Ensemble Learning

Making a whole greater than any one of its parts

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)

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
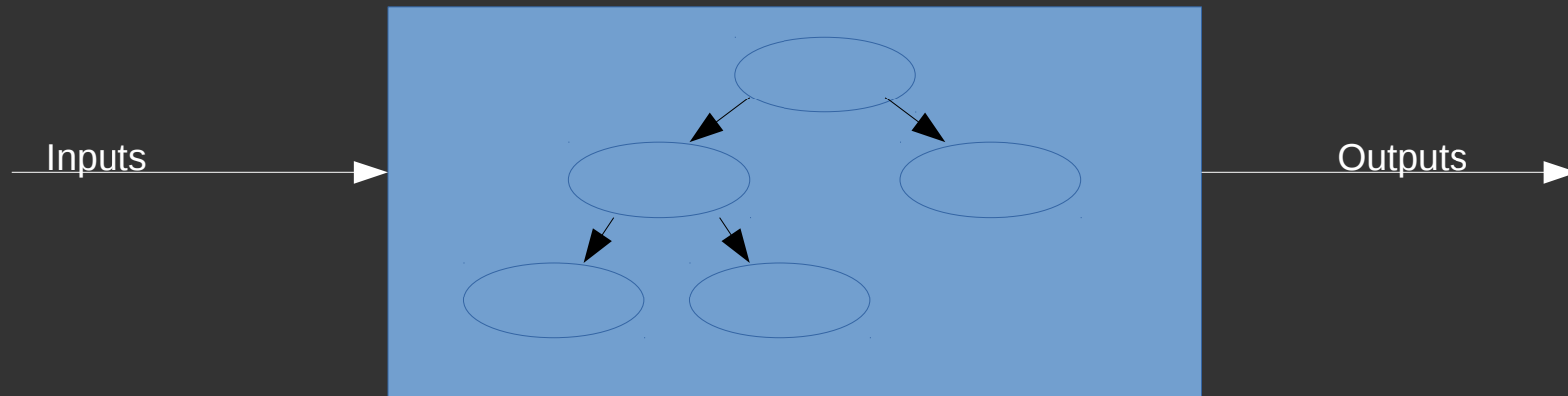allen@GrimmScience.com
@GrimmScientist

A Single Part

The supervised learning 'black box':

Inputs →

Outputs →

A Single Part

The supervised learning 'black box':

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

Inputs →

Outputs →

A Single Part

The supervised learning 'black box':

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

Inputs

Outputs
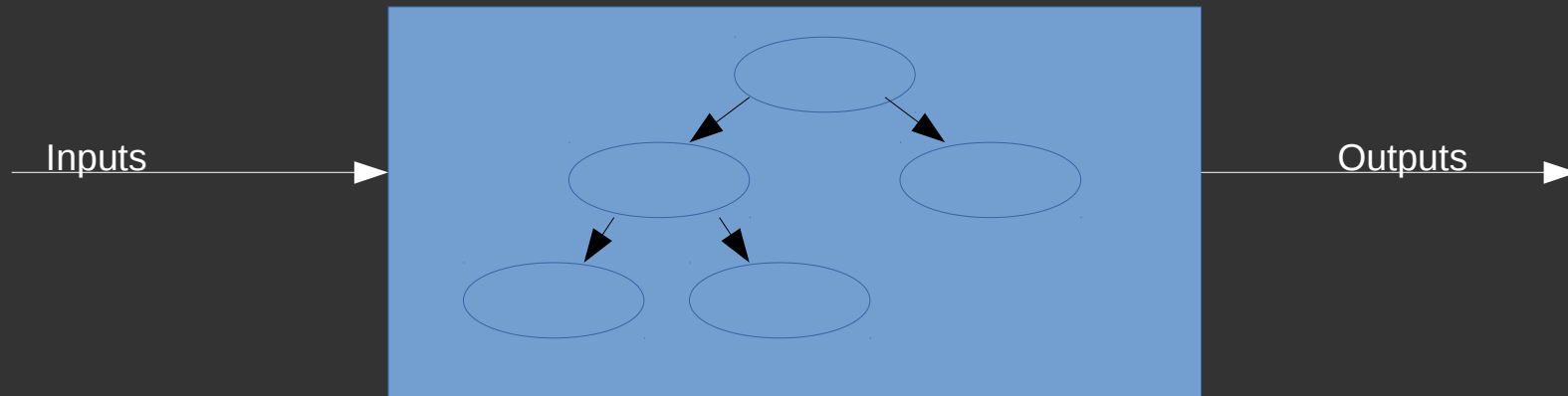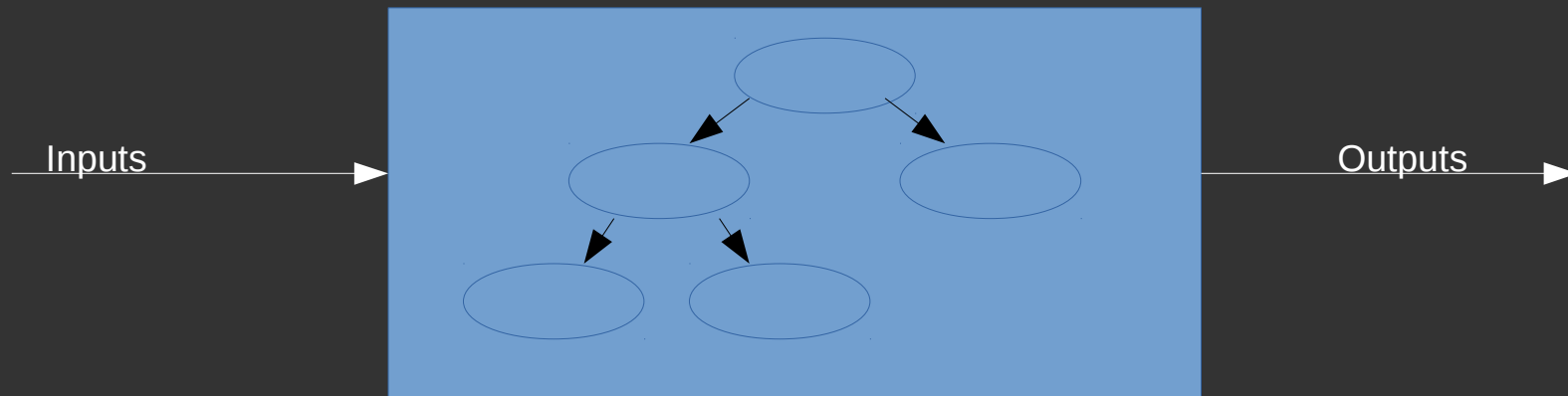
Advantages: Very fast. Intuitive results.

