

Ensemble Learning

Making a whole greater than
any one of its parts

Ensemble Learning

Making a whole greater than
any one of its parts

(where each part is a supervised learner)

Allen Grimm

4/15/2015

allen@GrimmScience.com

@GrimmScientist

Ensemble Learning

Making a whole greater than
any one of its parts

(where each part is a supervised learner)

(and so is the whole)

Allen Grimm

4/15/2015

allen@GrimmScience.com

@GrimmScientist

A Single Part

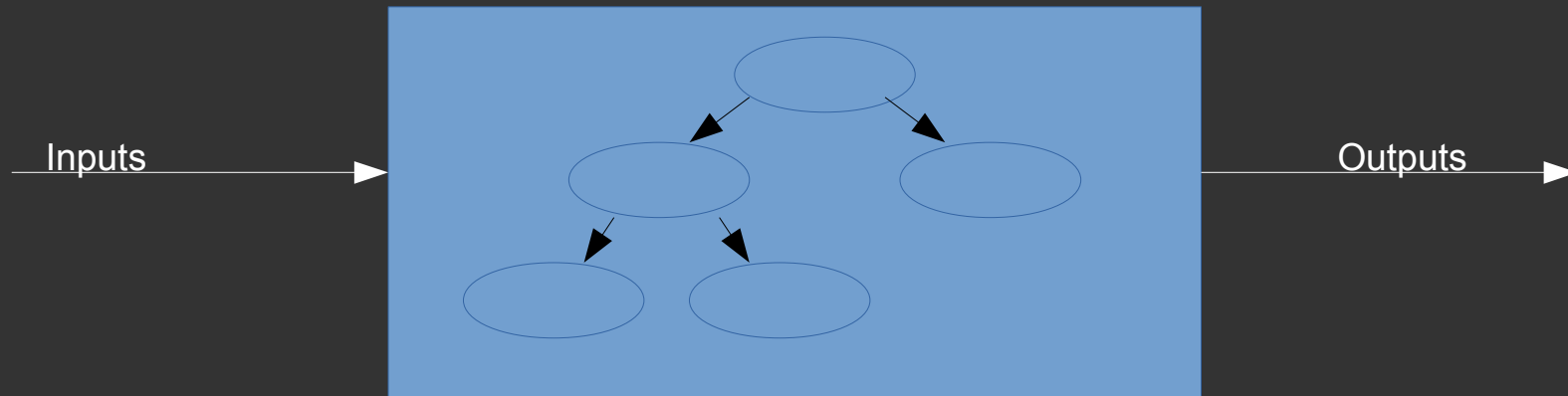
The supervised learning 'black box':



A Single Part

The supervised learning 'black box':

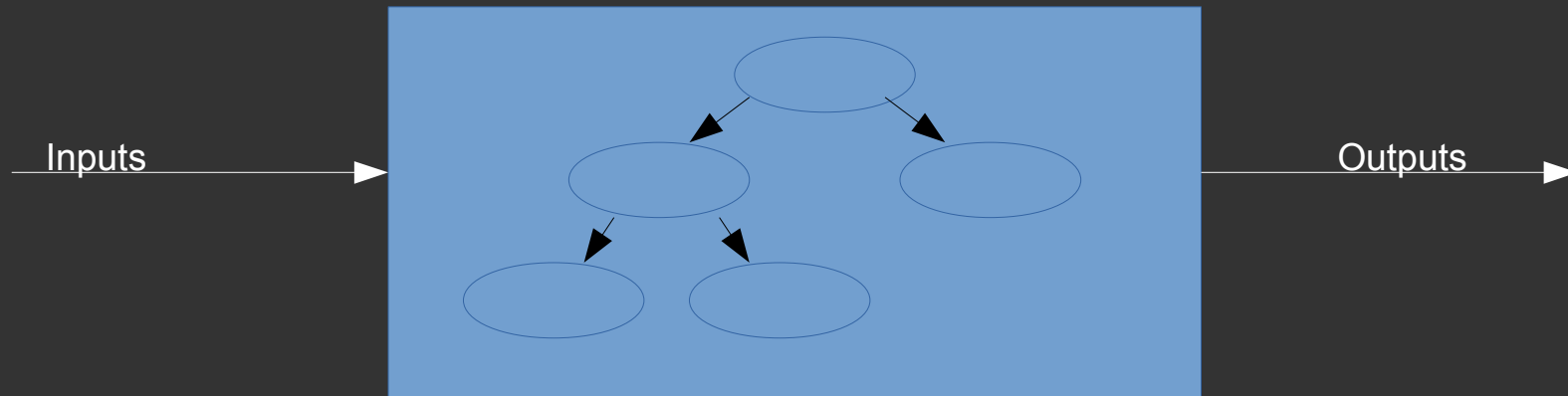
Let's say we use a decision tree...



A Single Part

The supervised learning 'black box':

Let's say we use a decision tree...

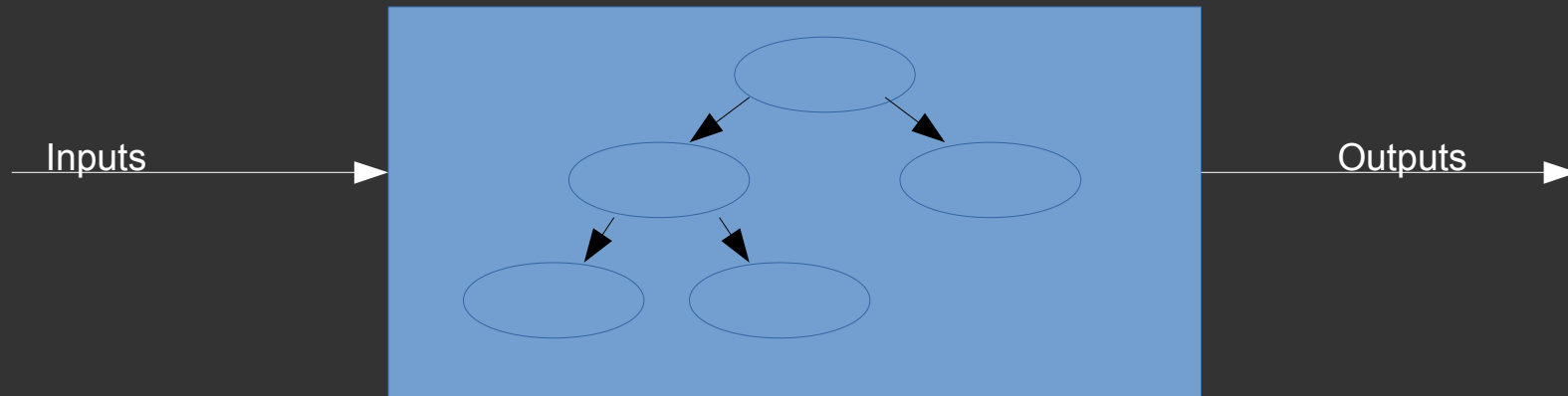


Advantages: Very fast. Intuitive results.

