Implementing Ensemble Learning Using Boosting Methods

Janani Ravi
CO-FOUNDER, LOONYCORN

www.loonycorn.com

Overview

Using a series of sequential learners in an ensemble

Adaptive boosting or AdaBoost

Gradient Boosting

Hyperparameters tuning of gradient boosting models with early stopping

“Build up your weaknesses until they become your strong points.”

Knute Rockne

Adaptive Boosting

*Ada*ptive *Boost*ing (AdaBoost)

Construct and train models sequentially

Misclassified training points from each model are up-weighted

Next model sees previously misclassified points more often

AdaBoost

Final model output is weighted output of individual predictor outputs

Thus, each individual predictor has weight

Higher weights assigned to more accurate predictors

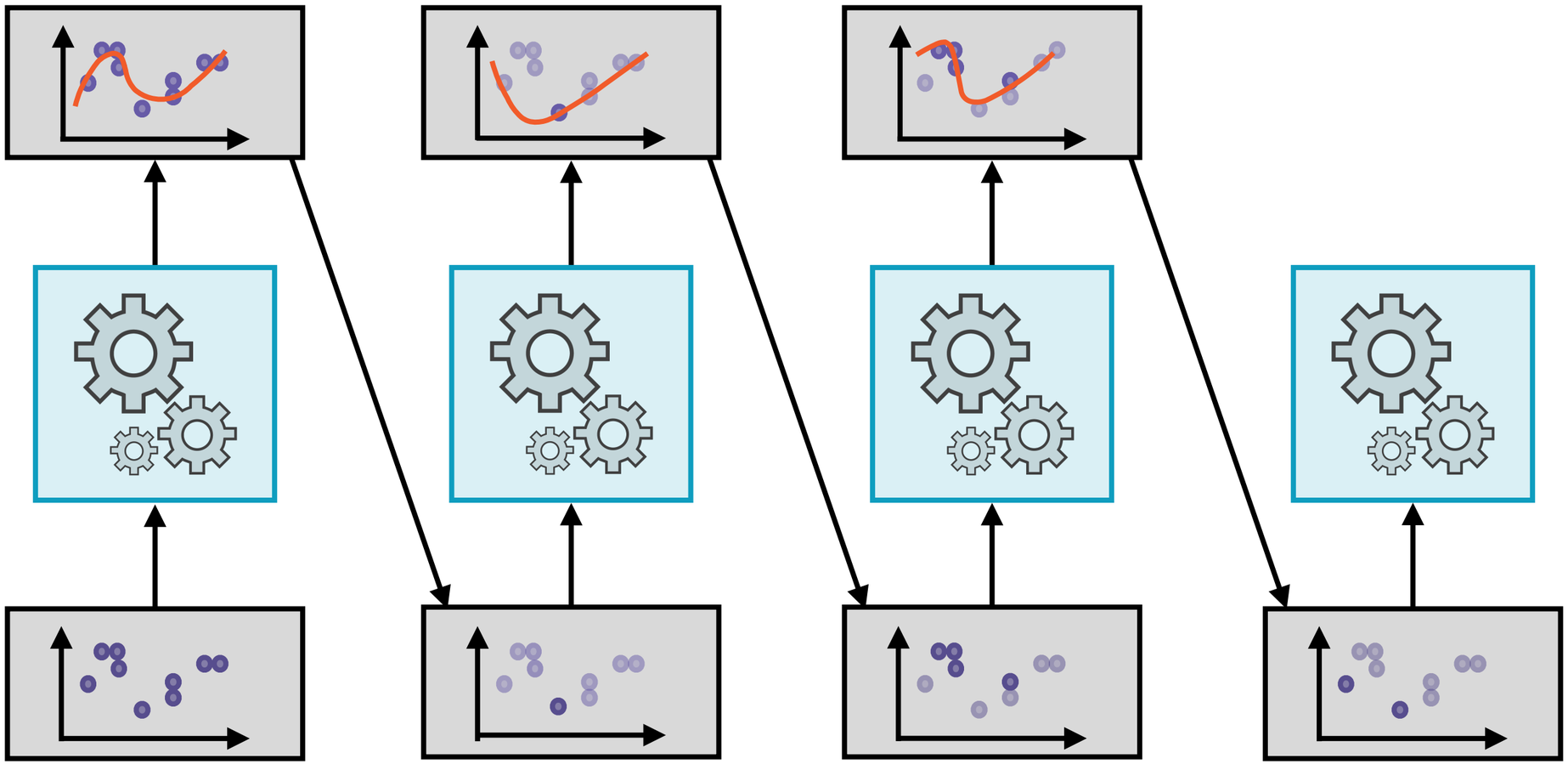
AdaBoost

Predictors that are wrong most often can have negative weights

