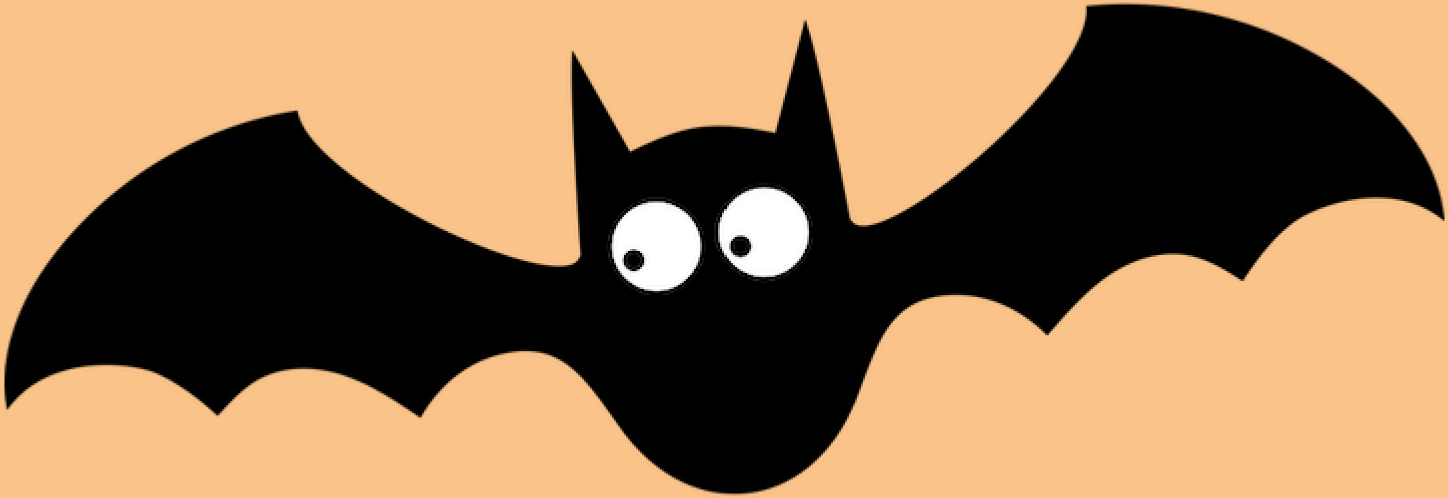


Ensemble Method and Boosting

Son Nguyen

Netflix Prize

- 1 million USD to any one that can improve Netflix's rating algorithm...

Netflix Prize

- 1 million USD to any one that can improve Netflix's rating algorithm **by 10%**

Netflix Prize

- The winning team:

Netflix Prize

- **The winning team:** BellKor's Pragmatic Chaos, used **ensemble models**"

Netflix Prize

- **The winning team:** BellKor's Pragmatic Chaos, used **ensemble models**"
- **The second-place team's name** is...

Netflix Prize

- **The winning team:** BellKor's Pragmatic Chaos, used **ensemble models**"
- **The second-place team's name** is "The Ensemble"

Netflix Prize

- "During the nearly 3 years of the Netflix competition, there were two main factors which improved the overall accuracy:
    - The quality of the individual algorithms, and
    - the **ensemble idea**"

Ensemble Success

> "XGBoost (an ensemble algorithm) is an algorithm that has recently been dominating applied machine learning and Kaggle competitions for structured or tabular data." Link

- List of machine learning winning solutions with XGBoost : https://github.com/dmlc/xgboost/tree/master/demo#machine-learning-challenge-winning-solutions

Ensemble Success

> Adaboost (an ensemble algorithm) won 2003 **Godel Prize**: AdaBoost demonstrated novel possibilities in analysing data and is a permanent contribution to science even beyond computer science. Link

Ensemble Success

> AdaBoost (with decision trees as the weak learners) is often referred to as the best out-of-the-box classifier. Link

Ensemble Success

> Leo Brieman, who invented "Bagging" and "Random Forest" crowned AdaBoost the "best off-the-shelf classifier in the world (2000).

