

A hand is shown pointing towards a complex, futuristic digital interface. The interface features a large circular gauge with multiple concentric rings, some of which are highlighted in green. Inside the gauge, there are several interlocking gears. The background is dark blue with various data visualizations, including bar charts and line graphs, some of which are also highlighted in green. The overall aesthetic is high-tech and modern.

Gradient Boosting

Cláudia Antunes

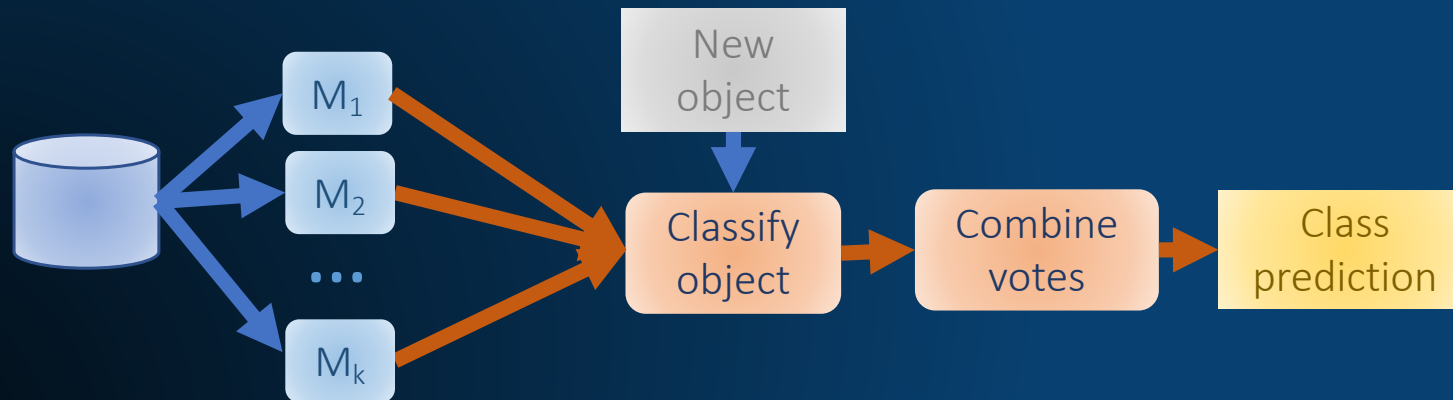
Instituto Superior Técnico – Universidade de Lisboa