A Single Part

The supervised learning 'black box':

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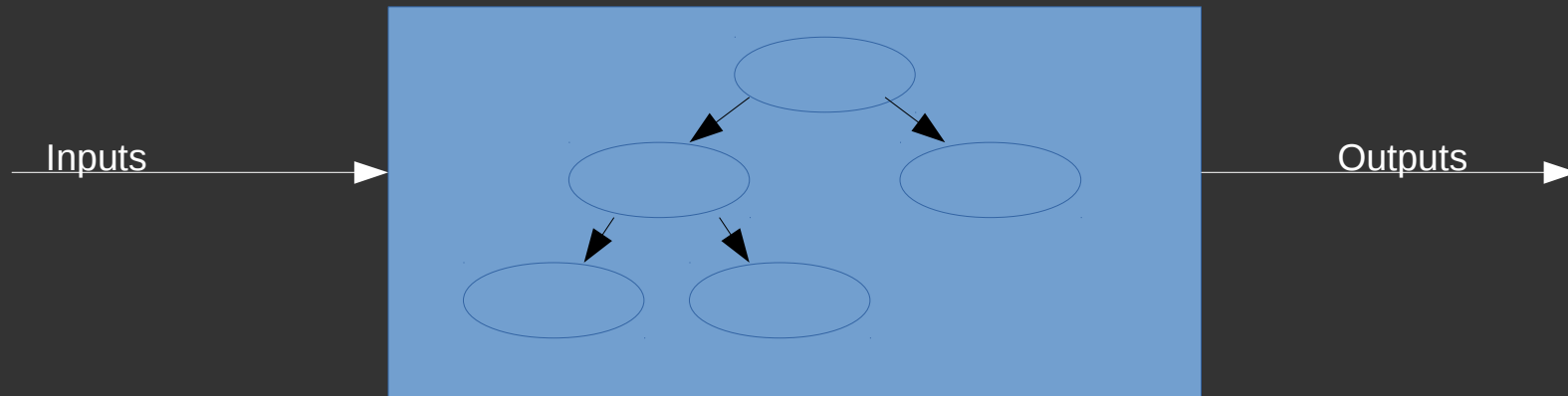
Inputs

Outputs

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...

Inputs →   [blue box with decision tree diagram]   Outputs →
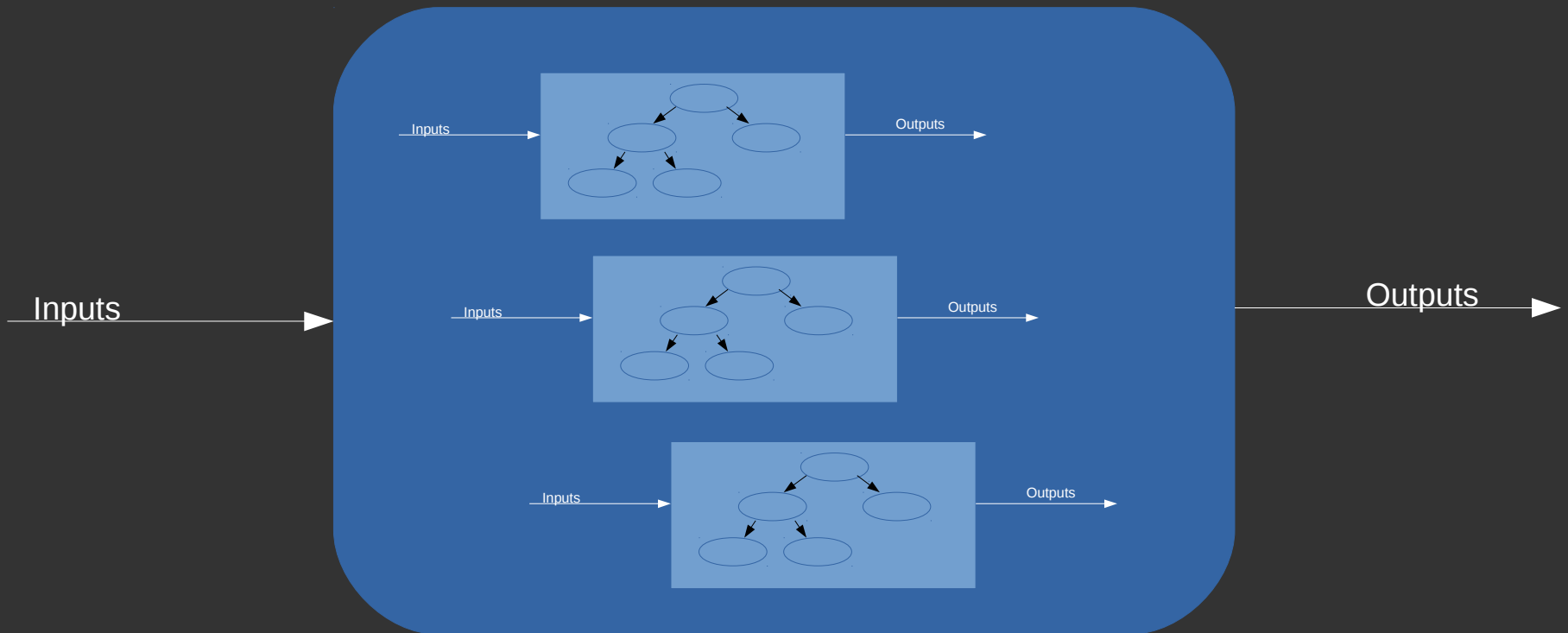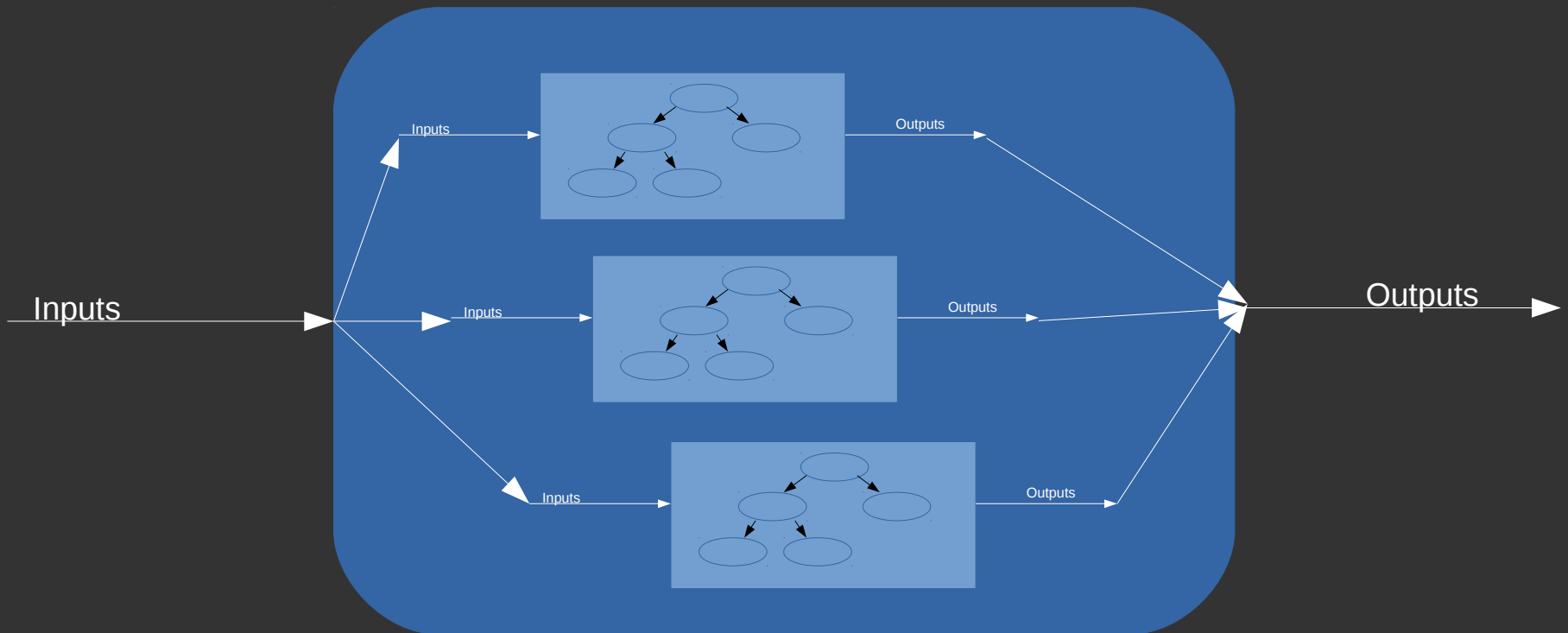
Advantages: Very fast. Intuitive results.

