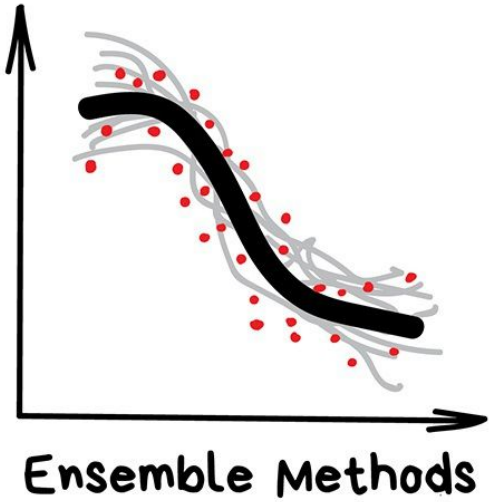
Ensemble Methods

Ramesh S

Introduction

- "*Bunch of stupid trees learning to correct errors of each other*"
- Nowadays is used for:
  - Search systems (★)
  - Computer vision
  - Object detection
  - Everything that fits classical algorithm approaches (but works better)

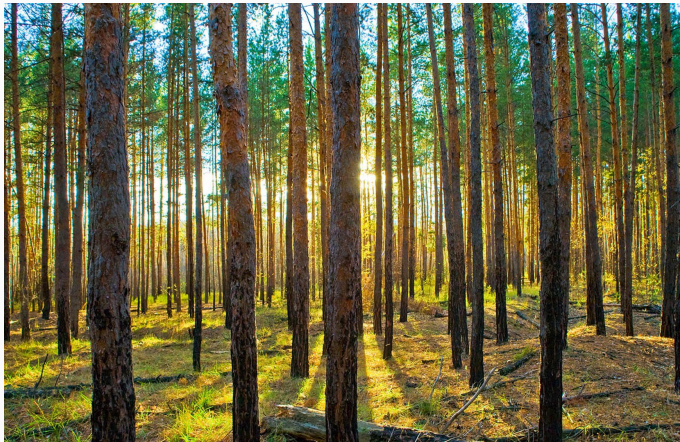


Ensemble Methods

- Suppose you ask a complex question to thousands of random people, then aggregate their answers.
- In many cases you will find that this aggregated answer is better than an expert's answer.
- This is called the *wisdom of the crowd*.

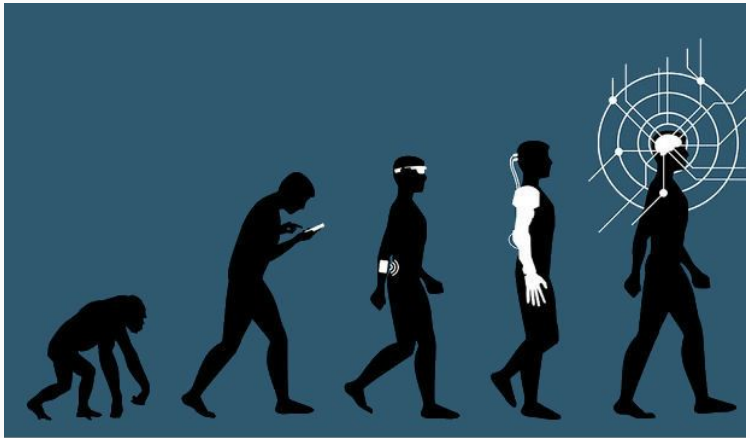
Ensemble Methods

- For example, you can 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.
- You’ve just built a Random Forest



Ensemble Methods

- It's time for modern, grown-up methods.
- Ensembles and neural networks are two main fighters paving our path to a singularity.
- Today they are producing the most accurate results and are widely used in production.



Ensemble Methods

- We can use any algorithm we know to create an ensemble.
- Just throw a bunch of classifiers, spice it up with regression and don't forget to measure accuracy.