ENSEMBLE CLASSIFIER

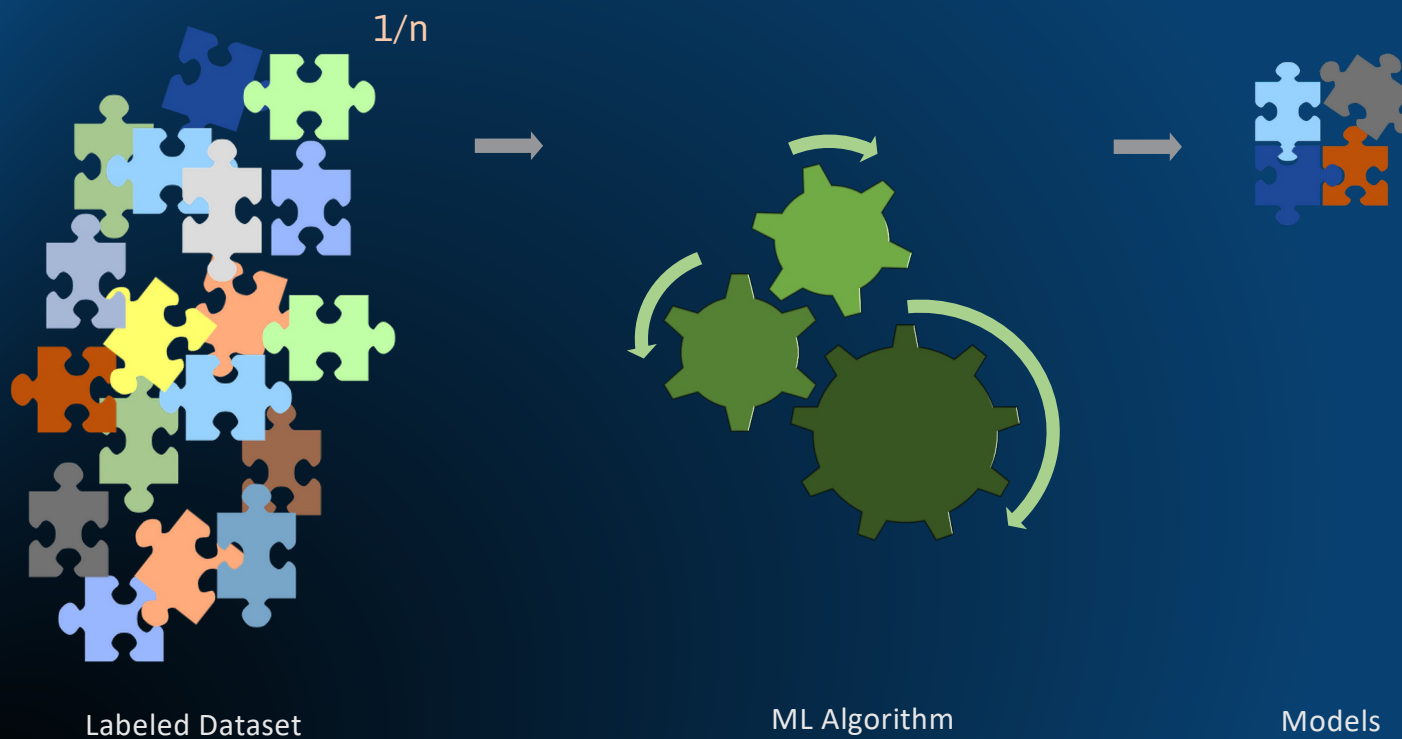


Training

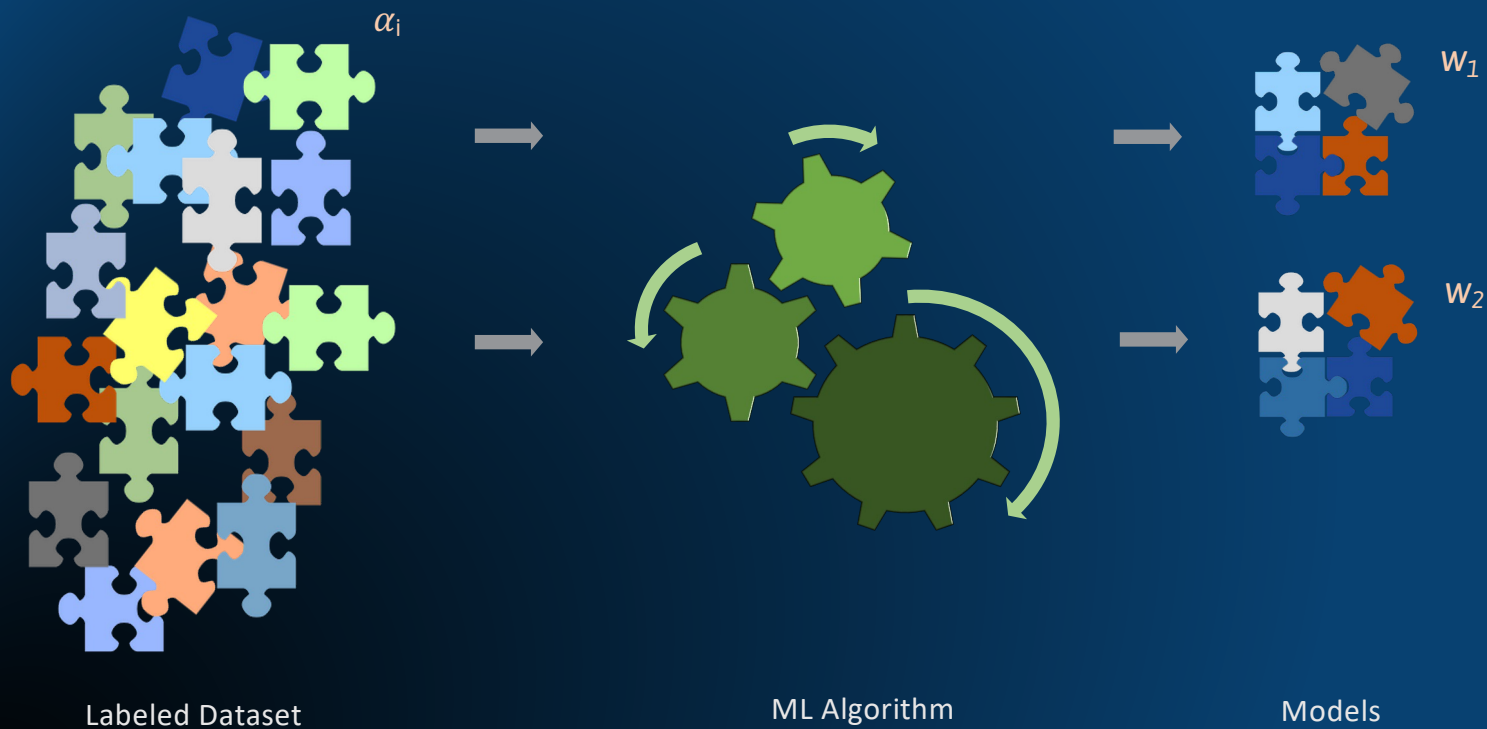
Model usage