Problem: Too quickly shifts from underfit to overfit

Possible solution: Train a few of them!

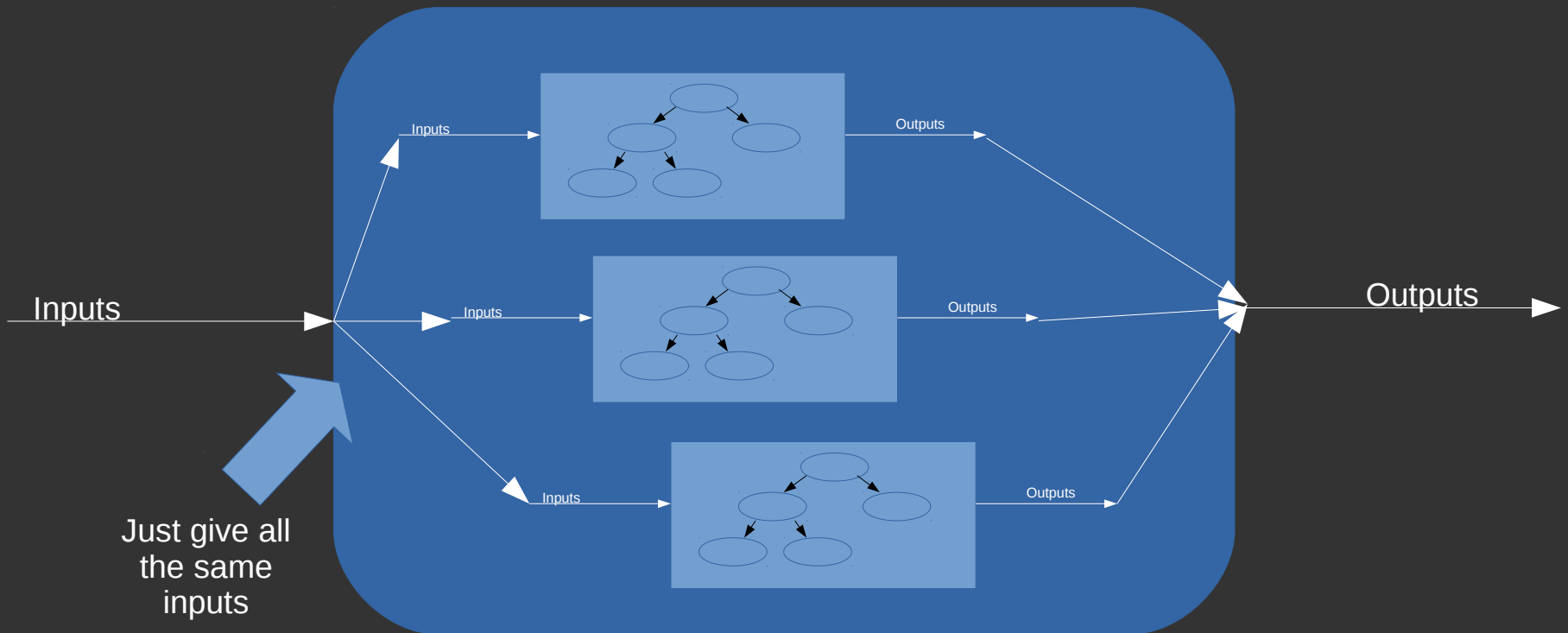# Build the Whole

Build the Whole

Build the Whole

Inputs

Inputs

Inputs

Outputs

Inputs

Outputs

Inputs

Outputs

Outputs

Just give all the same inputs

Build the Whole