A Single Part

The supervised learning 'black box':

Let's say we use a decision tree...



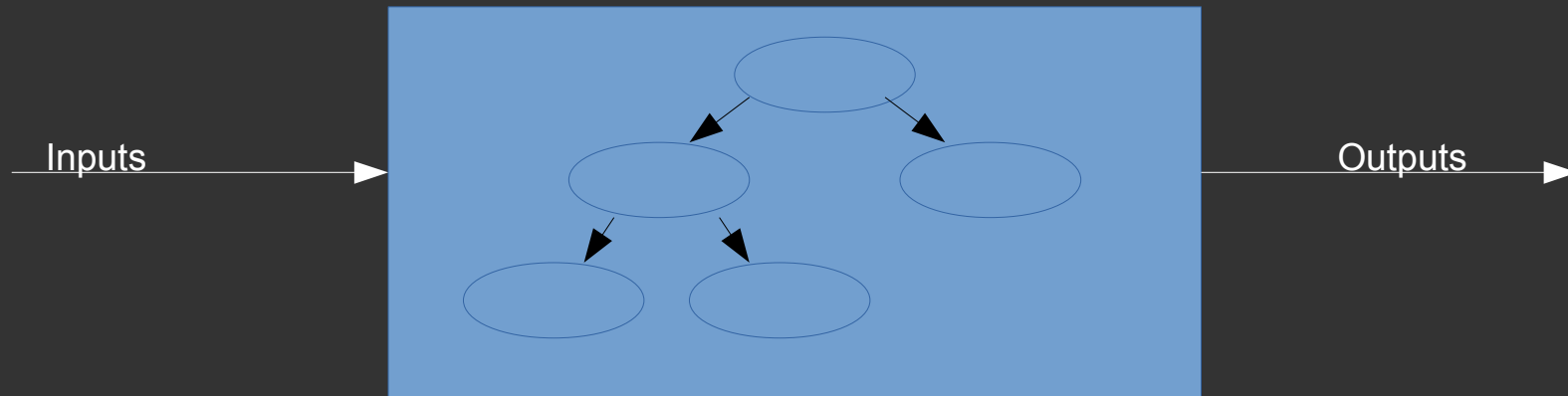
Advantages: Very fast. Intuitive results.

Problem: Too quickly shifts from underfit to overfit

A Single Part

The supervised learning 'black box':

Let's say we use a decision tree...

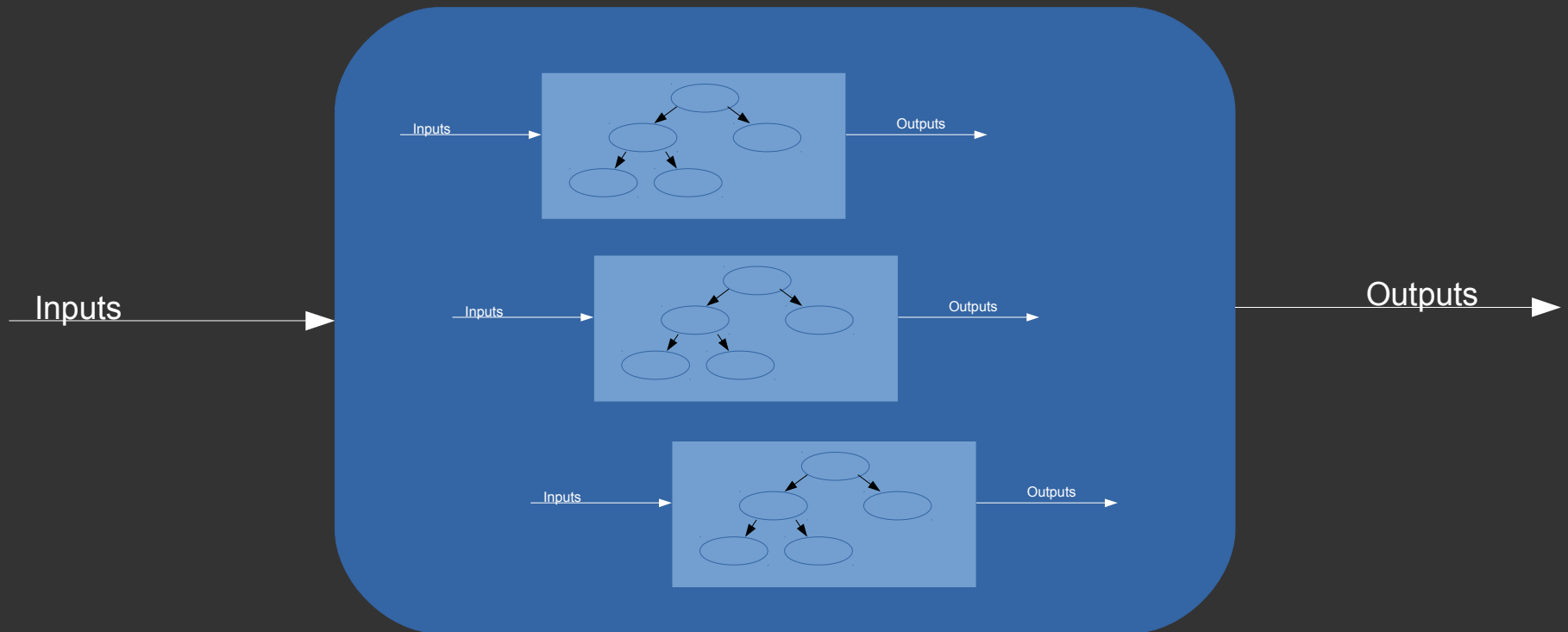


Advantages: Very fast. Intuitive results.