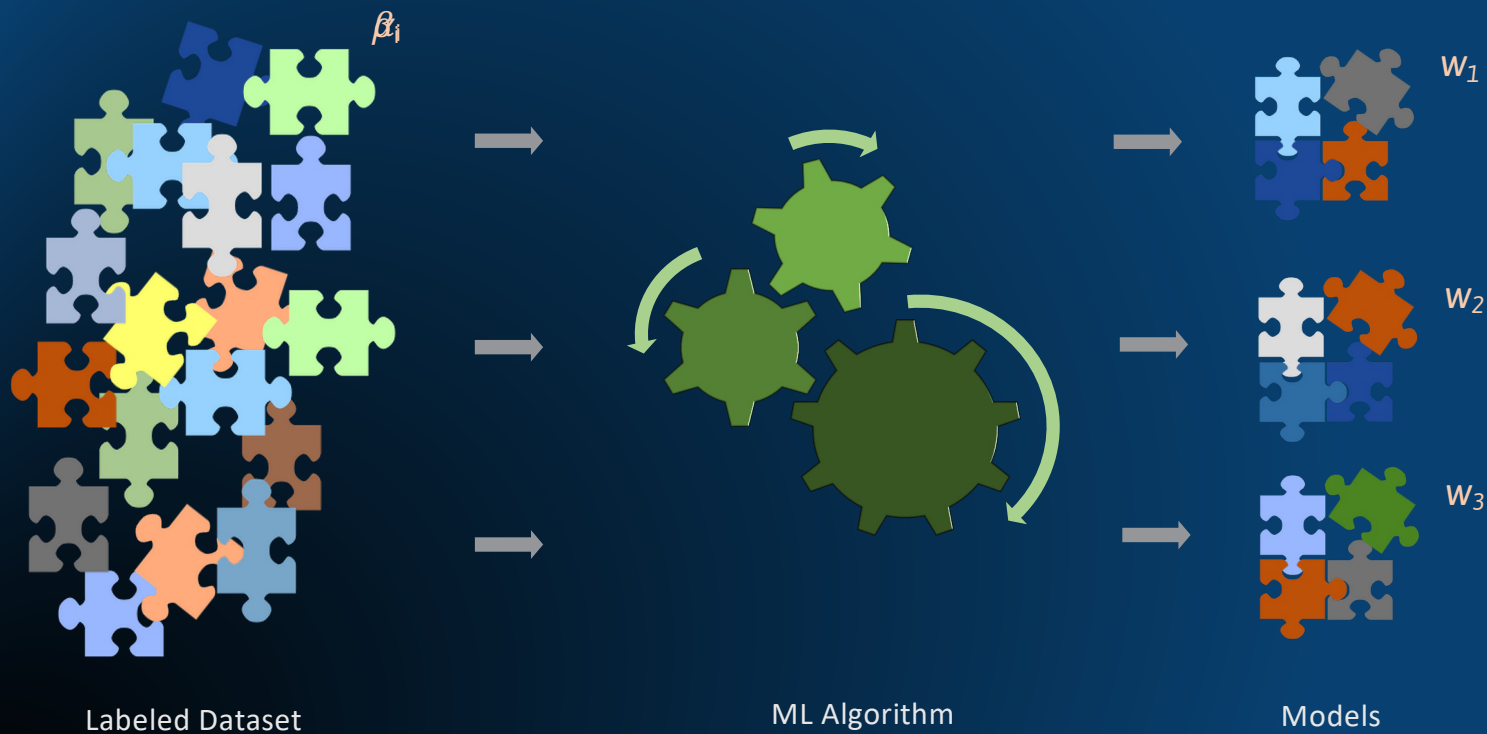
BOOSTING



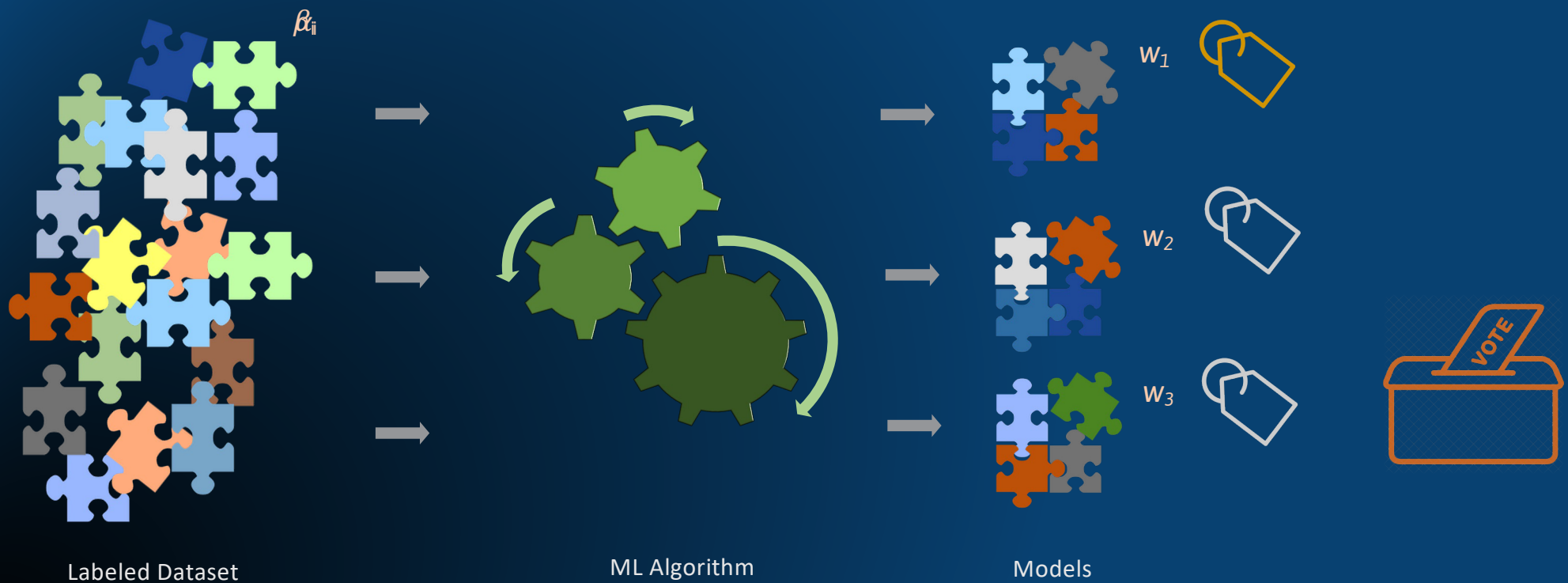
BOOSTING



BOOSTING