Inputs

Inputs

Inputs

Outputs

Inputs

Outputs

Outputs

Outputs

Just give all the same inputs

But what when the outputs disagree?
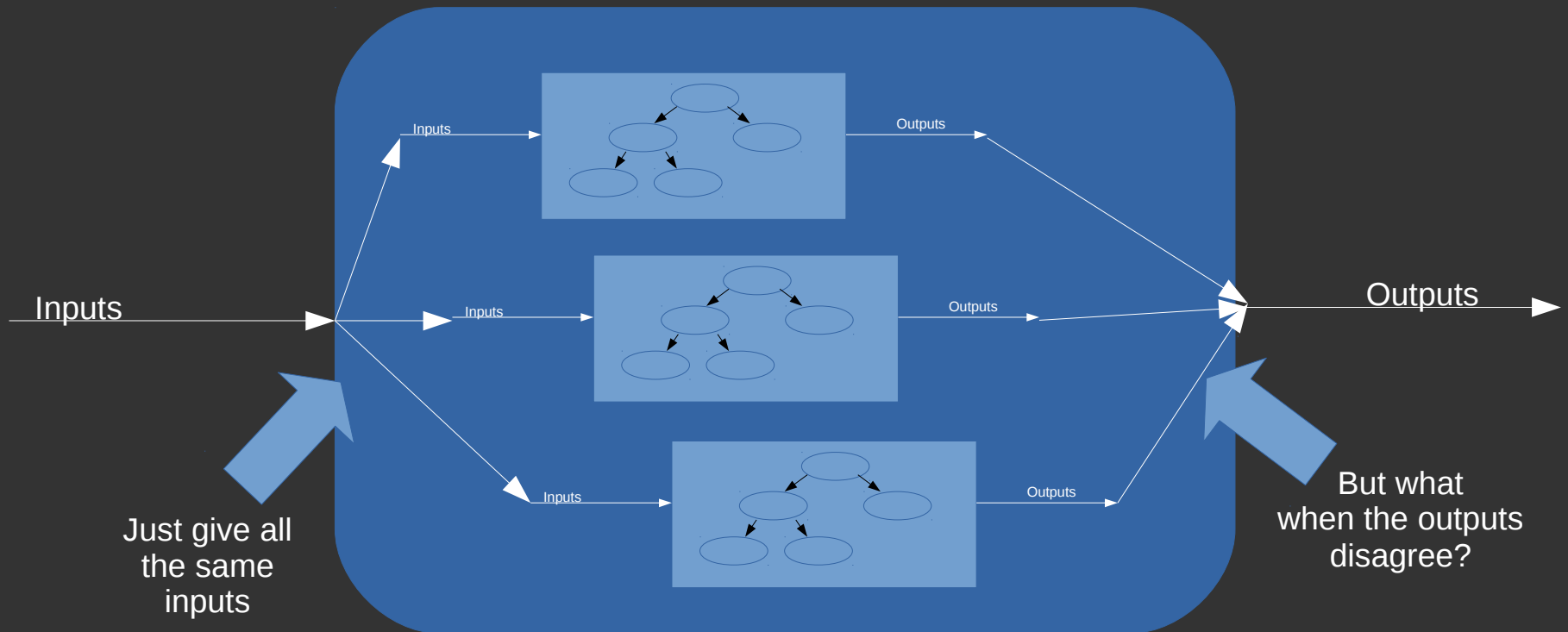
Ensemble Learning Strategies

- Bagging

- Boosting

- Bayes Optimal Classifier
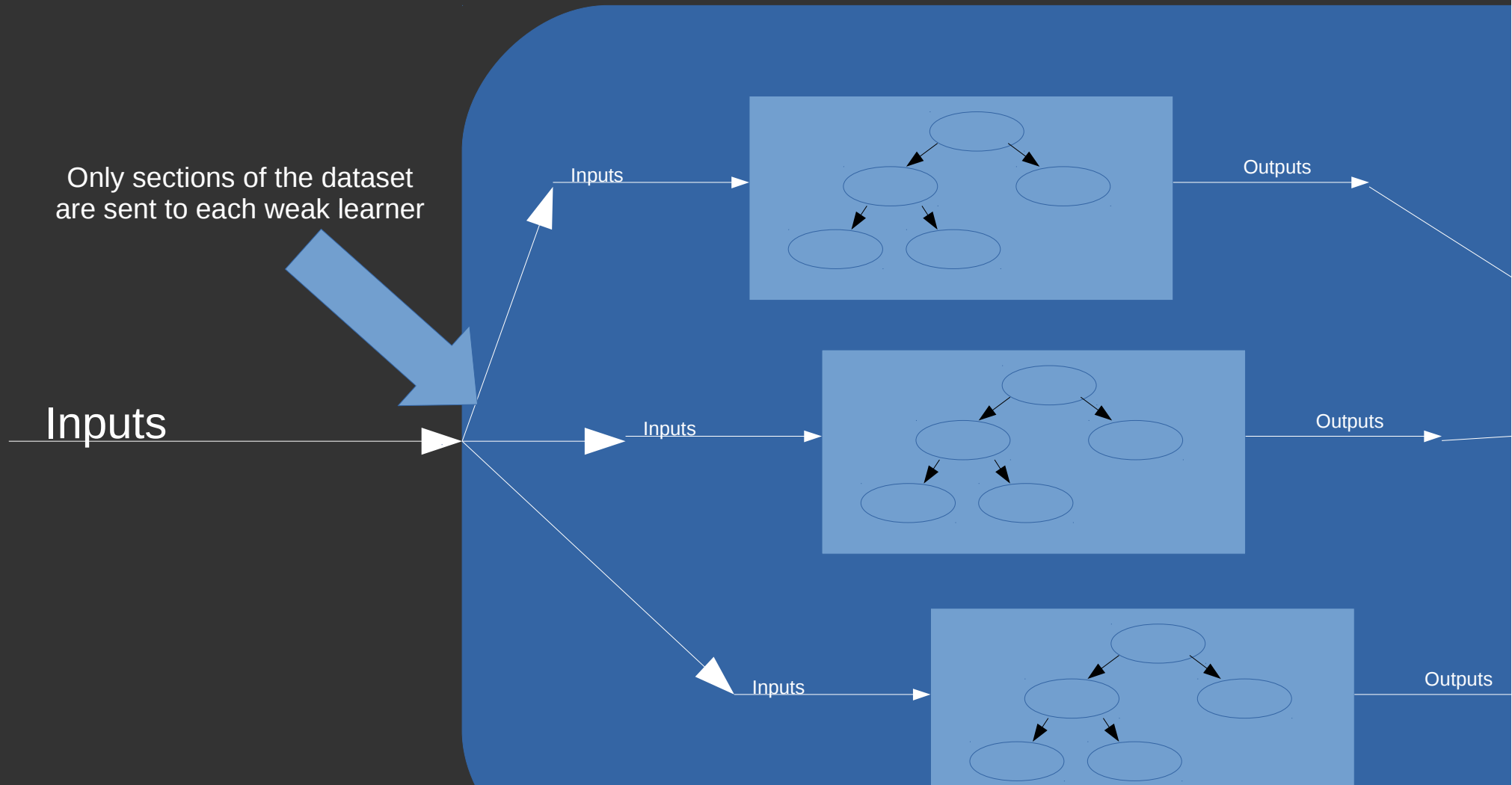
- Stacking

- ...