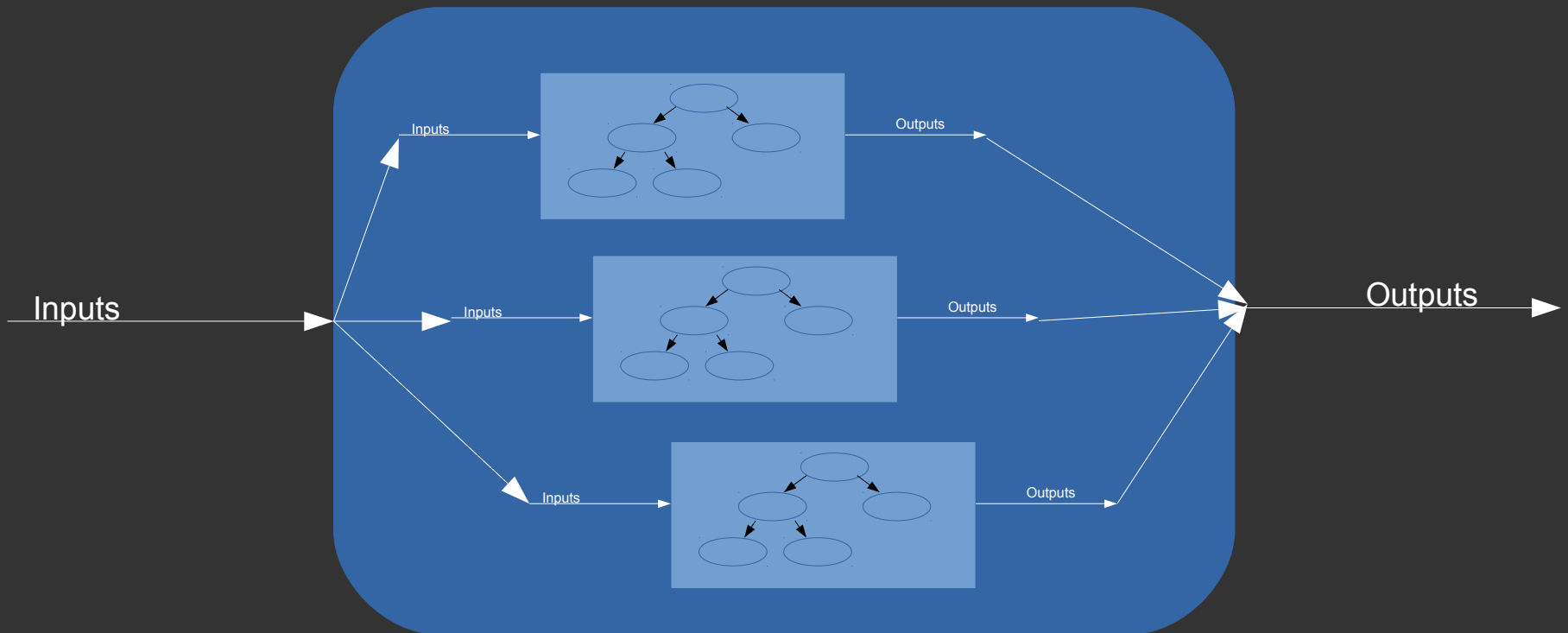
Problem: Too quickly shifts from underfit to overfit

Possible solution: Train a few of them!