Ensemble Methods

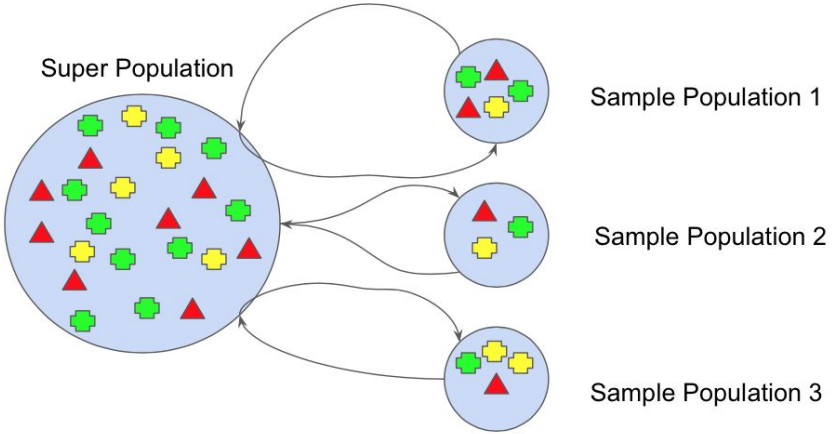
- If you aggregate the predictions of a group of predictors (such as classifiers or regressors), you will often get better predictions than with the best individual predictor.
- A group of predictors is called an *ensemble*; thus, *Ensemble Learning*
- An Ensemble Learning algorithm is called an Ensemble method.