Ensemble Learning Strategies

- Bagging

- Boosting

- Bayes Optimal Classifier
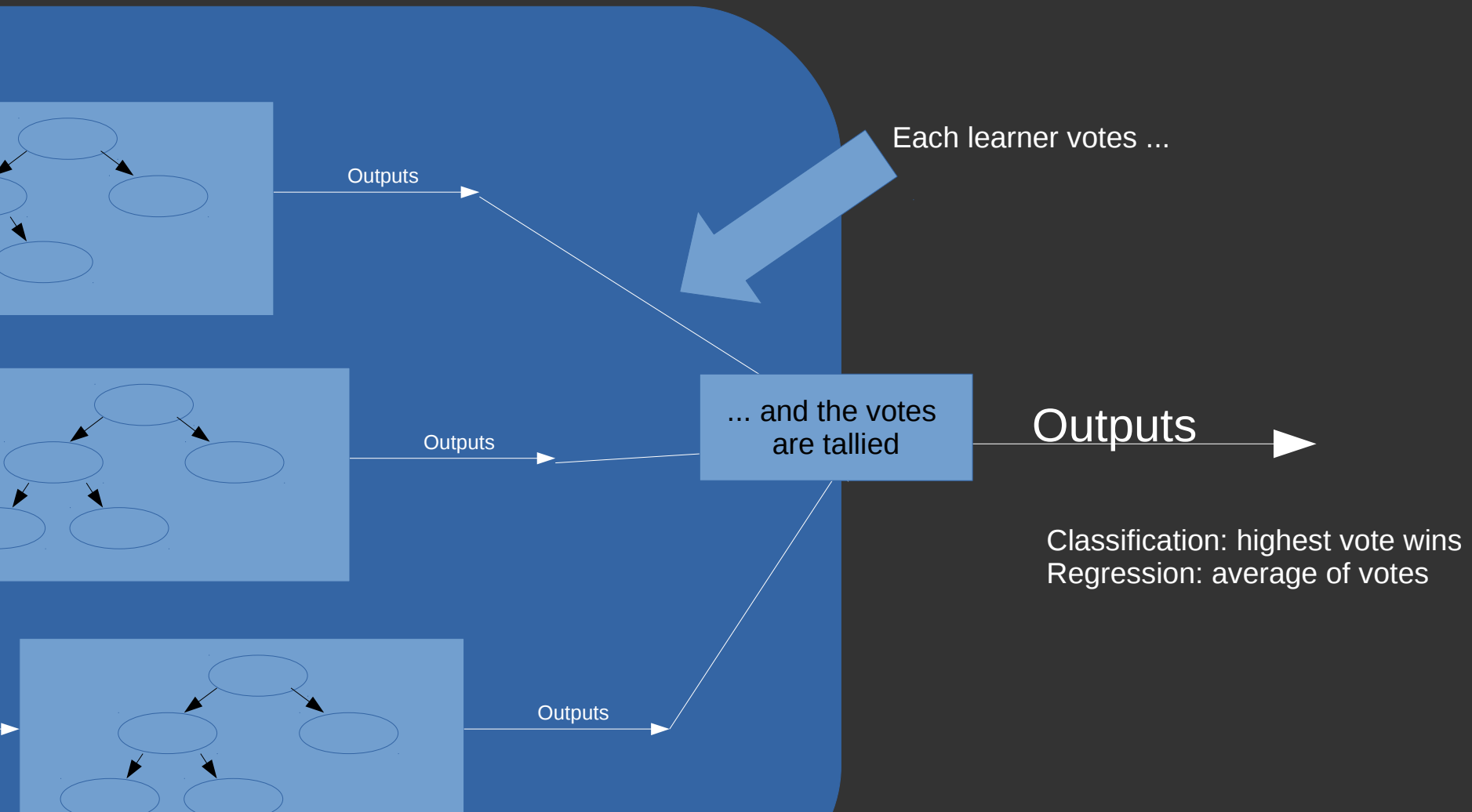
- Stacking

- ...

And then, to random forests!