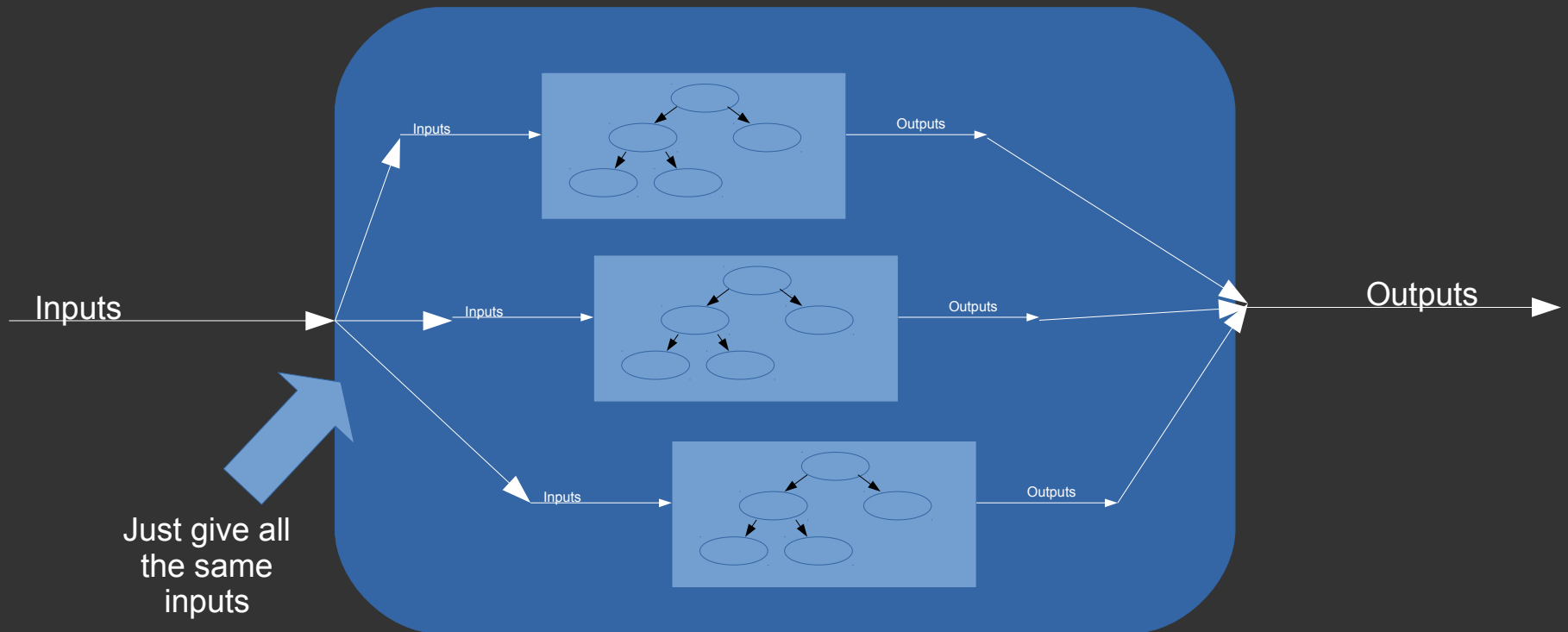
Build the Whole



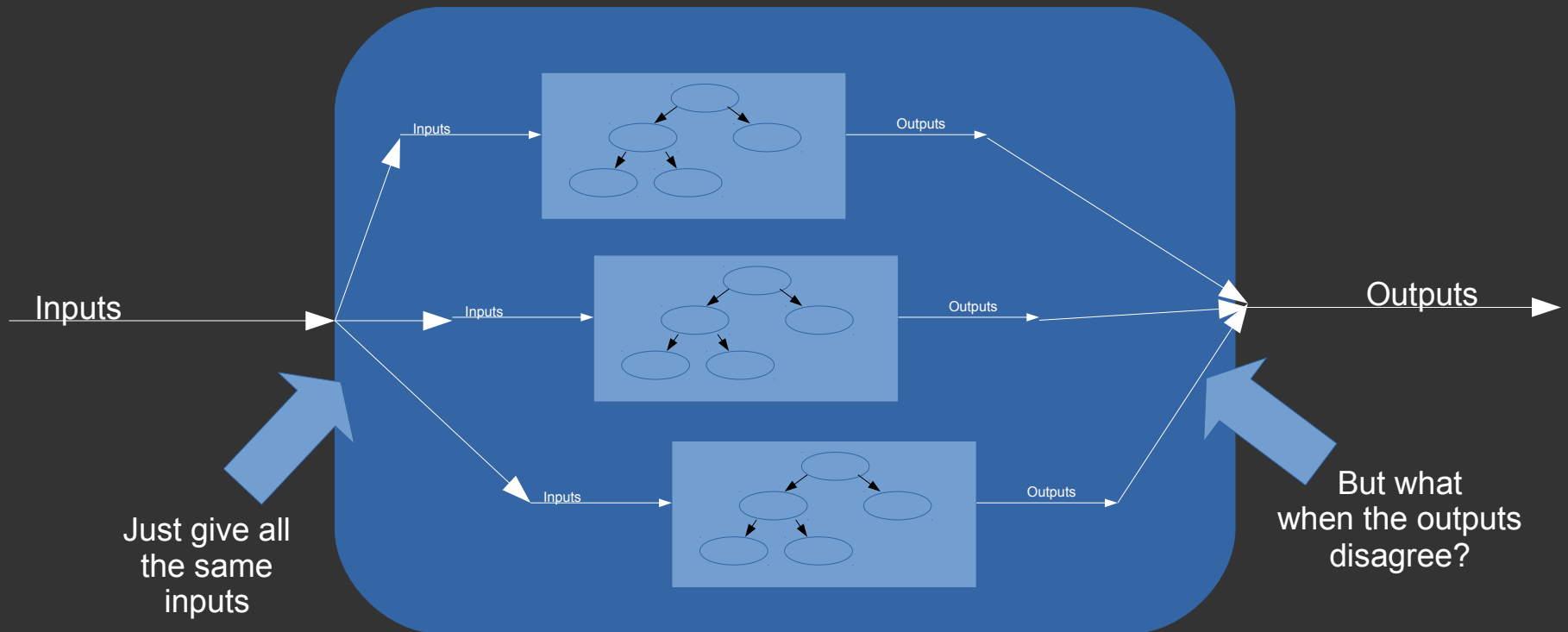
Build the Whole



Build the Whole



Build the Whole



Ensemble Learning Strategies

- Bagging
- Boosting
- Bayes Optimal Classifier
- Stacking
- ...