Ensemble Methods - Types

- Bagging (Random Forests)
- Boosting
- Stacking and few others

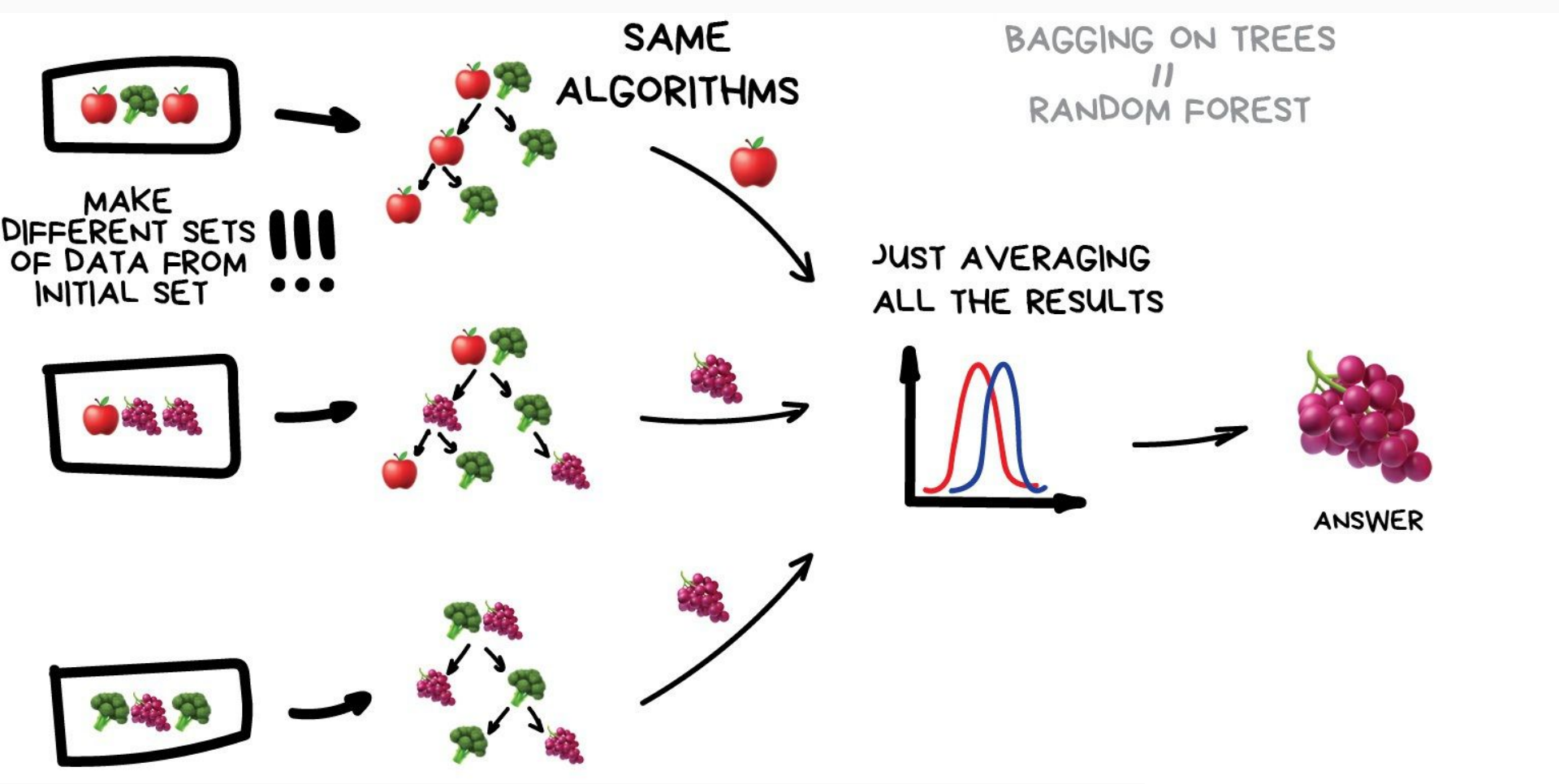
Bagging

- **B**ootstrap **Agg**regat**ing** aka Bagging uses the same algorithm but trains it on different subsets of original data.
- In the end — just average answers.



Bagging

- Data in random subsets may repeat.
- For example, from a set like "1-2-3" we can get subsets like "2-2-3", "1-2-2", "3-1-2" and so on.
- We use these new datasets to teach the same algorithm several times and then predict the final answer via simple majority voting.

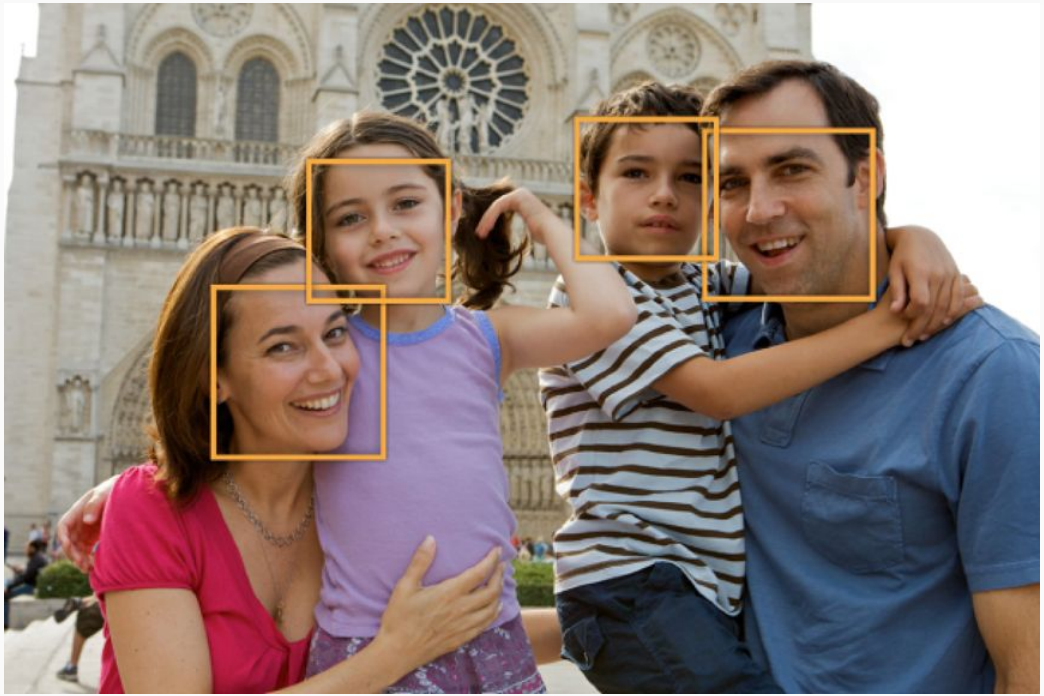

SAME ALGORITHMS
BAGGING ON TREES
=
RANDOM FOREST
MAKE DIFFERENT SETS OF DATA FROM INITIAL SET
!!!
JUST AVERAGING ALL THE RESULTS
ANSWER