Ensemble Machine Learning Approach

- An ensemble is a composite model, combining a series of low performing **classifiers**(classification models) or **learners** with the aim of creating an improved classifier.

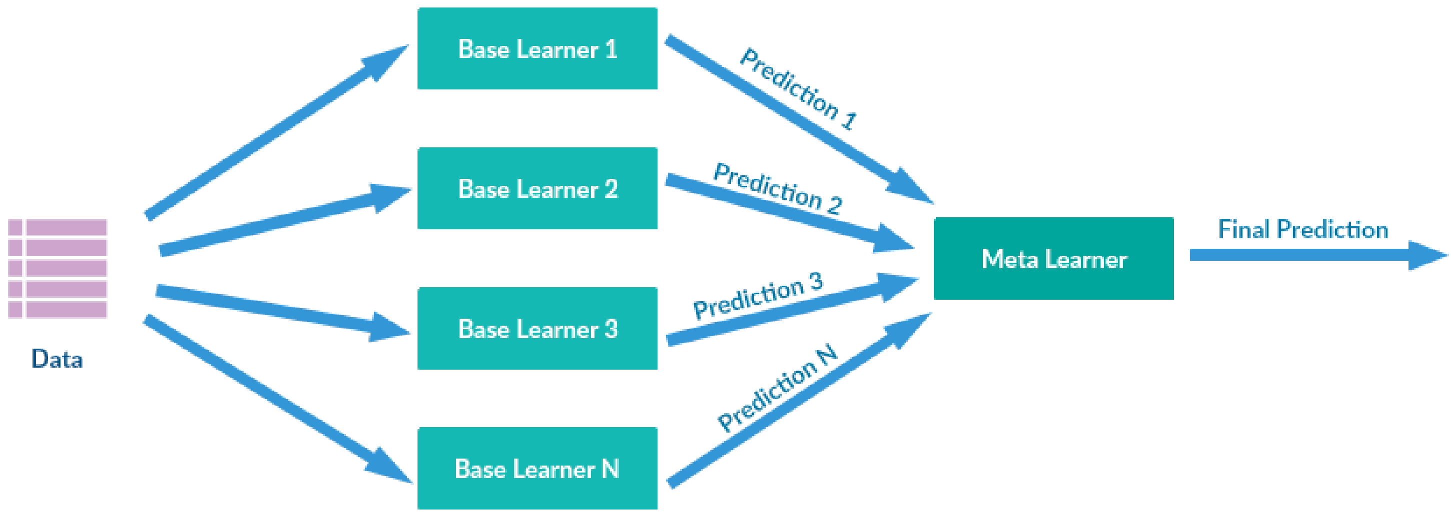
Ensemble Machine Learning Approach

- An ensemble is a composite model, combining a series of low performing **classifiers**(classification models) or **learners** with the aim of creating an improved classifier.

- Three common ensemble:
  - Stacking
  - Bagging
  - Boosting

Stacking

- Stacking combines multiple base learners predictions into a new data set.
- This new data are treated as the **input data** for another learner (meta learner).