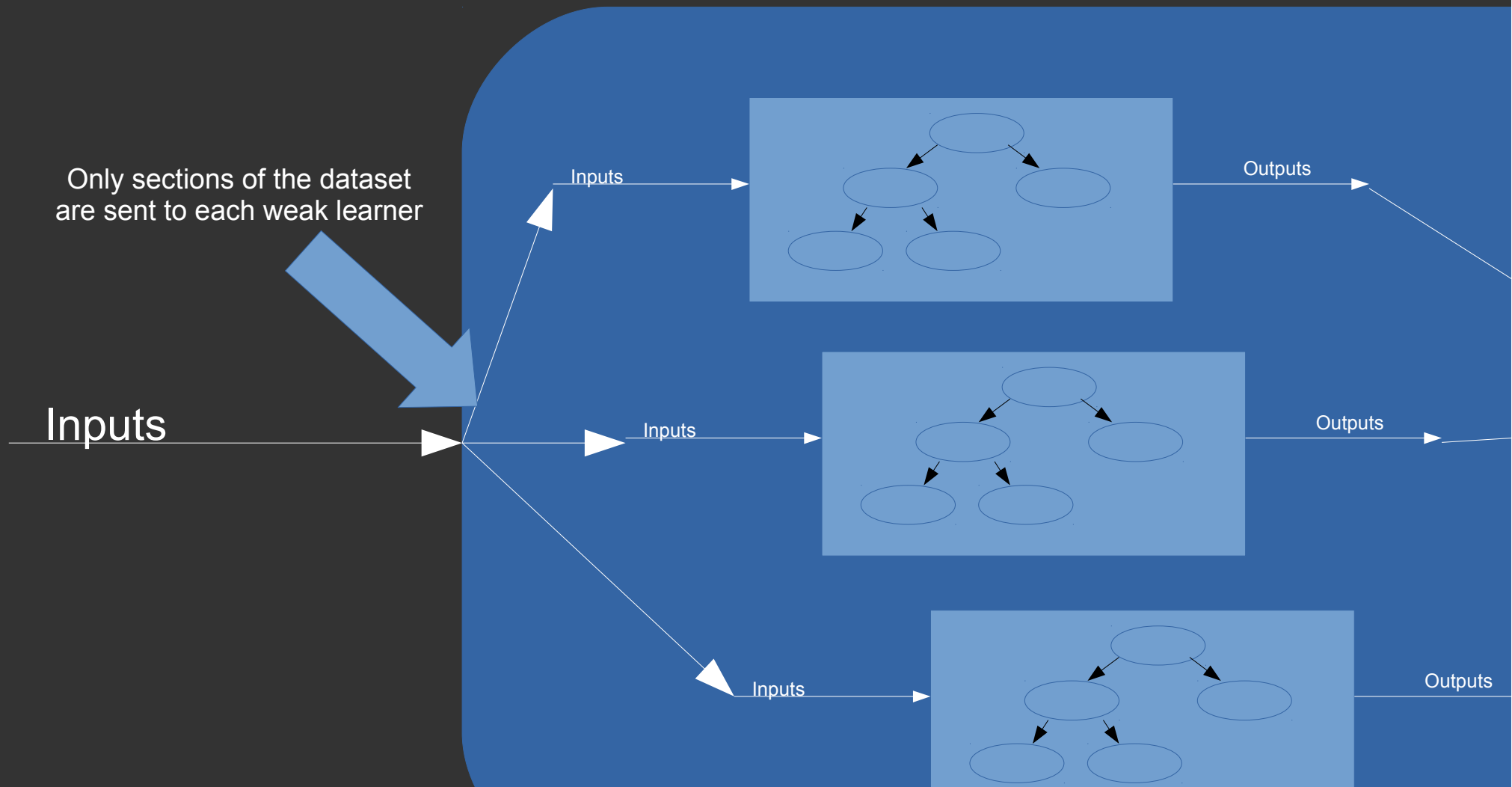
Ensemble Learning Strategies

- Bagging
- Boosting
- Bayes Optimal Classifier
- Stacking
- ...

And then, to random forests!