Bagging

Bagging

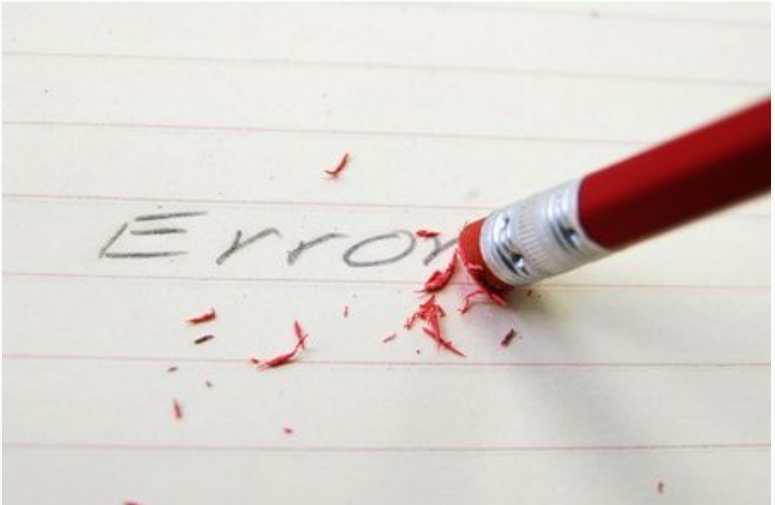
- The most famous example of bagging is the Random Forest algorithm, which is simply bagging on the decision trees.
- When you open your phone's camera app and see it drawing boxes around people's faces — it's probably the results of Random Forest work.
- Neural networks would be too slow to run real-time yet bagging is ideal given it can calculate trees on all the shaders of a video card or on these new fancy ML processors.