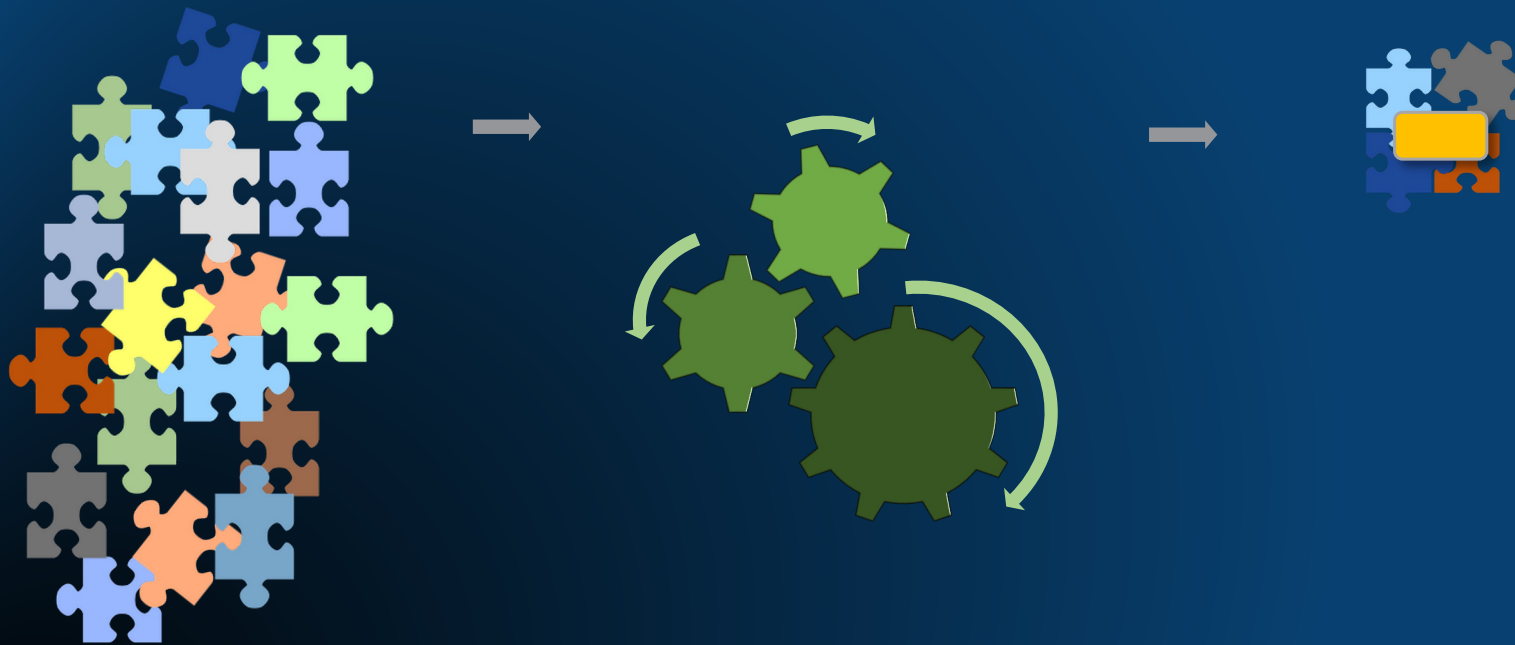
BOOSTING



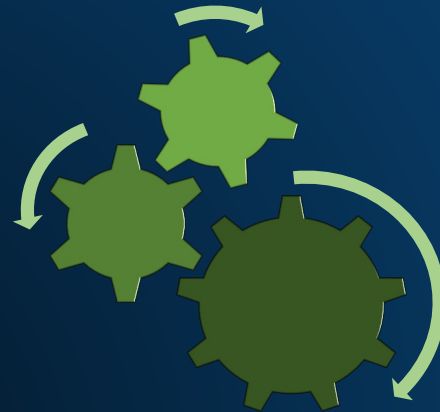


Gradient Boosting

GRADIENT BOOSTING



GRADIENT BOOSTING



residuals

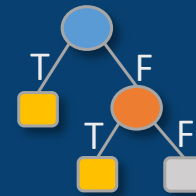
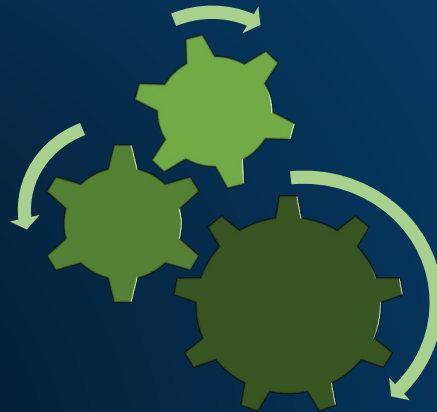