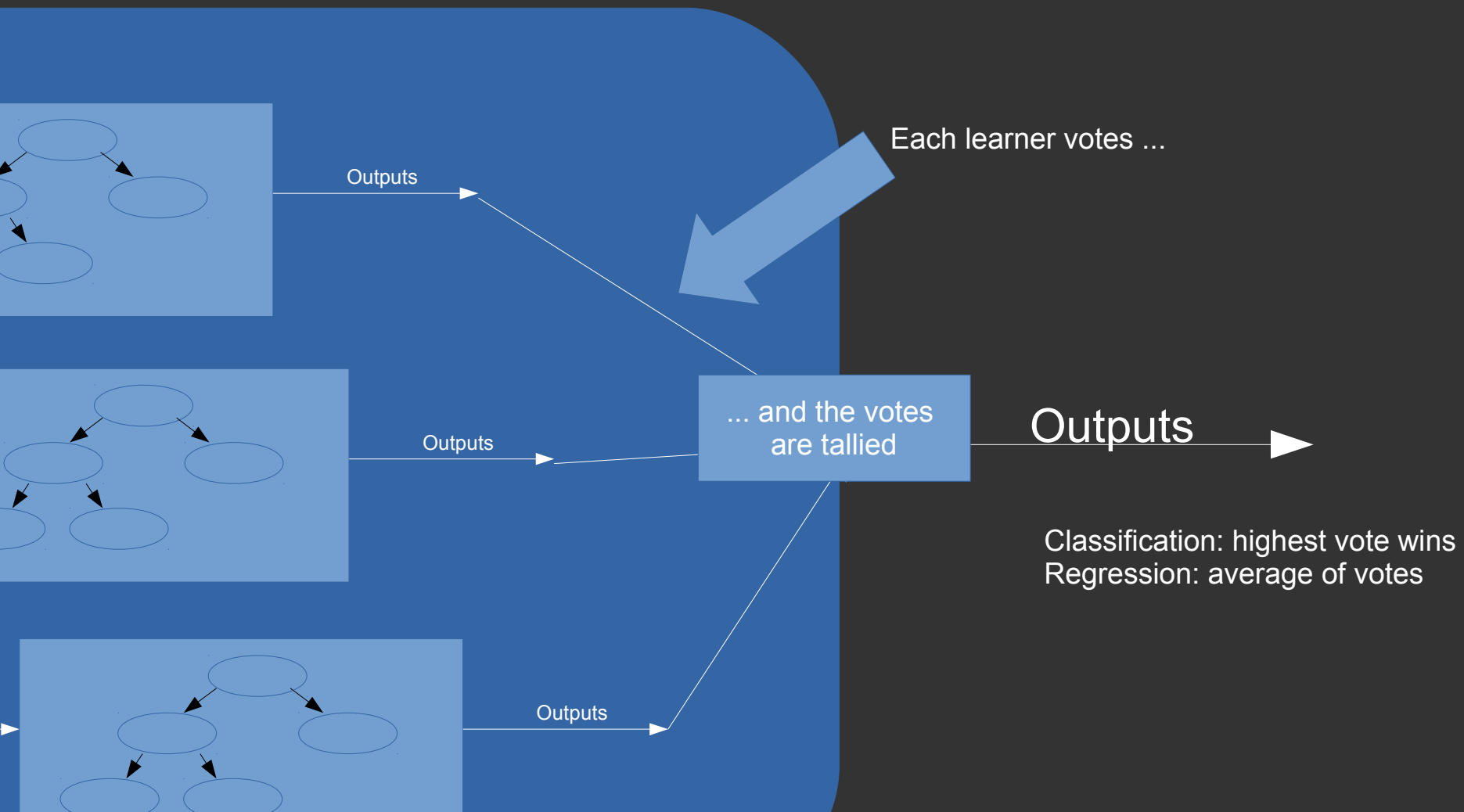
Bootstrap Aggregating

“Bootstrap” means to sample with replacement



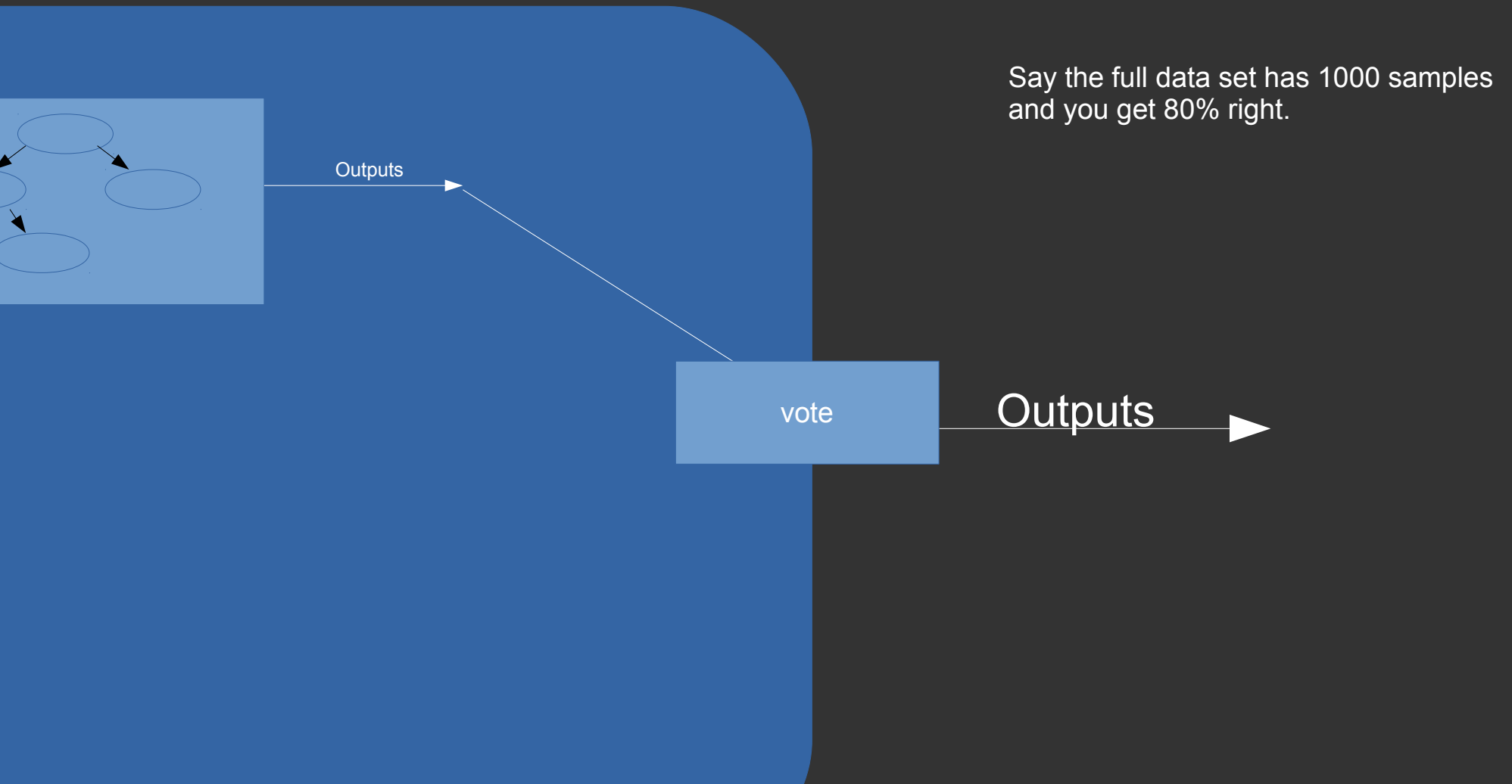
Bootstrap Aggregating

Usually abbreviated to “Bagging”



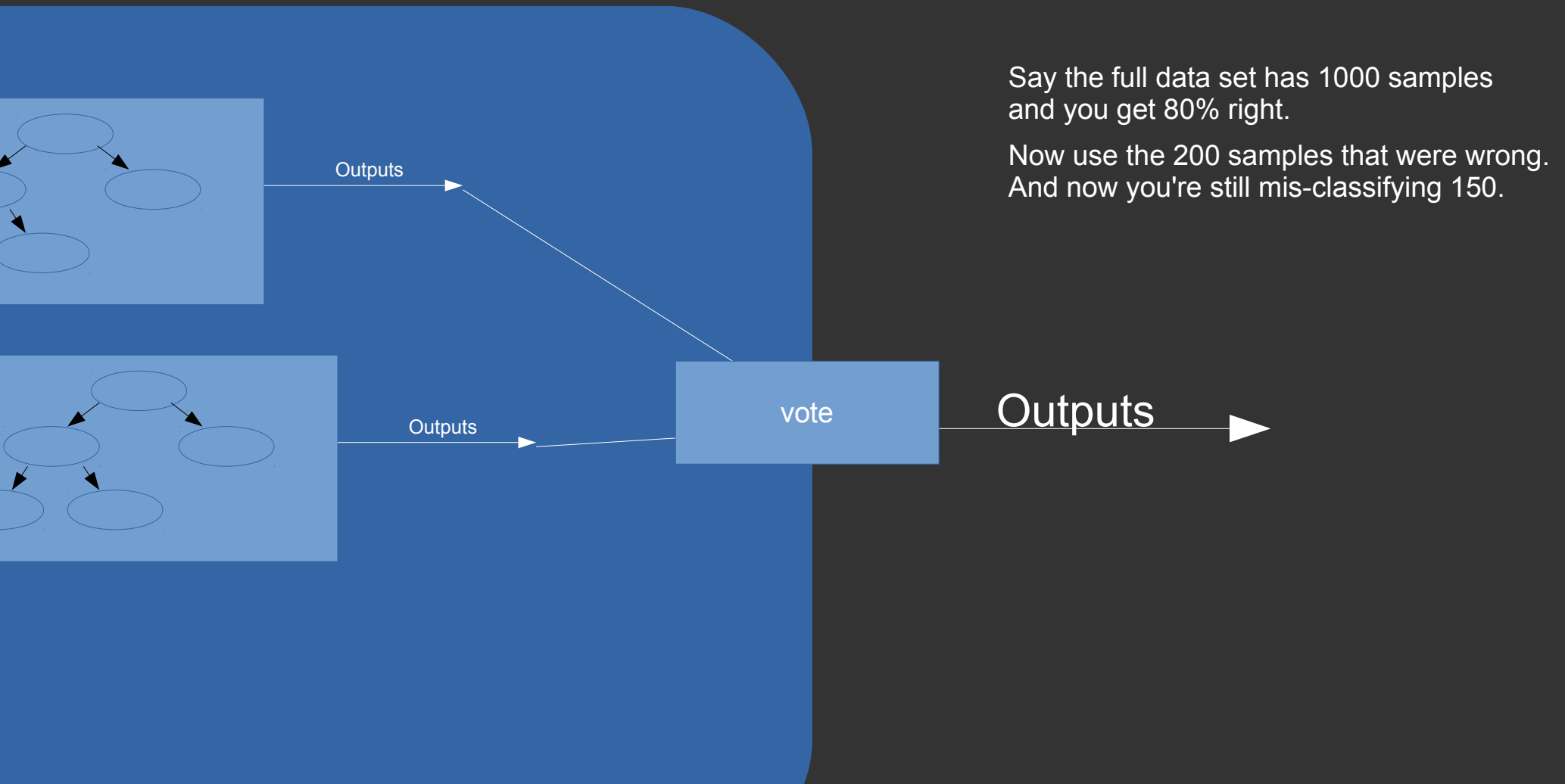
Boosting

Current performance not good enough?



