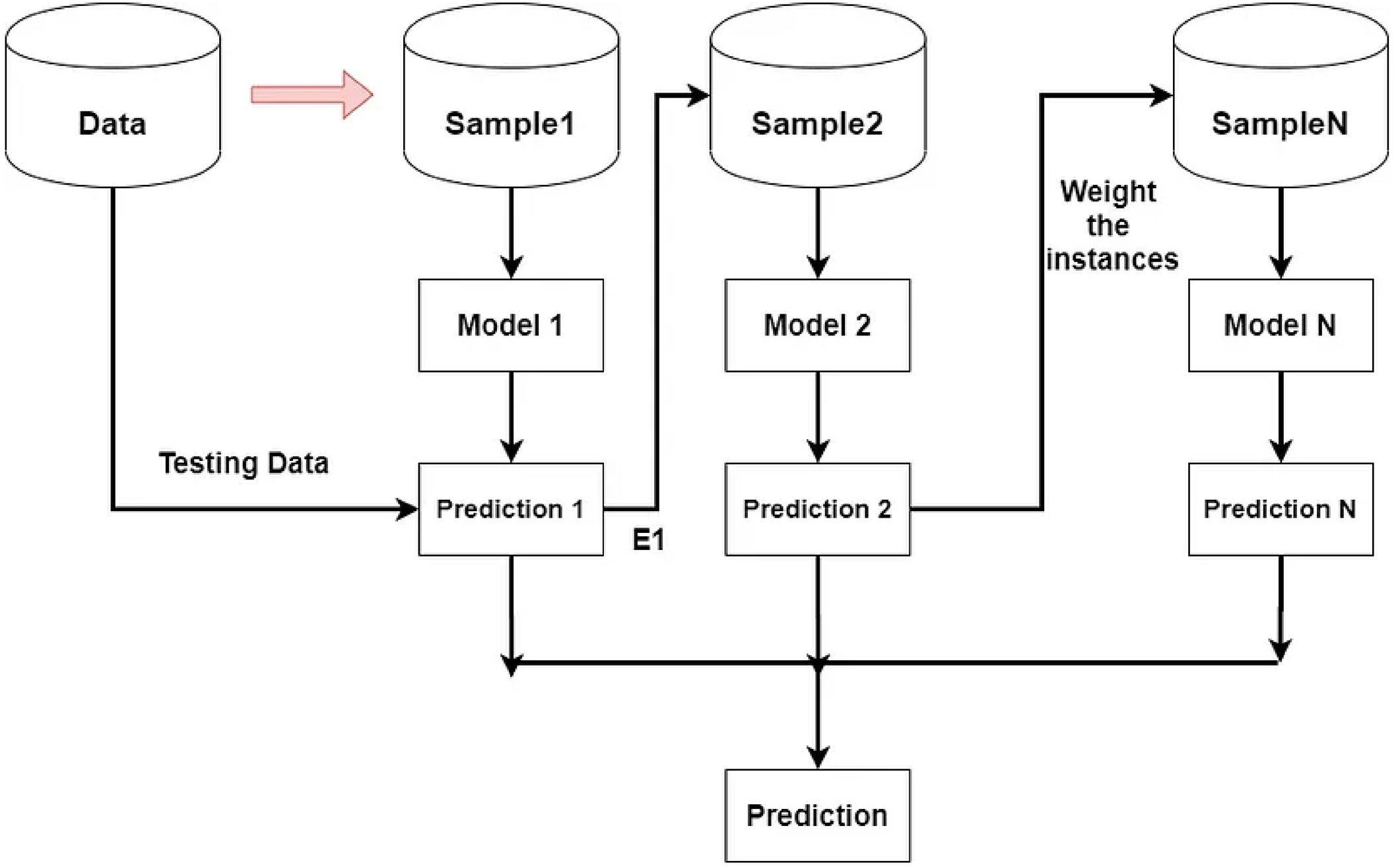
Ensemble Learning

- Multiple ML model used together for prediction






BAGGING
BOOTSTRAP AGGREGATION
Data
Bootstrapped Data
Models
Voting
Outcome
Train model
DATA
DATA
DATA
DATA
MODEL
MODEL
MODEL
Voting
VOTE
OUTPUT
Sampled with replacement
Chris Albon