x_1	x_2	....	x_K	$\rightarrow$	M1	M2	M3

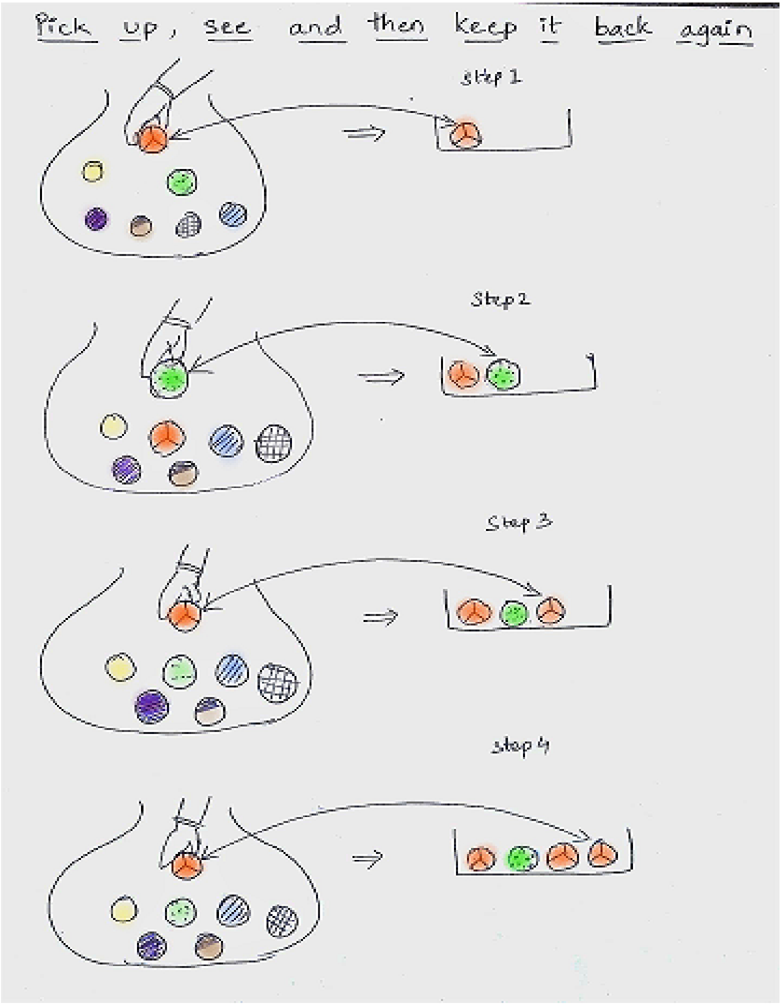
<u>Stacking</u>

x_1	x_2	...	x_K	M_1 x_{k+1}	M_2 x_{k+2}	M_3 x_{k+3}	Y	Meta model

Data to train the "meta" model

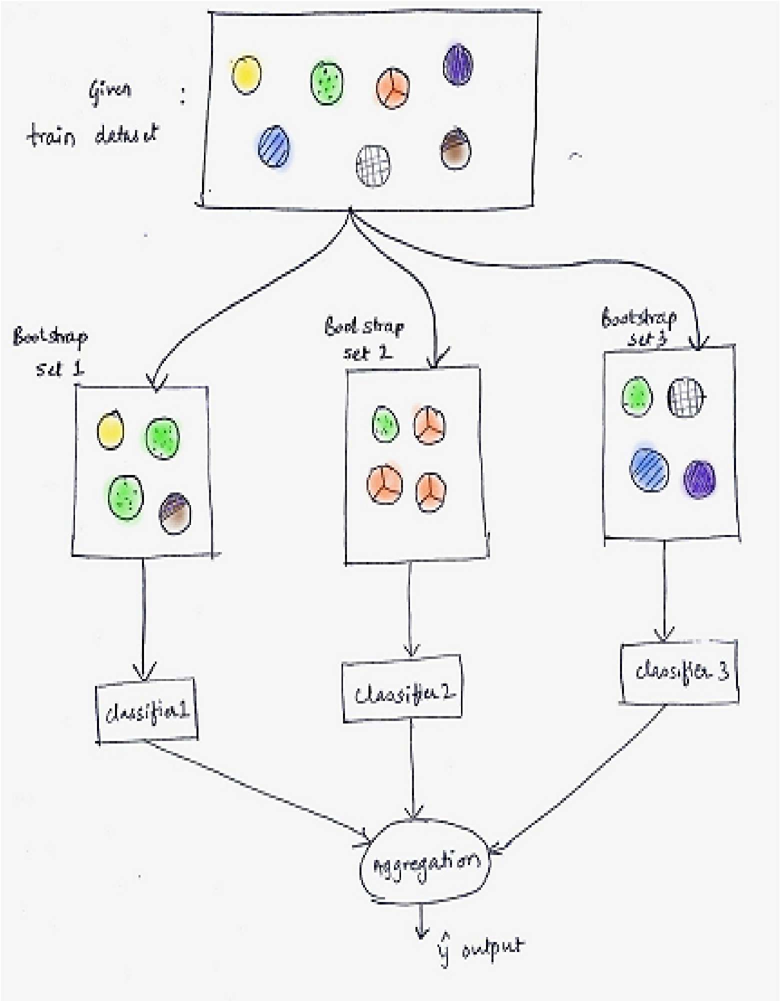
Bagging = Bootstrap Aggregating