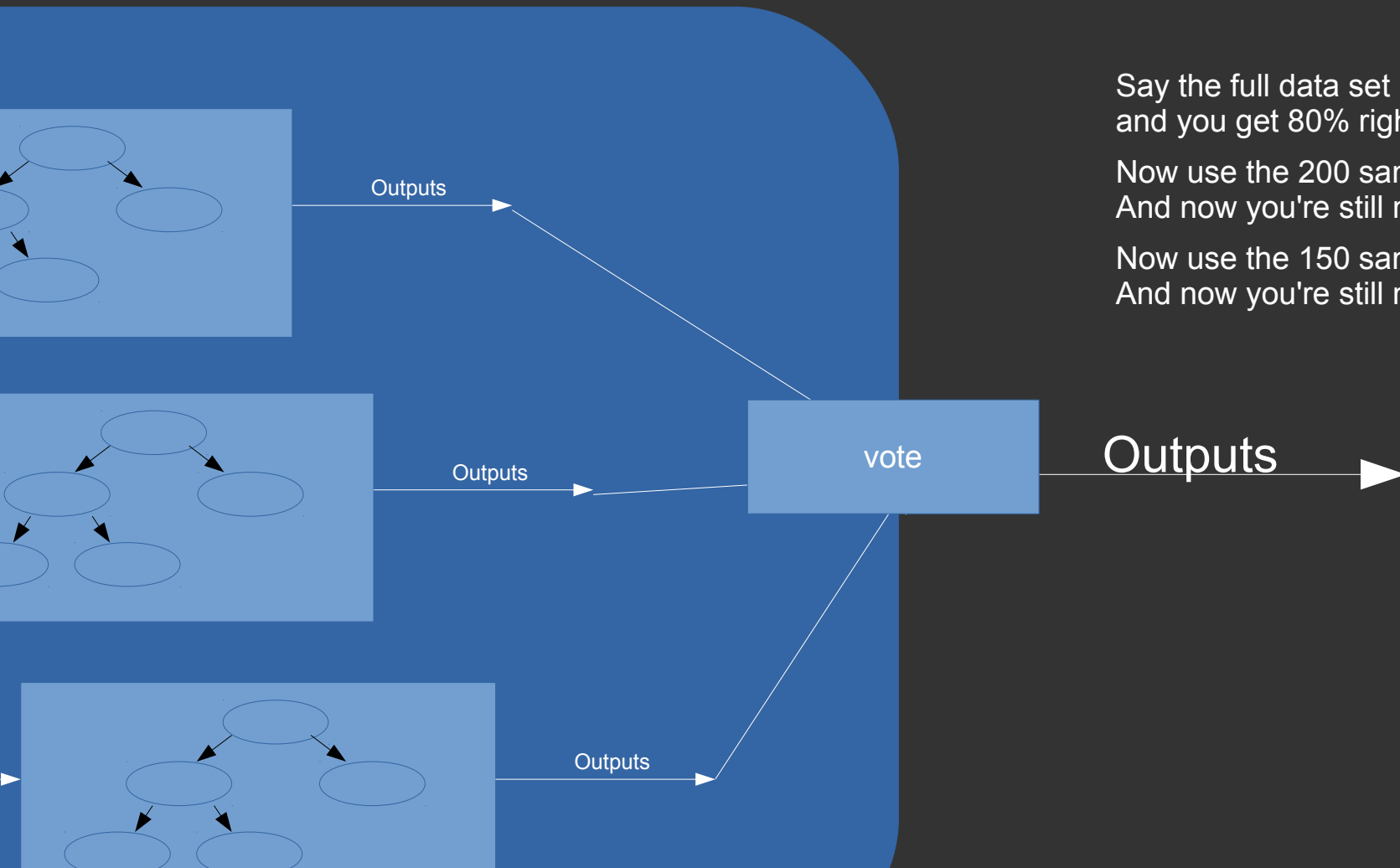
Boosting

Current performance not good enough?
Add another weak learner!
(and only train it on the hard samples)



Boosting

Current performance not good enough?
Add another weak learner!
(and only train it on the hard samples)



Say the full data set has 1000 samples
and you get 80% right.

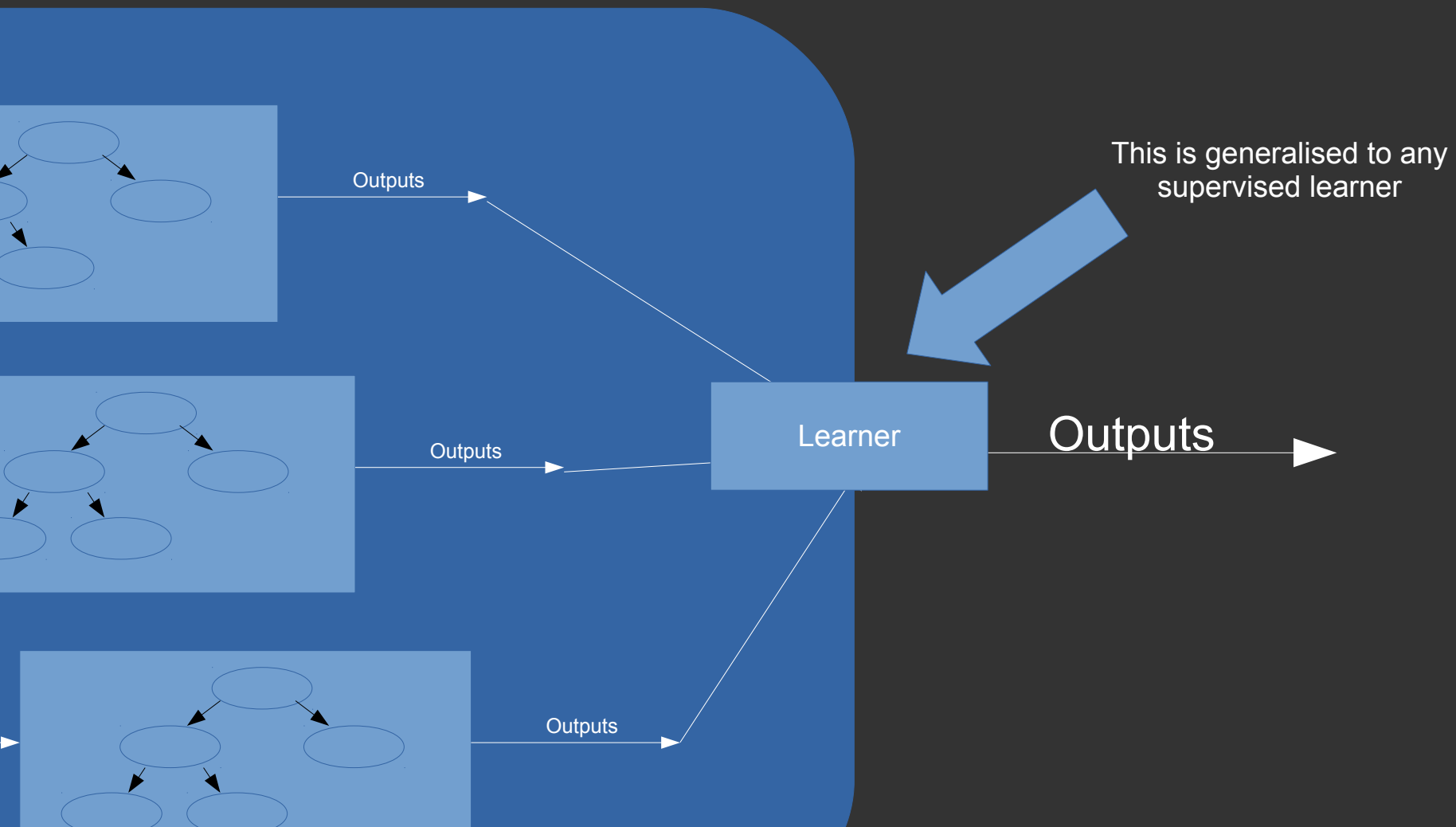
Now use the 200 samples that were wrong.
And now you're still mis-classifying 150.