Data Science by Cláudia Antunes

GRADIENT BOOSTING



residuals



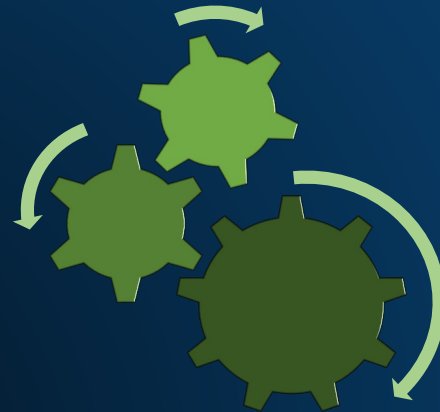
+

w_1

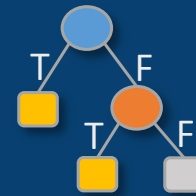
GRADIENT BOOSTING



residuals

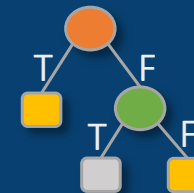


+



w_1

+



w_2



DIVERSITY THROUGH GRADIENT SEARCH



$$L(y, f(x))$$

DIVERSITY THROUGH GRADIENT SEARCH



$$L(y, f(x))$$

$$M_i = \operatorname{argmin}_f \sum_{j=1}^n L(y_j, f(x_j))$$

$$\frac{\partial \sum_{j=1}^n L(y_j, M_i(x_j))}{\partial M_i(x_j)} = \frac{\partial L(y_j, M_j(j))}{\partial M_i(x_j)}$$

DIVERSITY THROUGH GRADIENT SEARCH



$$L(y, f(x))$$

$$L_{MSE}(y, f(x)) = \frac{1}{2}(y - f(x))^2$$

$$M_i = \operatorname{argmin}_f \sum_{j=1}^n L(y_j, f(x_j))$$

$$\frac{\partial \sum_{j=1}^n L(y_j, M_i(x_j))}{\partial M_i(x_j)} = \frac{\partial L(y_j, M_i(x_j))}{\partial M_i(x_j)}$$

$$\frac{\partial L_{MSE}(y_j, M_i(x_j))}{\partial M_i(x_j)} = -(y_j - M_i(x_j))$$

$$y_j - M_i(x_j) = \begin{cases} 1, & y_j \neq M_i(x_j) \\ 0, & \text{otherwise} \end{cases}$$

residuals

DIVERSITY THROUGH GRADIENT SEARCH



$$L(y, f(x))$$

$$L_{MSE}(y, f(x)) = \frac{1}{2}(y - f(x))^2$$

$$M_i = \operatorname{argmin}_f \sum_{j=1}^n L(y_j, f(x_j))$$

$$\frac{\partial \sum_{j=1}^n L(y_j, M_i(x_j))}{\partial M_i(x_j)} = \frac{\partial L(y_j, M_j(j))}{\partial M_i(x_j)}$$

$$r_{ij} = - \frac{\partial L(y_j, M_i(x_j))}{\partial M_i(x_j)}$$

$$\frac{\partial L_{MSE}(y_j, M_i(x_j))}{\partial M_i(x_j)} = -(y_j - M_i(x_j))$$

$$y_j - M_i(x_j) = \begin{cases} 1, & y_j \neq M_i(x_j) \\ 0, & \text{otherwise} \end{cases}$$

residuals

MODELS' COMPOSITION



$$M_1 = \operatorname{argmin}_f \sum_{j=1}^n L(y_j, f(x_j))$$

$$M_{i+1} = M_i + \eta \operatorname{argmin}_f \sum_{j=1}^n L(y_j, M_i(x_j) + f(x_j))$$