Predictors that guess at random have weight close to zero

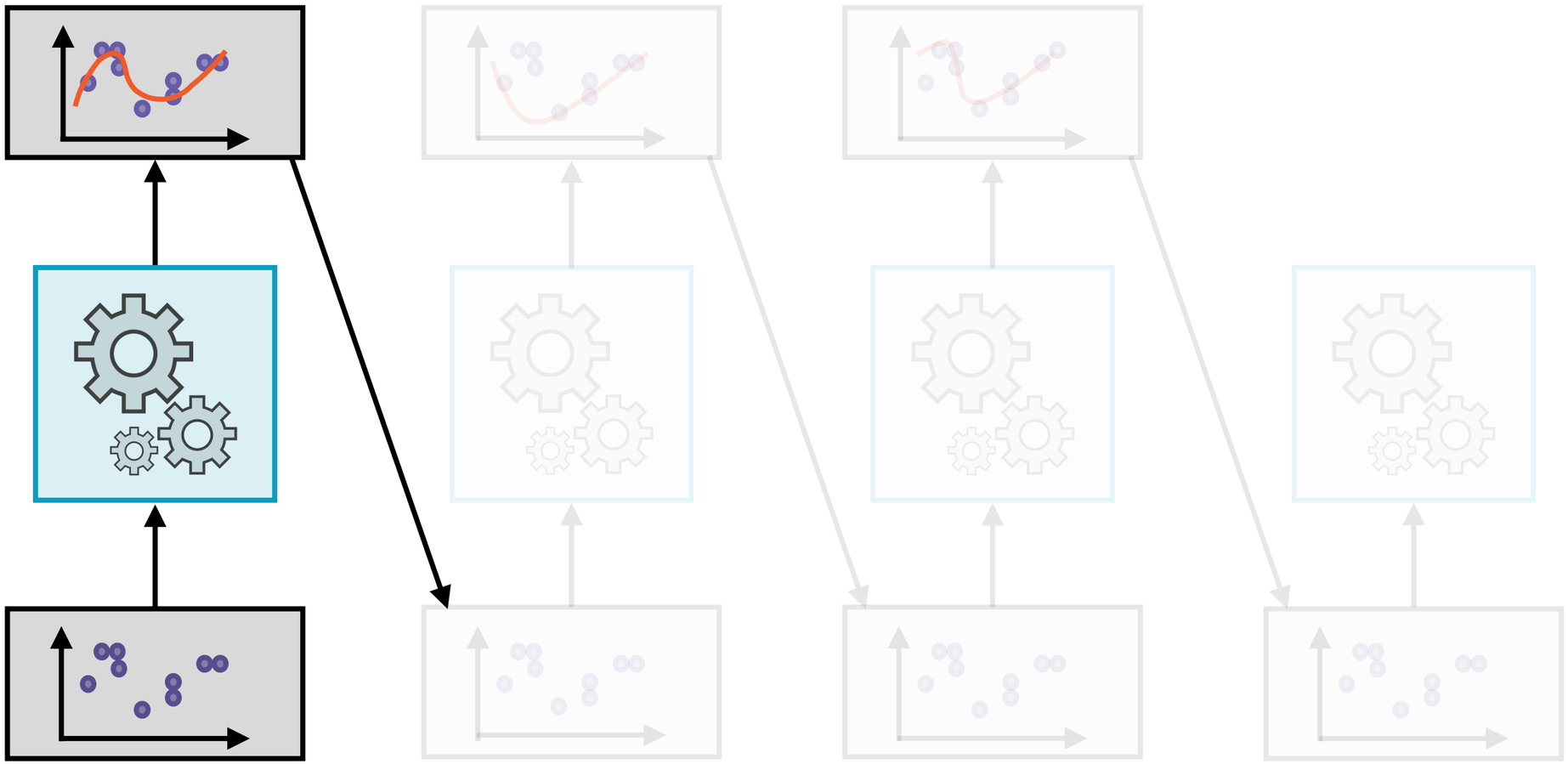
AdaBoost

Sequential training



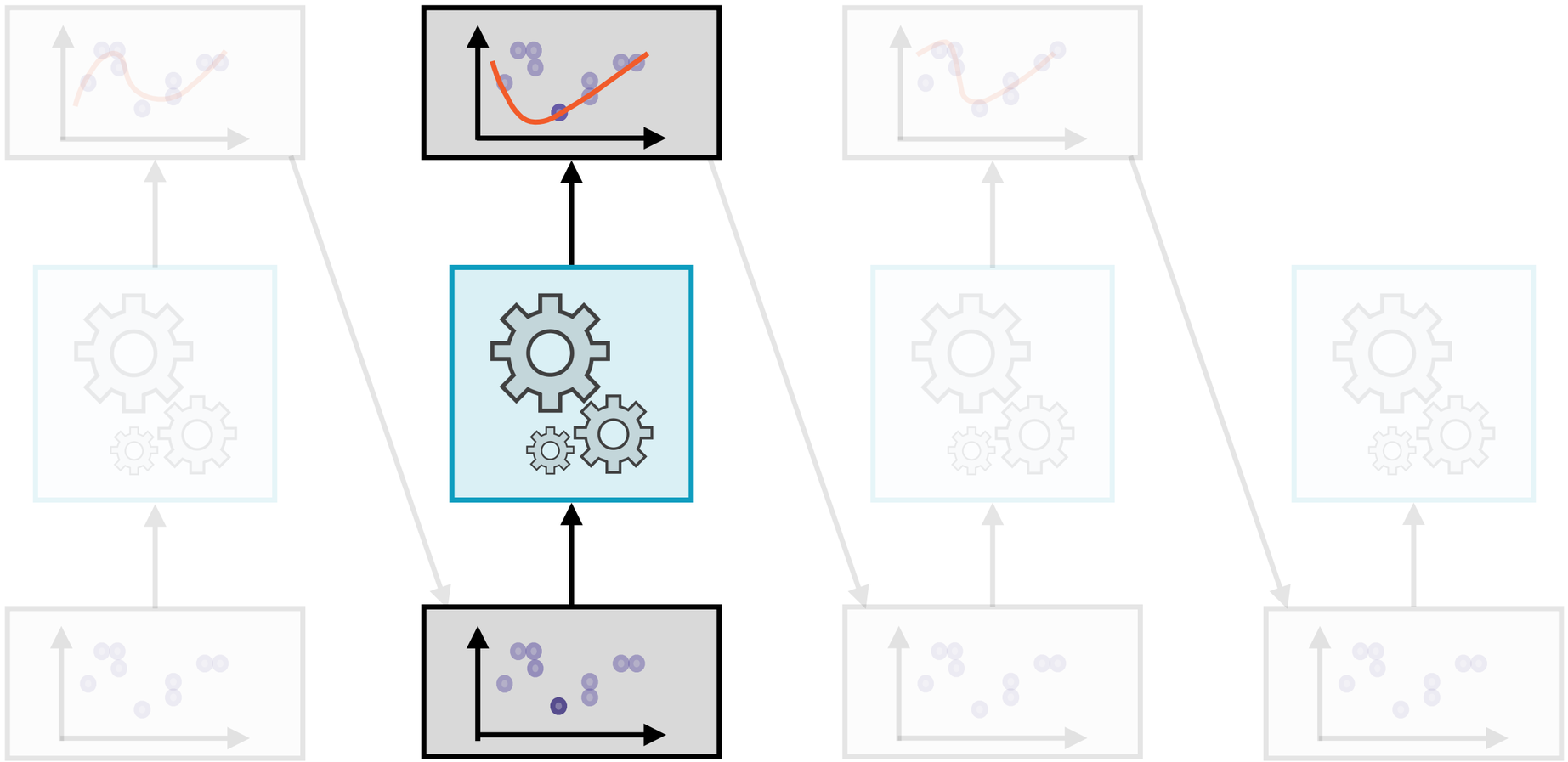
AdaBoost

Train the first predictor in a sequence



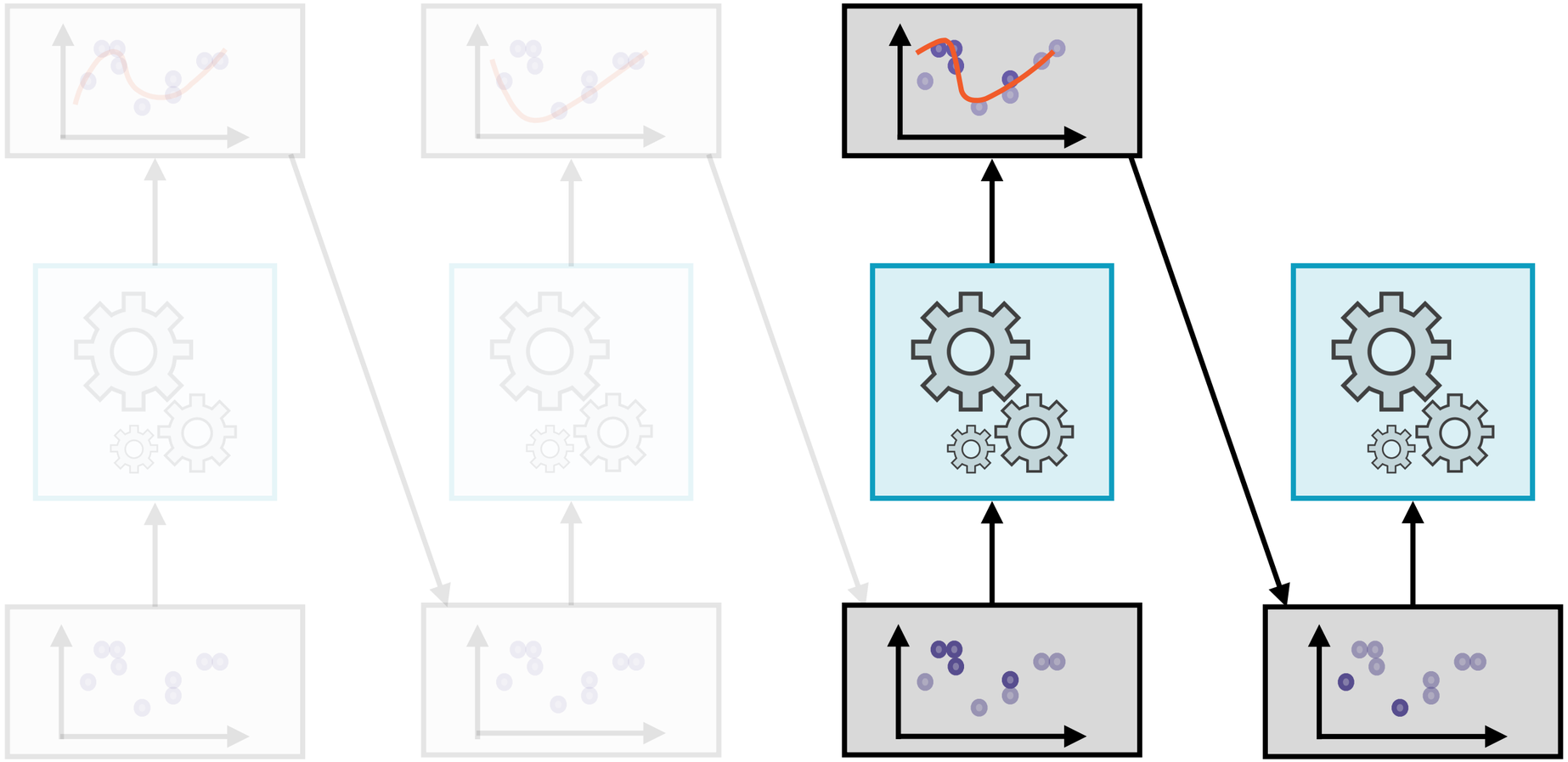
AdaBoost