- Step 1: Boostrapping



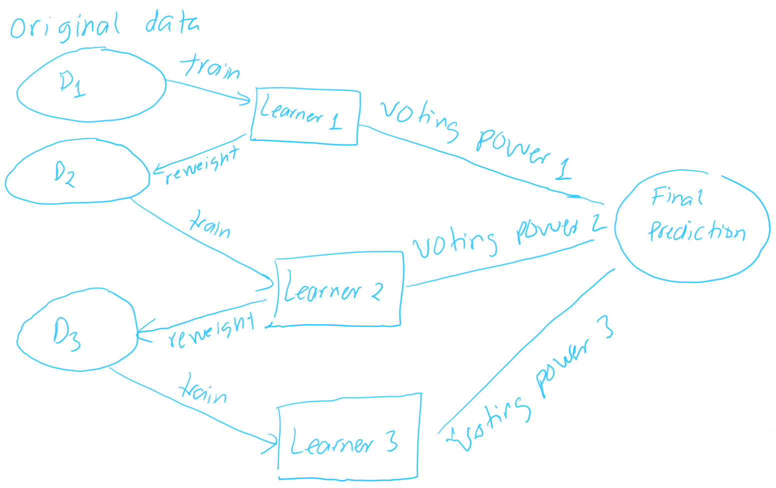
Bagging = Bootstrap Aggregating

- Step 2: Aggregating



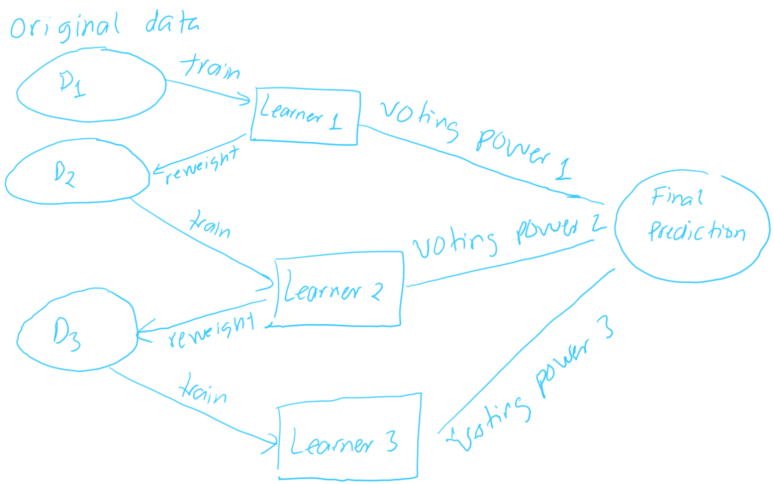
Boosting

- Weak learners are sequentially converted into a strong learner.