Data
Sample1
Sample2
SampleN
Weight the instances
Model 1
Model 2
Model N
Testing Data
Prediction 1
E1
Prediction 2
Prediction N
Prediction

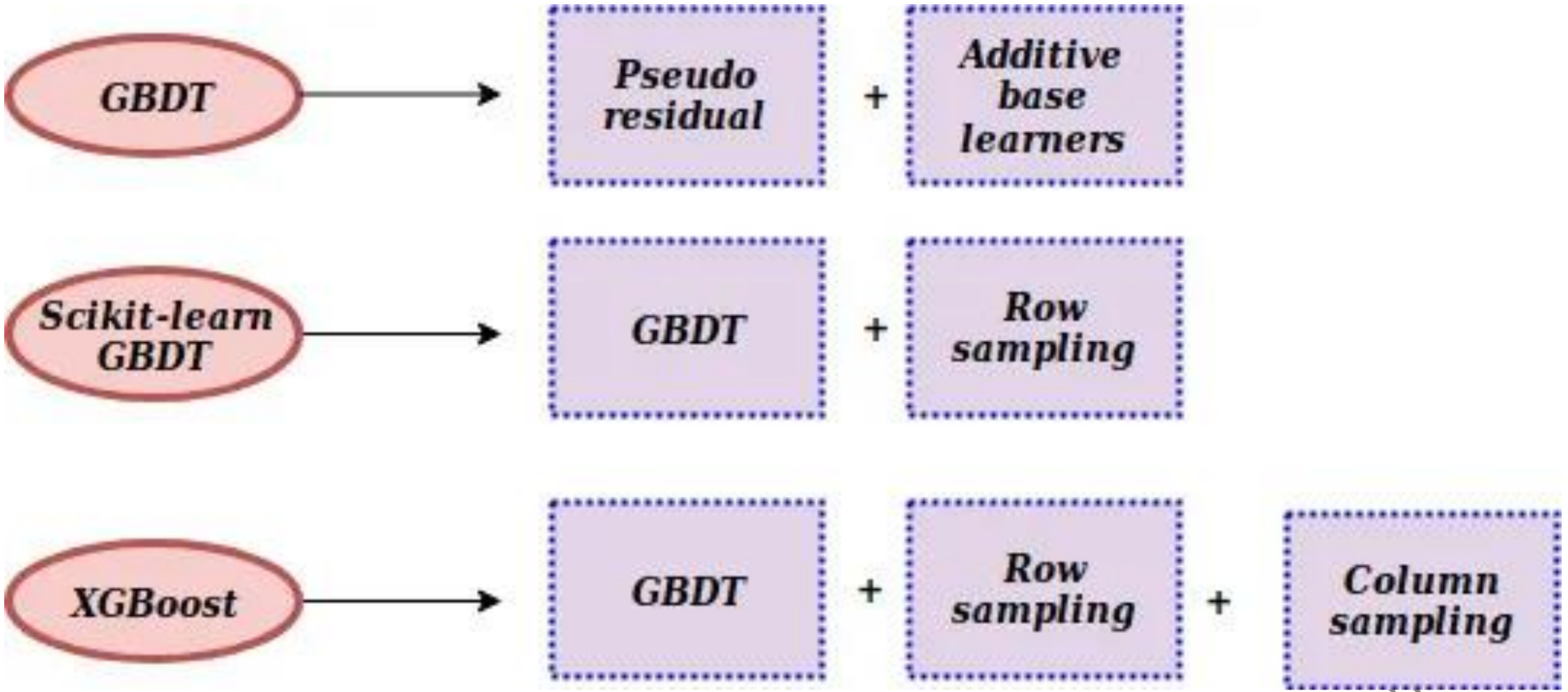
Boosting and Bias/Variance

- Boosting reduces Bias
- Hence low variance models are used as individual learner
  - E.g. Decision Tree with very less depth

- Combine high bias low variance models (weak learners) to make low bias low variance models

Boosting

- AdaBoost
- GradientBoost
- XGBoost



Helpful videos on boosting (Optional)

- AdaBoost
  - https://www.youtube.com/watch?v=LsK-xG1cLYA
- Gradient Boost (4 parts)
  - https://www.youtube.com/watch?v=3CC4N4z3GJc
- XGBoost (4 parts)
  - https://www.youtube.com/watch?v=OtD8wVaFm6E


QUESTIONS