Increase relative weights of those instances that the first predictor got wrong

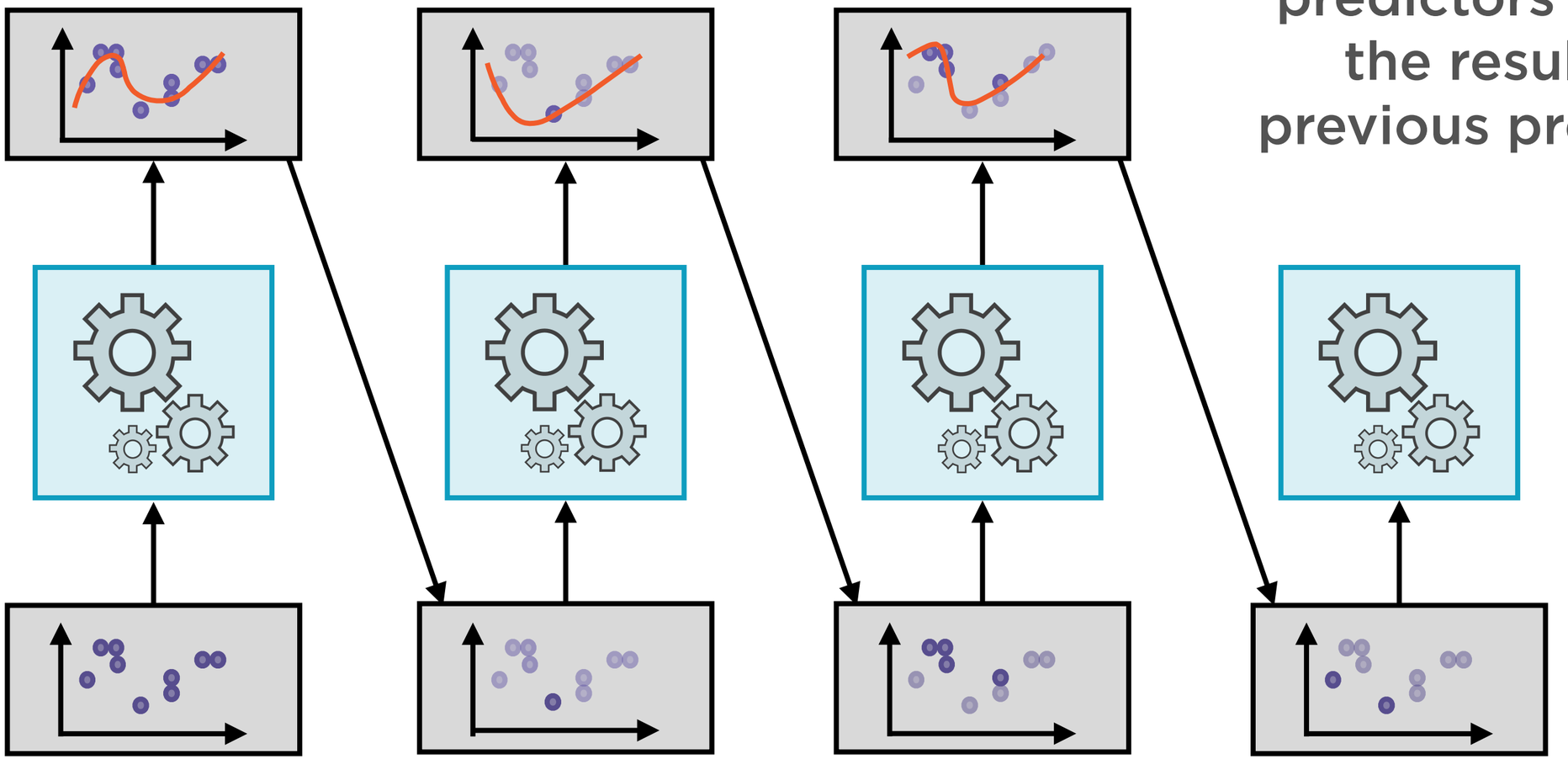


AdaBoost

Increase relative weights of those instances that previous predictors got wrong



AdaBoost



Limited room for parallelism as predictors require the results of previous predictors