Bootstrap Aggregating

"Bootstrap" means to sample with replacement

Only sections of the dataset
are sent to each weak learner

Inputs

Inputs

Outputs

Inputs

Outputs

Inputs

Outputs

Bootstrap Aggregating

Usually abbreviated to "Bagging"
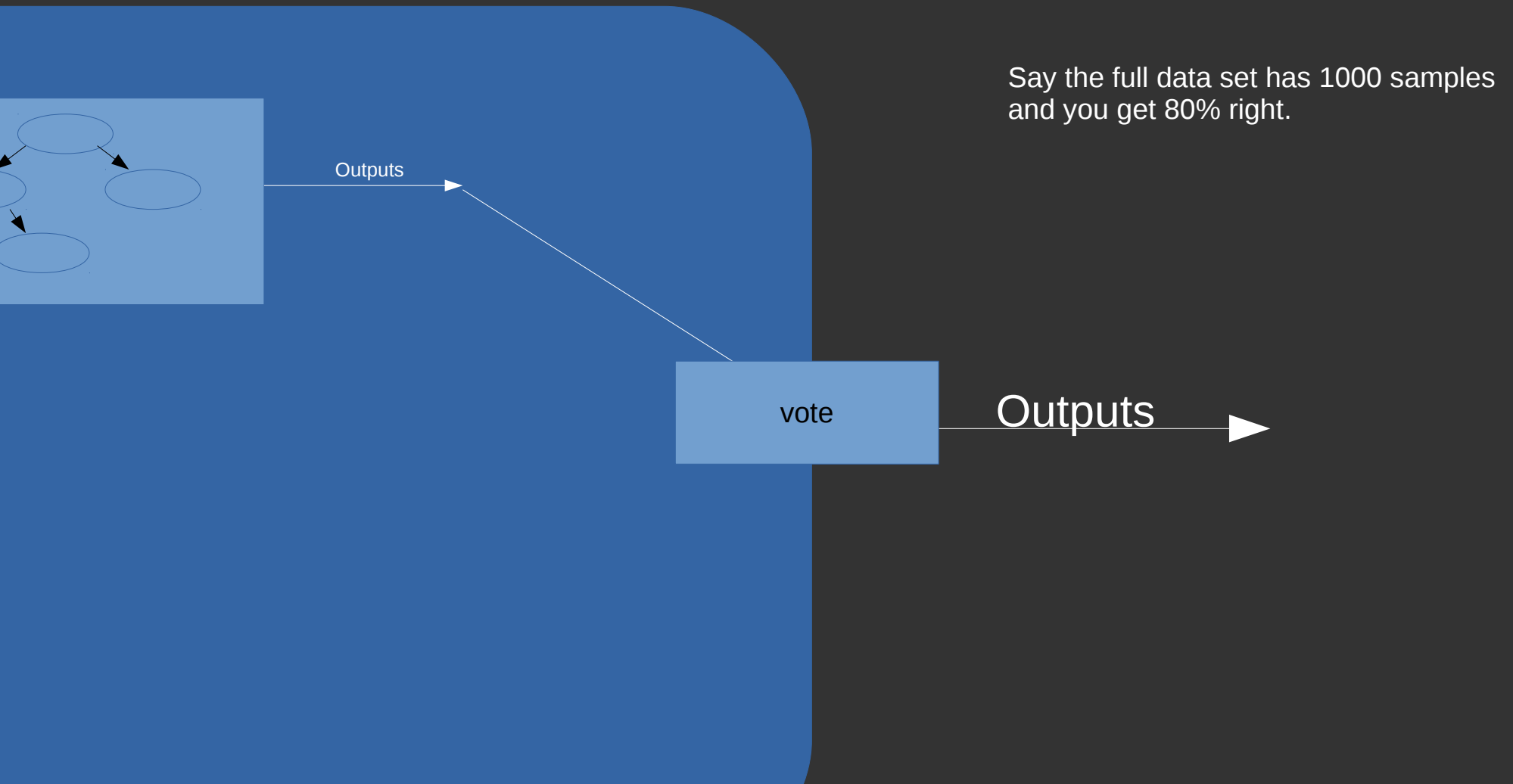
Outputs

Each learner votes ...

Outputs

... and the votes are tallied

Outputs

Outputs

Classification: highest vote wins
Regression: average of votes

Boosting

Current performance not good enough?

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

Outputs

vote

Outputs

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.
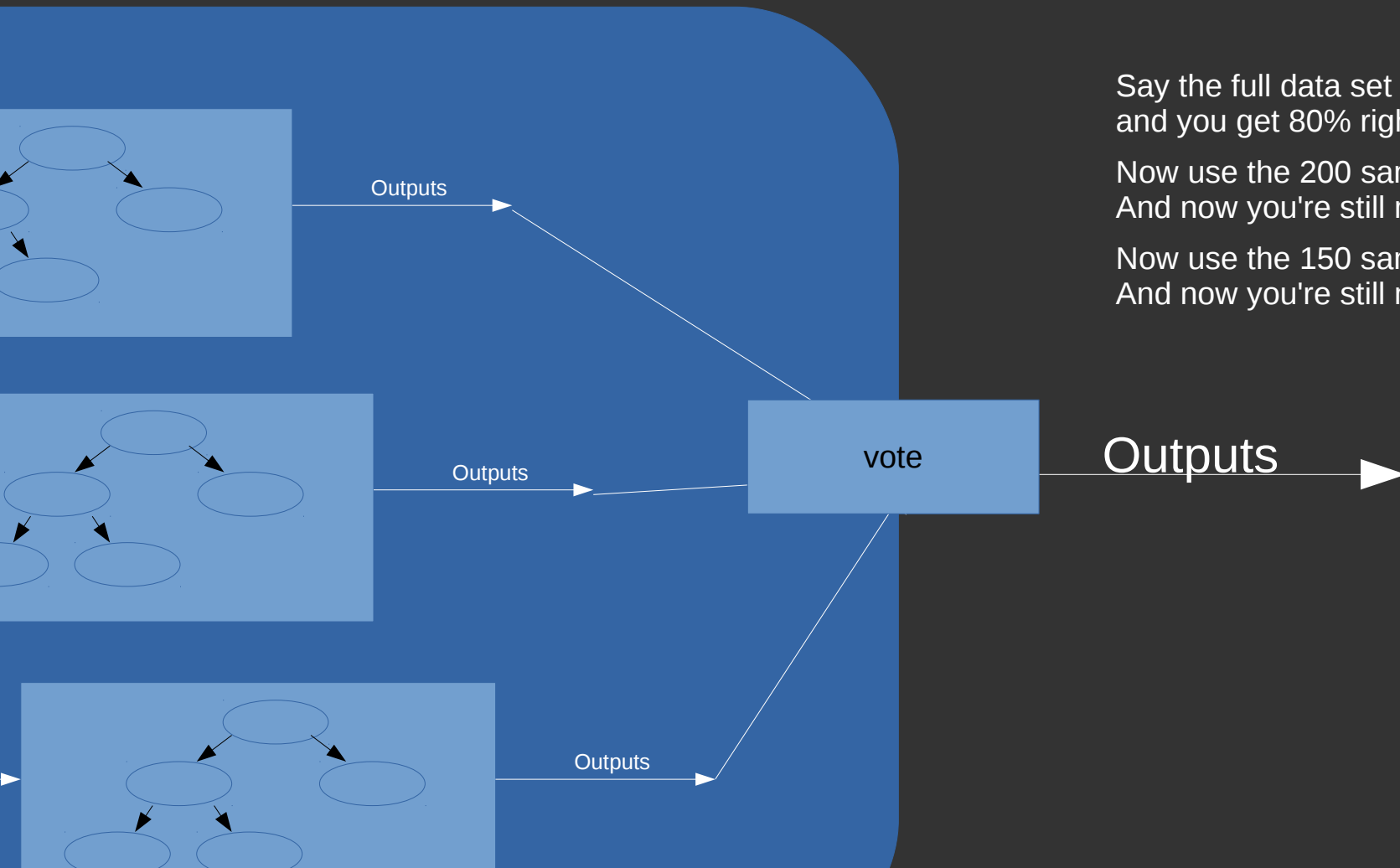And now you're still mis-classifying 150.