Bagging



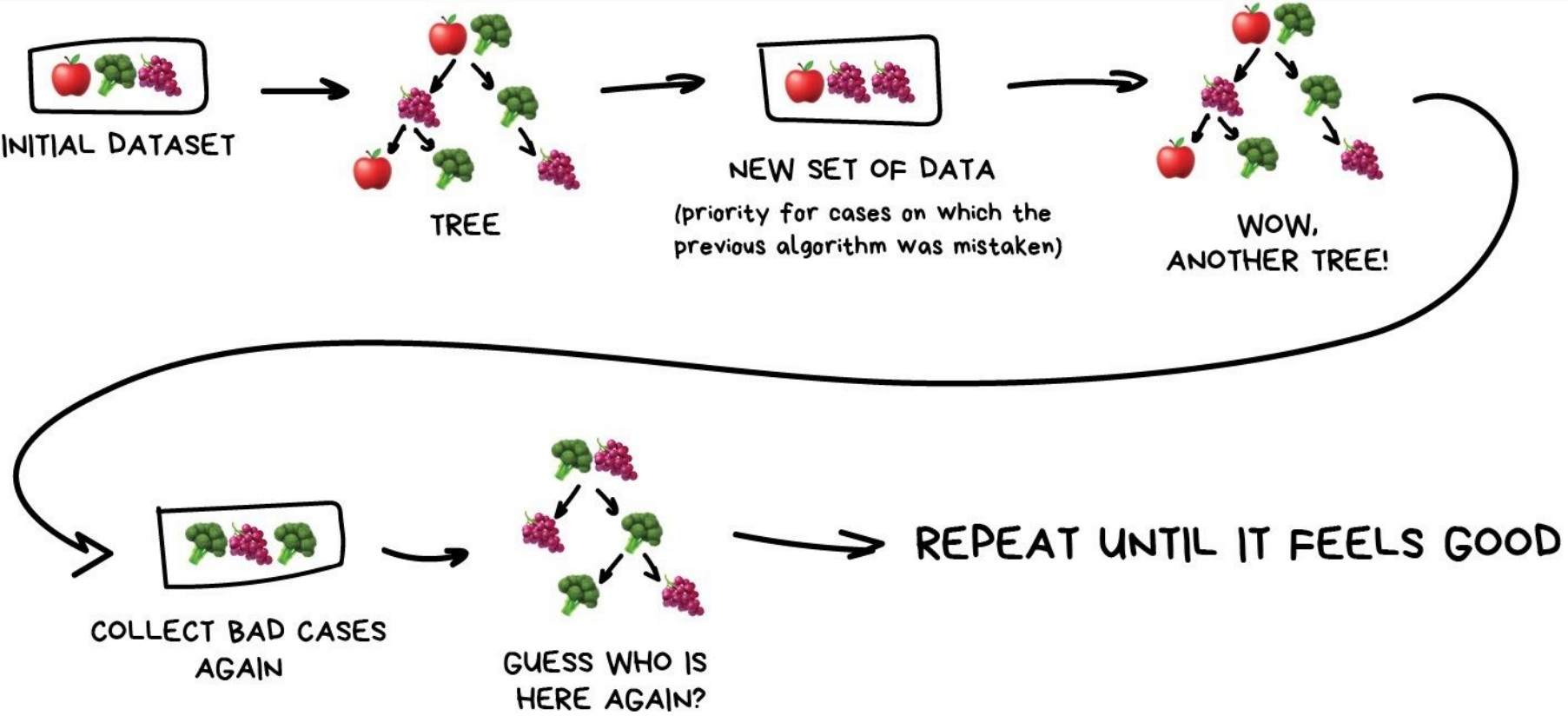
Boosting

- Boosting Algorithms are trained one by one sequentially.
- Each subsequent one paying most of its attention to data points that were mispredicted by the previous one.
- Repeat until you are happy.



Boosting

- Same as in bagging, we use subsets of our data but this time they are not randomly generated.
- Now, in each subsample we take a part of the data the previous algorithm failed to process.
- Thus, we make a new algorithm learn to fix the errors of the previous one.


INITIAL DATASET
TREE
NEW SET OF DATA
(priority for cases on which the previous algorithm was mistaken)
WOW, ANOTHER TREE!
REPEAT UNTIL IT FEELS GOOD
GUESS WHO IS HERE AGAIN?
COLLECT BAD CASES AGAIN

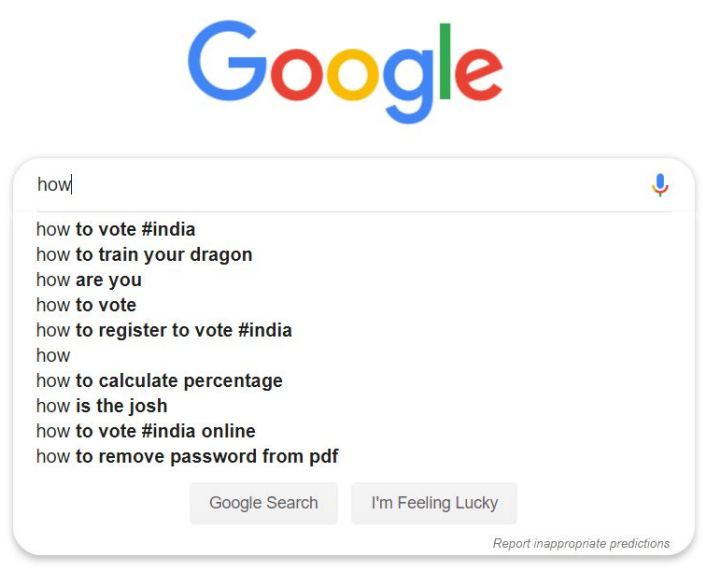
Boosting

Boosting

- The main advantage here — a very high precision of classification.
- The cons were already called out — it doesn't parallelize.
- But it's still faster than neural networks.
- It's like a race between a dump truck and a racecar.
- The truck can do more, but if you want to go fast — take a car.