Demo

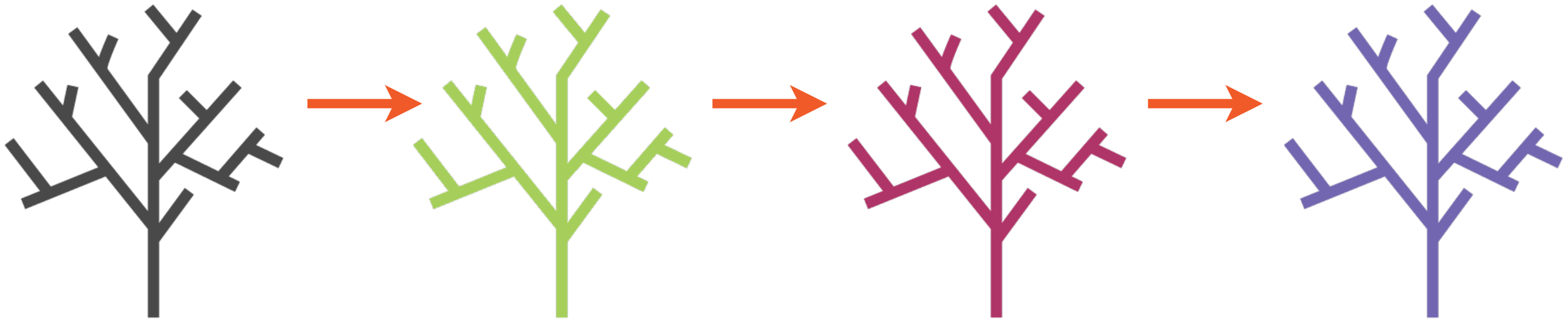
Performing regression using adaptive boosting

Demo

Performing classification using adaptive boosting

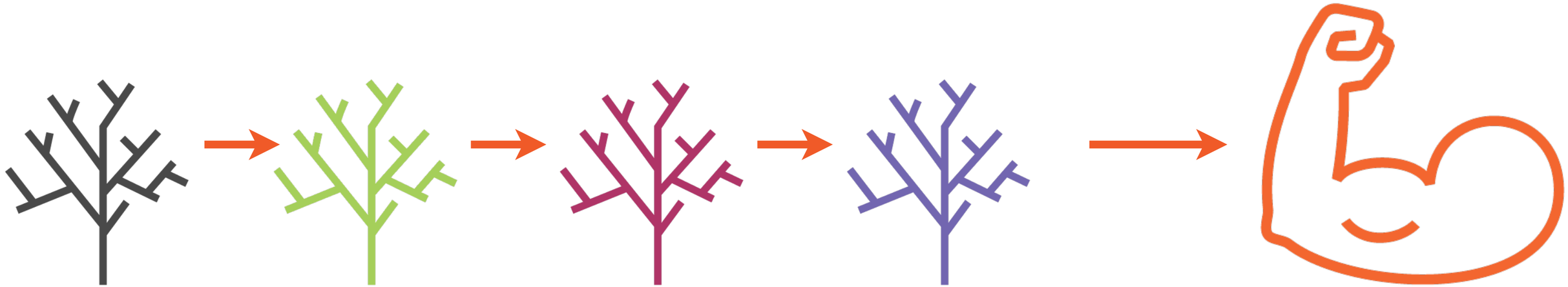
Gradient Boosting

Gradient Boosting



Many machine learning models come together to work on the training data

Gradient Boosting



Many weak learners

Gradient Boosting

Model 1:

$$y = A_1 + B_1x + e_1$$

Model 2:

$$e_1 = A_2 + B_2x + e_2$$

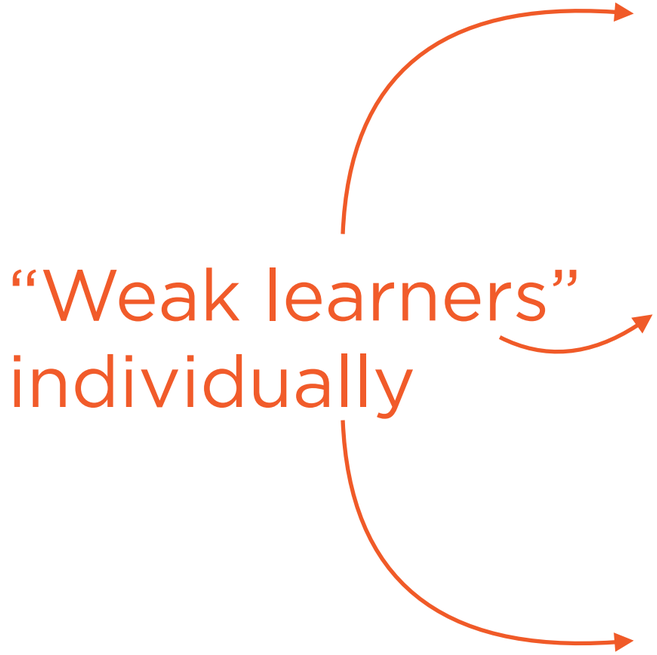
Model 3:

$$e_2 = A_3 + B_3x + e_3$$

Combined Model:

$$y = A_1 + A_2 + A_3 + (B_1 + B_2 + B_3)\,x + e_3$$

Gradient Boosting



Model 1:

$$y = A_1 + B_1x + e_1$$

Model 2:

$$e_1 = A_2 + B_2x + e_2$$

Model 3:

$$e_2 = A_3 + B_3x + e_3$$

Combined Model:

$$y = A_1 + A_2 + A_3 + (B_1 + B_2 + B_3)\,x + e_3$$

Gradient Boosting

Model 1:

$$y = A_1 + B_1x + e_1$$

Model 2:

$$e_1 = A_2 + B_2x + e_2$$

Model 3:

$$e_2 = A_3 + B_3x + e_3$$

"Strong learner" when combined

Combined Model:

$$y = A_1 + A_2 + A_3 + (B_1 + B_2 + B_3)\,x + e_3$$

Gradient Boosting

Model 1:

$$y = A_1 + B_1x + e_1$$

Residuals from Model 1

Model 2:

$$e_1 = A_2 + B_2x + e_2$$

Model 3:

$$e_2 = A_3 + B_3x + e_3$$

Combined Model:

$$y = A_1 + A_2 + A_3 + (B_1 + B_2 + B_3)\,x + e_3$$

Gradient Boosting

Model 1:

$$y = A_1 + B_1x + e_1$$

Residuals from Model 1

What Model 1 fails to learn'

Model 2:

$$e_1 = A_2 + B_2x + e_2$$

Model 3:

$$e_2 = A_3 + B_3x + e_3$$

Combined Model:

$$y = A_1 + A_2 + A_3 + (B_1 + B_2 + B_3)\,x + e_3$$

Gradient Boosting

Model 1:

$$y = A_1 + B_1x + e_1$$

Residuals from Model 1

Model 2:

$$e_1 = A_2 + B_2x + e_2$$

What Model 1 fails to learn'

Model 3:

$$e_2 = A_3 + B_3x + e_3$$

Learn in Model 2

Combined Model:

$$y = A_1 + A_2 + A_3 + (B_1 + B_2 + B_3)\,x + e_3$$

Gradient Boosting

Model 1:

$$y = A_1 + B_1x + e_1$$

Focuses on what previous model failed to learn

Model 2:

$$\mathbf{e_1 = A_2 + B_2x + e_2}$$

Model 3:

$$e_2 = A_3 + B_3x + e_3$$

Combined Model:

$$y = A_1 + A_2 + A_3 + (B_1 + B_2 + B_3)\,x + e_3$$

Gradient Boosting

Model 1:

$$y = A_1 + B_1x + e_1$$

Model 2:

$$e_1 = A_2 + B_2x + \mathbf{e_2}$$

Residuals from Model 2

Model 3:

$$e_2 = A_3 + B_3x + e_3$$

Combined Model:

$$y = A_1 + A_2 + A_3 + (B_1 + B_2 + B_3)\,x + e_3$$

Gradient Boosting

Model 1:

$$y = A_1 + B_1x + e_1$$

Model 2:

$$e_1 = A_2 + B_2x + e_2$$

Residuals from Model 2

Model 3:

$$e_2 = A_3 + B_3x + e_3$$

What Model 2 fails to learn'

Combined Model:

$$y = A_1 + A_2 + A_3 + (B_1 + B_2 + B_3)\,x + e_3$$

Gradient Boosting

Model 1:

$$y = A_1 + B_1x + e_1$$

Model 2:

$$e_1 = A_2 + B_2x + e_2$$

Focuses on what previous model failed to learn

Model 3:

$$e_2 = A_3 + B_3x + e_3$$

Combined Model:

$$y = A_1 + A_2 + A_3 + (B_1 + B_2 + B_3)\,x + e_3$$

Gradient Boosting

Model 1:

$$y = A_1 + B_1x + e_1$$

Model 2:

$$e_1 = A_2 + B_2x + e_2$$

Model 3:

$$e_2 = A_3 + B_3x + \mathbf{e_3}$$

These residuals are now unlearnt

Combined Model:

$$y = A_1 + A_2 + A_3 + (B_1 + B_2 + B_3)\,x + \mathbf{e_3}$$

Gradient Boosting

Model 1:

$$y = A_1 + B_1x + e_1$$

Model 2:

$$e_1 = A_2 + B_2x + e_2$$

Model 3:

$$e_2 = A_3 + B_3x + \mathbf{e_3}$$

Only these residuals are now unlearnt

Combined Model:

$$\mathbf{y = A_1 + A_2 + A_3 + (B_1 + B_2 + B_3)\,x + e_3}$$

Gradient Boosting

Model 1:

$$y = A_1 + B_1x + e_1$$

Model 2:

$$e_1 = A_2 + B_2x + e_2$$

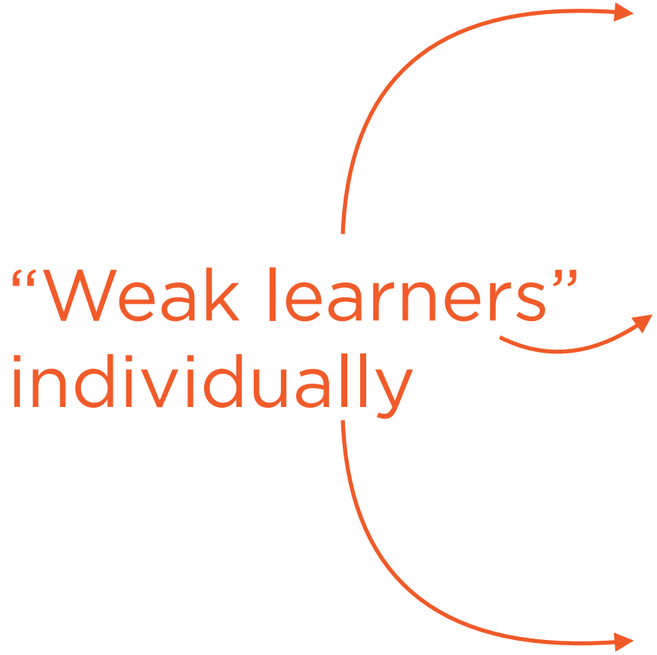
Model 3:

$$e_2 = A_3 + B_3x + e_3$$

Combined Model:

$$y = A_1 + A_2 + A_3 + (B_1 + B_2 + B_3)\,x + e_3$$

Gradient Boosting



Model 1:

$$y = A_1 + B_1x + e_1$$

Model 2:

$$e_1 = A_2 + B_2x + e_2$$

Model 3:

$$e_2 = A_3 + B_3x + e_3$$

Combined Model:

$$y = A_1 + A_2 + A_3 + (B_1 + B_2 + B_3)\,x + e_3$$

Gradient Boosting

"Ensemble" of learners

Model 1:

$$y = A_1 + B_1x + e_1$$

Model 2:

$$e_1 = A_2 + B_2x + e_2$$

Model 3:

$$e_2 = A_3 + B_3x + e_3$$

Combined Model:

$$y = A_1 + A_2 + A_3 + (B_1 + B_2 + B_3)\,x + e_3$$

Gradient Boosting

Model 1:

$$y = A_1 + B_1x + e_1$$

In practice:
100-200 weak learners, each learning from previous mistakes

Model 2:

$$e_1 = A_2 + B_2x + e_2$$

Model 3:

$$e_2 = A_3 + B_3x + e_3$$

Combined Model:

$$y = A_1 + A_2 + A_3 + (B_1 + B_2 + B_3)\,x + e_3$$

Gradient Boosting

Model 1:

$$y = A_1 + B_1x + e_1$$

Model 2:

$$e_1 = A_2 + B_2x + e_2$$

Model 3:

$$e_2 = A_3 + B_3x + e_3$$

"Strong learner" when combined

Combined Model:

$$y = A_1 + A_2 + A_3 + (B_1 + B_2 + B_3)\,x + e_3$$

Gradient Boosting

Model 1:

$$y = A_1 + B_1x + e_1$$

Model 2:

$$e_1 = A_2 + B_2x + e_2$$

Model 3:

$$e_2 = A_3 + B_3x + e_3$$

"Strong learner" when combined

Combined Model:

y = Sum of outputs of weak learners

Demo

Performing regression using gradient boosting

Demo

Hyperparameter tuning of the number of estimators in gradient boosting models

Summary

Using a series of sequential learners in an ensemble

Adaptive boosting or AdaBoost

Gradient Boosting

Hyperparameters tuning of gradient boosting models with early stopping