Outputs

Outputs

vote

Outputs
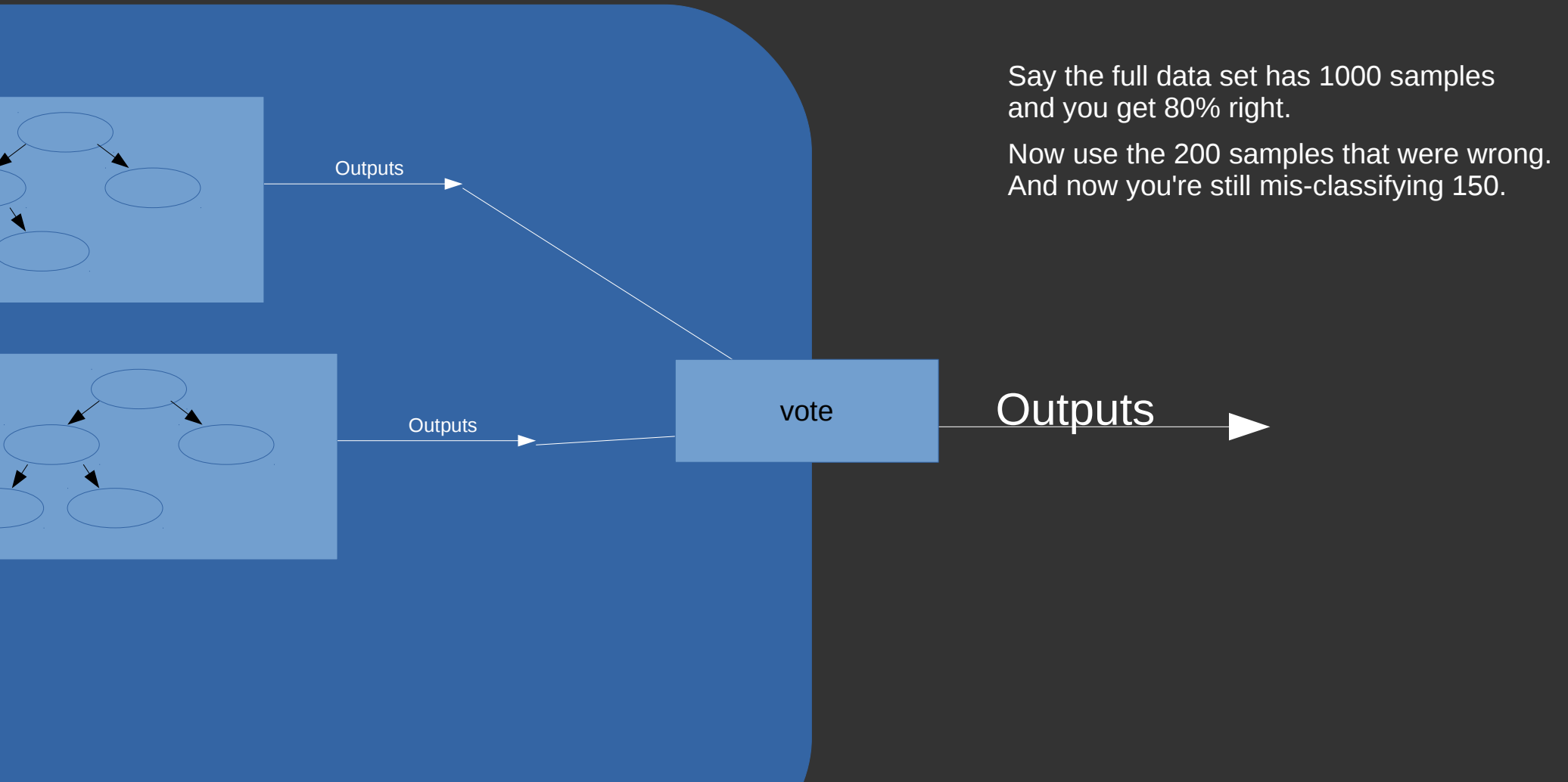
Adaptive Boosting

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...

Outputs

Outputs

Outputs

vote

Outputs

Adaptive Boosting

- Called "AdaBoost"