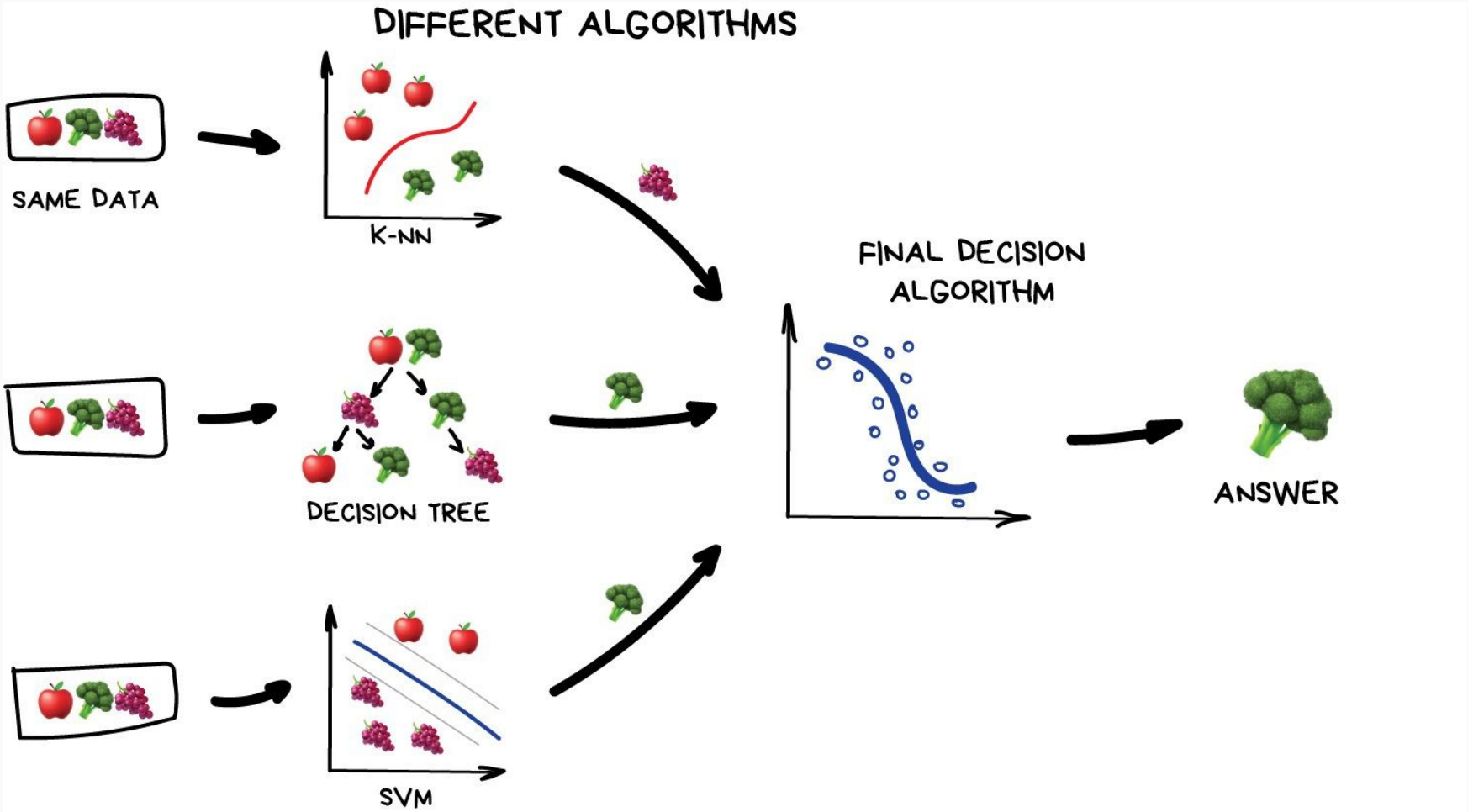
Boosting

- If you want a real example of boosting — open Facebook or Google and start typing in a search query.
- Can you hear an army of trees roaring and smashing together to sort results by relevance?
- That's because they are using boosting.



Stacking

- Output of ***several*** parallel models is passed as input to the last one which makes a final decision.
- Like that girl who asks her girlfriends whether to meet with you in order to make the final decision herself.


DIFFERENT ALGORITHMS
SAME DATA
K-NN
DECISION TREE
SVM
FINAL DECISION ALGORITHM
ANSWER

Stacking

Stacking

- Emphasis here on the word "different".
- Mixing the same algorithms on the same data would make no sense.
- The choice of algorithms is completely up to you.
- However, for final decision-making model, regression is usually a good choice.