Now use the 150 samples that were wrong.
And now you're still mis-classifying 120...

Adaptive Boosting

- Called “AdaBoost”
- Instead of dropping samples, weight them less
- Each iteration updates weights and trains a new classifier

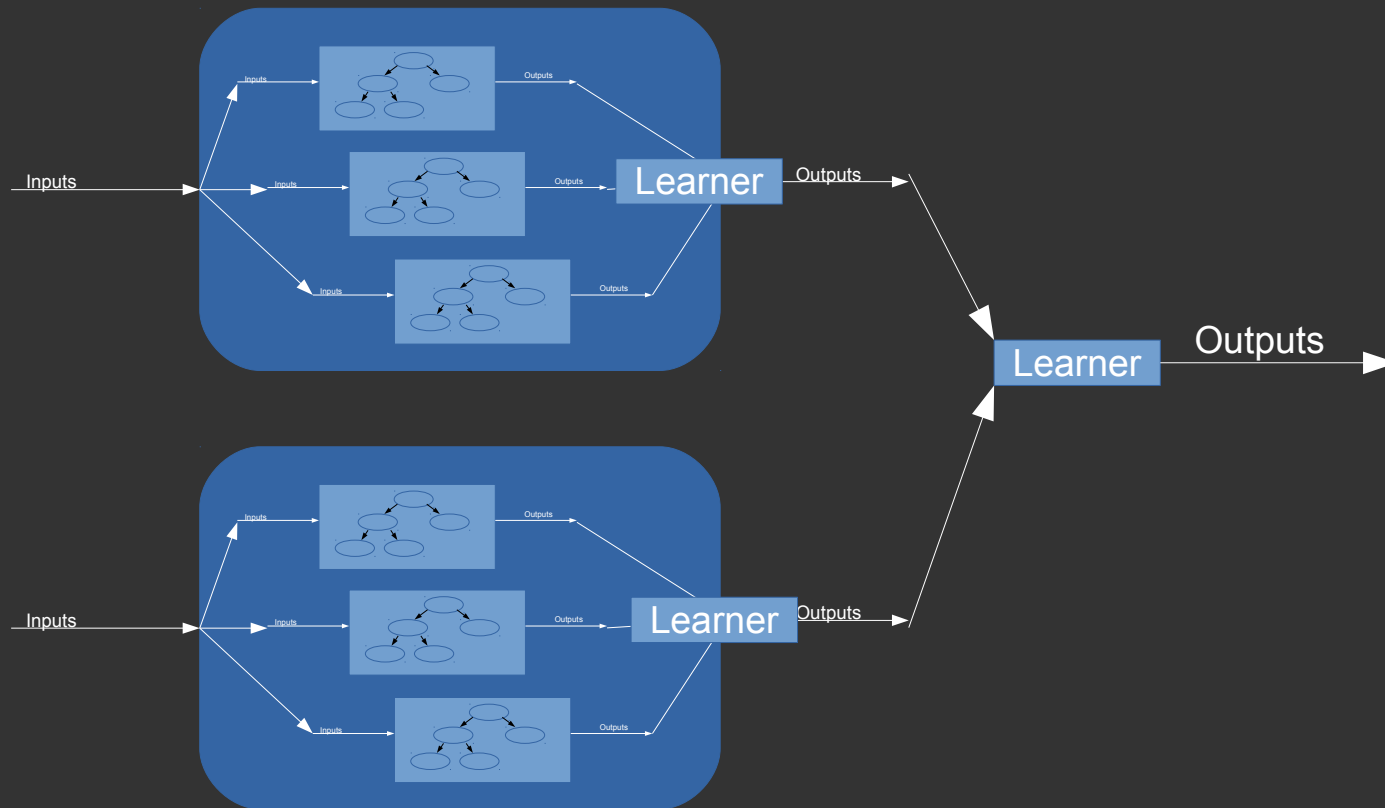
Stacking



Bayesian Model Averaging

- Ensemble of hypotheses
- Each hypothesis is given a weighted vote
- Weighting = $p(\text{data}|\text{hypothesis})$
- Usually computationally infeasible
- Bayesian Model Averaging:
 - Selects hypothesis from sample of hypotheses, rather than full set
- Bayesian Model Combination
 - Same, but select the most likely set of hypotheses

... and more



Back To Python!

- Digits dataset
- Decision Trees
- Random Forests
 - Feature sampling
 - Sample sampling
 - Stumping
 - Random learners
 - and on...