- Instead of dropping samples, weight them less

- Each iteration updates weights and trains a new classifier

Stacking

Outputs

Outputs

This is generalized to any supervised learner

Learner

Outputs

Outputs
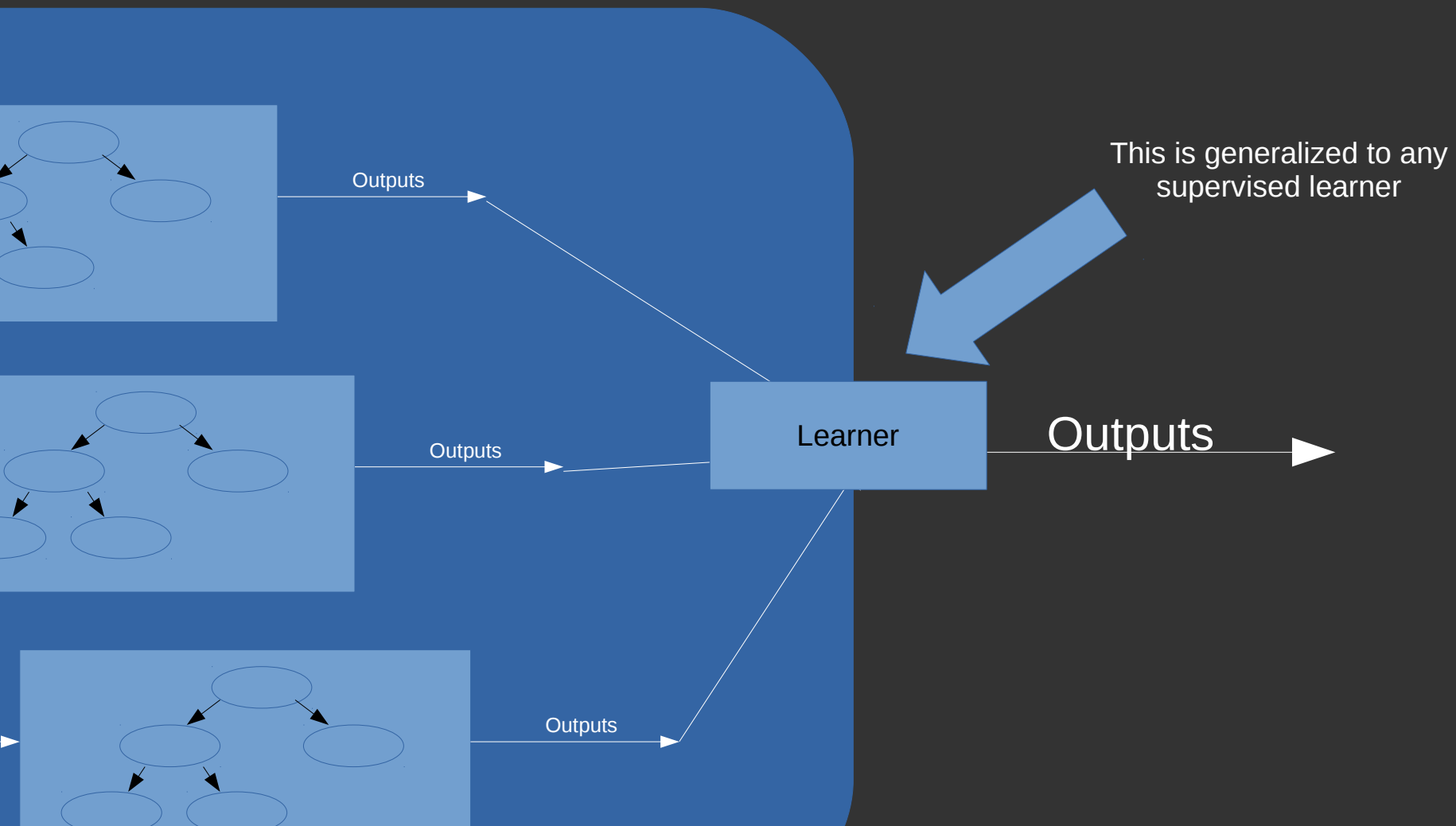
Bayesian Model Averageing

- Ensemble of hypotheses
- Each hypothesis is given a weighted vote
- Weighting = p(data|hypothesis)
- Usually compuationally infeasable
- Bayesian Model Averaging:
  - Selects hypothesis from sample of hypotheses, rather than full set
- Bayesian Model Combination
  - Same, but selecst the most likely set of hypotheses

… and more
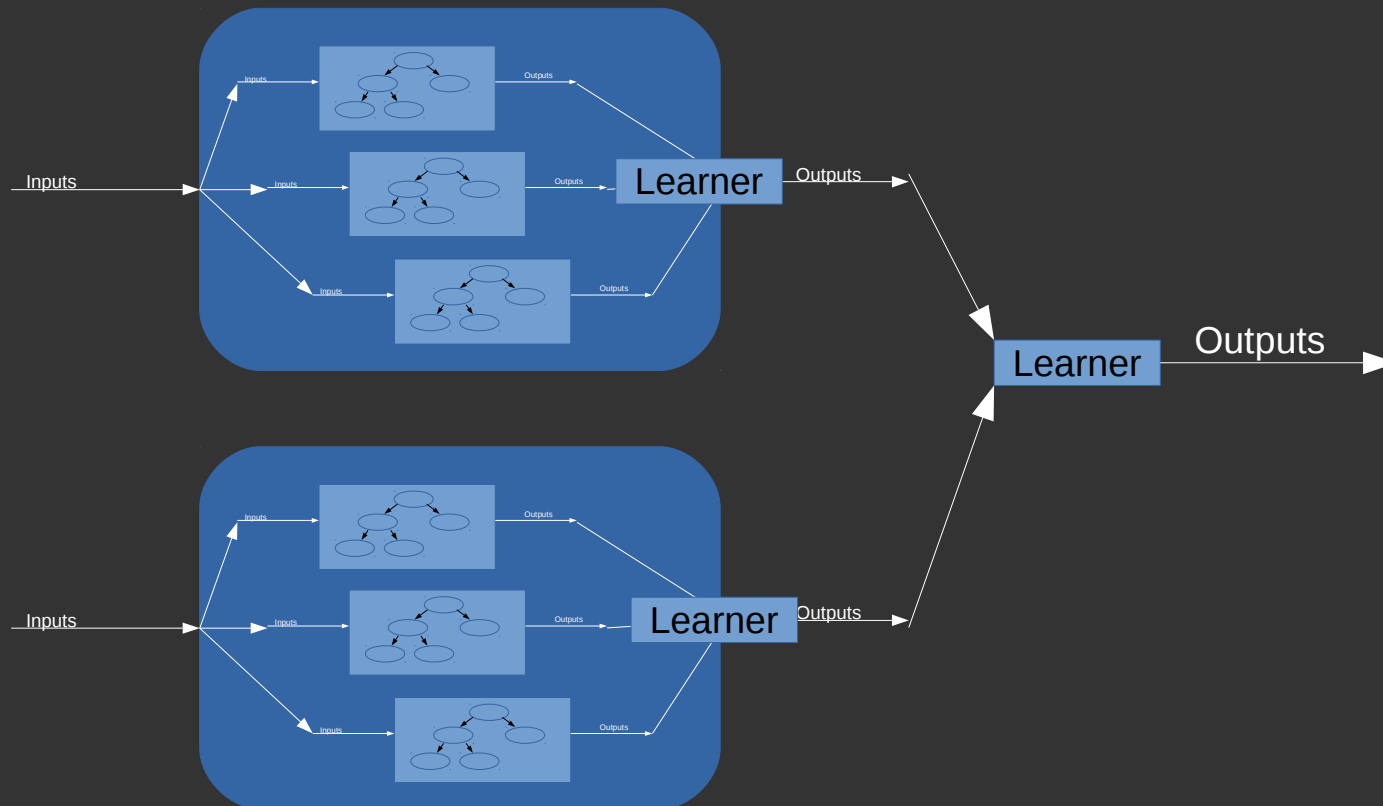
Back To Python!

- Digits dataset

- Decision Trees

- Random Forests

  - Feature sampling

  - Sample sampling

  - Stumping

  - Random learners

  - and on...