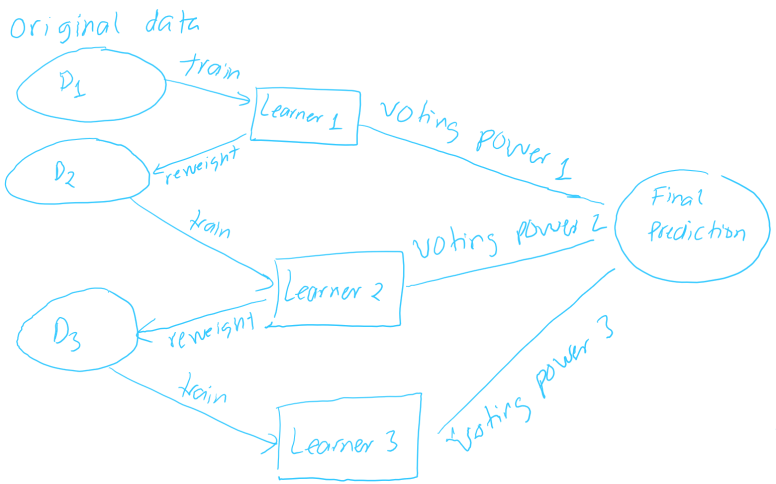
Boosting

- Weak learners are sequentially converted into a strong learner.
- Each leaner in the sequence tries to correct its predecessor.



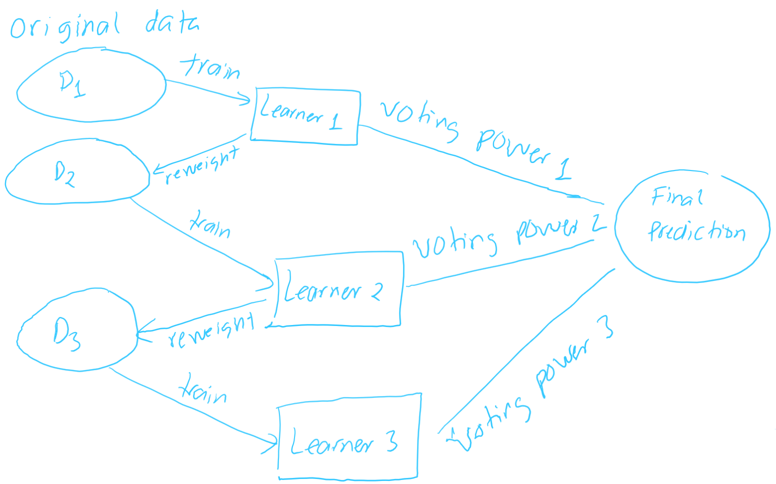
Boosting

- Weak learners are sequentially converted into a strong learner.
- Each leaner in the sequence tries to correct its predecessor.
- Individually, Learn 1, Learner 2 and Learner 3 are **weak leaners** (a little better than a coin toss).



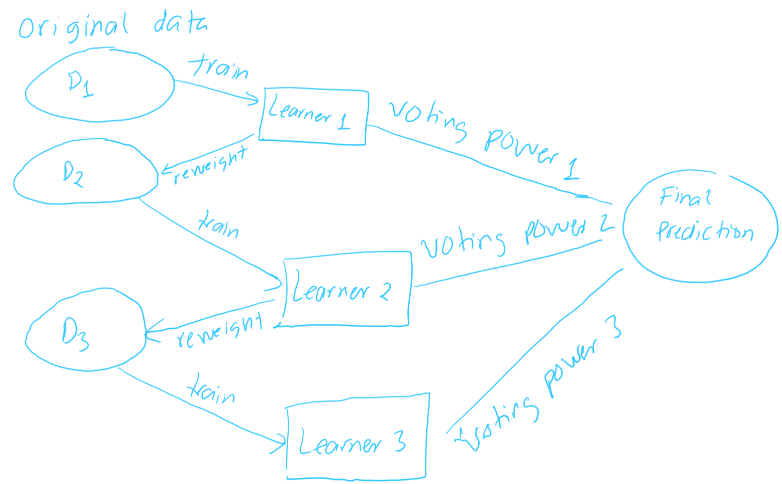
Boosting

- Weak learners are sequentially converted into a strong learner.
- Each leaner in the sequence tries to correct its predecessor.
- Individually, Learn 1, Learner 2 and Learner 3 are **weak leaners** (a little better than a coin toss).
- The learners are usually a tree with 2 leaves (**Stump**).

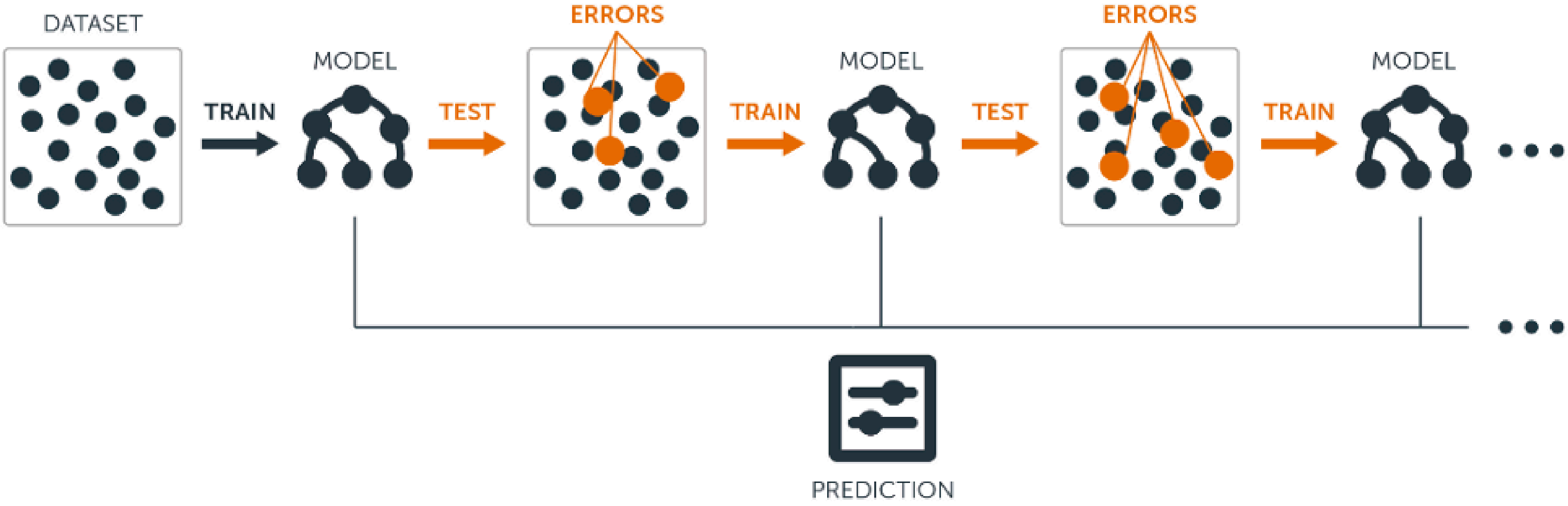


Boosting

- Weak learners are sequentially converted into a strong learner.
- Each leaner in the sequence tries to correct its predecessor.
- Individually, Learn 1, Learner 2 and Learner 3 are **weak leaners** (a little better than a coin toss).
- The learners are usually a tree with 2 leaves (**Stump**).
- In D2, the wrong misclassified of Learner 1 gets higher weights.
- In D3, the wrong misclassified of Learner 2 gets higher weights.

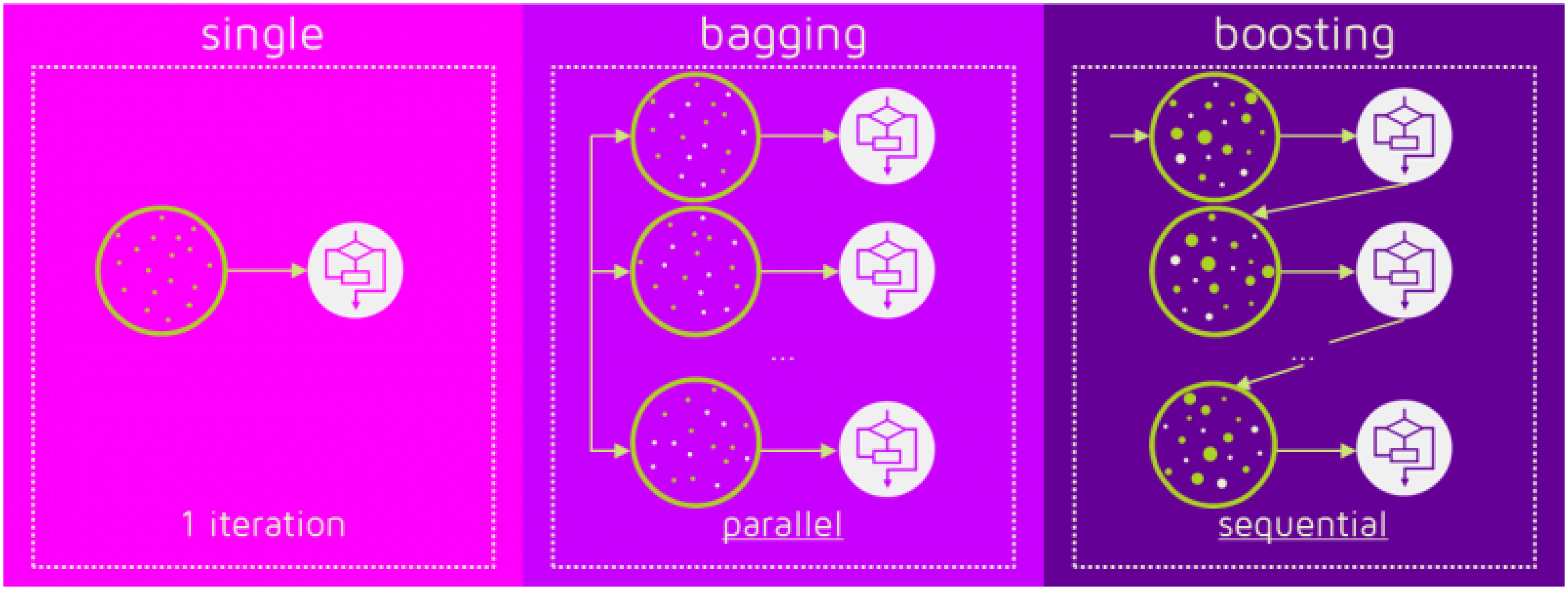


Boosting

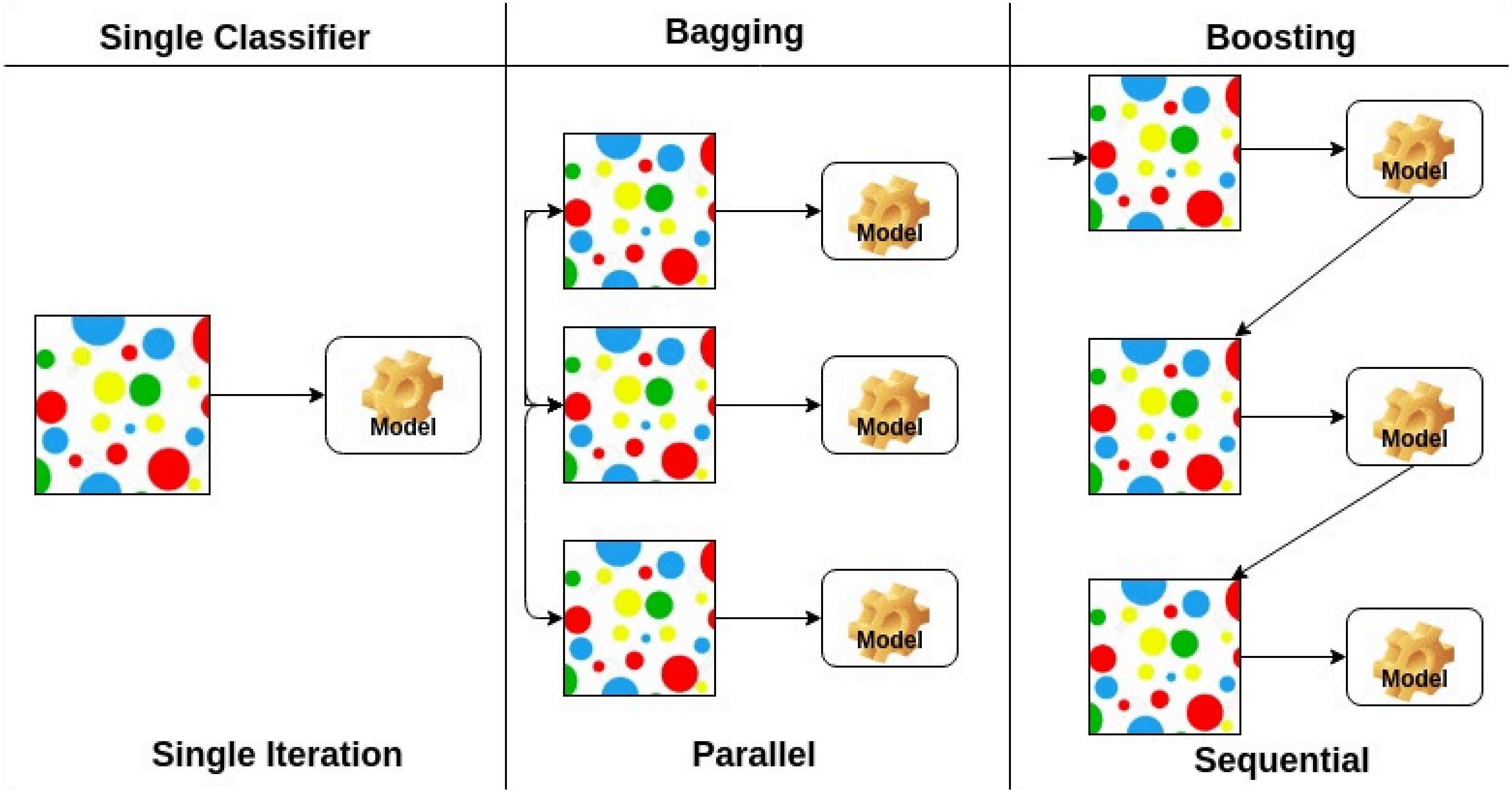


DATASET
TRAIN
MODEL
TEST
ERRORS
TRAIN
MODEL
TEST
ERRORS
TRAIN
MODEL
PREDICTION

Bagging vs. Boosting



Bagging vs. Boosting



Types of Boosting

- Adaboost
- Gradient Boosting

Adaboost

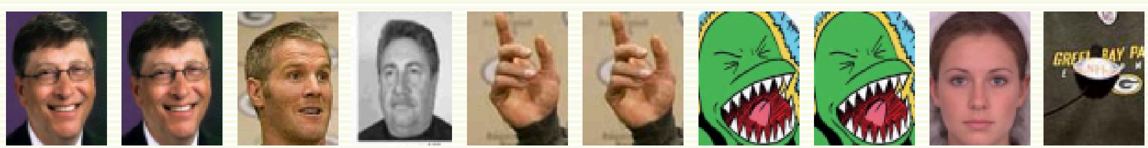
Idea Behind Ada Boost

- Examples of high weight are shown more often at later rounds
- Face/nonface classification problem:

Round 1

	1/7	1/7	1/7	1/7	1/7	1/7	1/7
best weak classifier:	✓	✗	✓	✓	✗	✓	✗
change weights:	1/16	1/4	1/16	1/16	1/4	1/16	1/4

Round 2

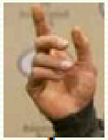


best weak classifier:	✓	✓	✓	✗	✗	✗	✓	✓	✓	✓
change weights:		1/8	1/32	11/32		1/2		1/8	1/32	1/32

Adaboost

Idea Behind Ada Boost

Round 3

- out of all available weak classifiers, we choose the one that works best on the data we have at round 3
- we assume there is always a weak classifier better than random (better than 50% error)
-  image is half of the data given to the classifier
- chosen weak classifier **has to** classify this image correctly

Adaboost, Clearly Explained

- Demonstration by StatQuest
- Link