MODELS' COMPOSITION



$$M_1 = \operatorname{argmin}_f \sum_{j=1}^n L(y_j, f(x_j))$$

$$M_{i+1} = M_i + \eta \operatorname{argmin}_f \sum_{j=1}^n L(y_j, M_i(x_j) + f(x_j))$$

learning rate



Input

Output M_k

$$D = \{(x_j, y_j) : 1 \leq j \leq n\}, \quad k, c, \eta, L(y, M(x))$$

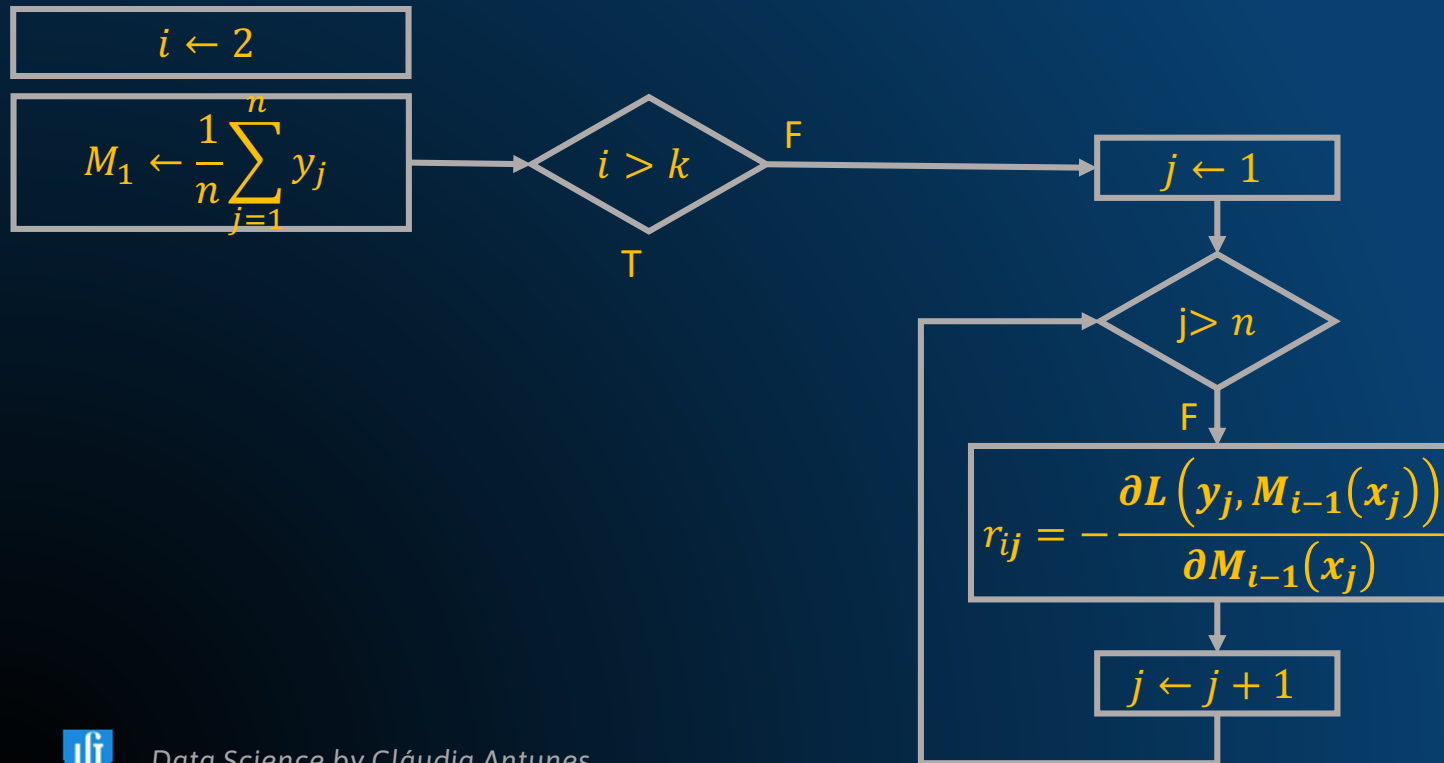
$$i \leftarrow 2$$

$$M_1 \leftarrow \frac{1}{n} \sum_{j=1}^n y_j$$

Input

$$D = \{(x_j, y_j) : 1 \leq j \leq n\}, k, c, \eta, L(y, M(x))$$

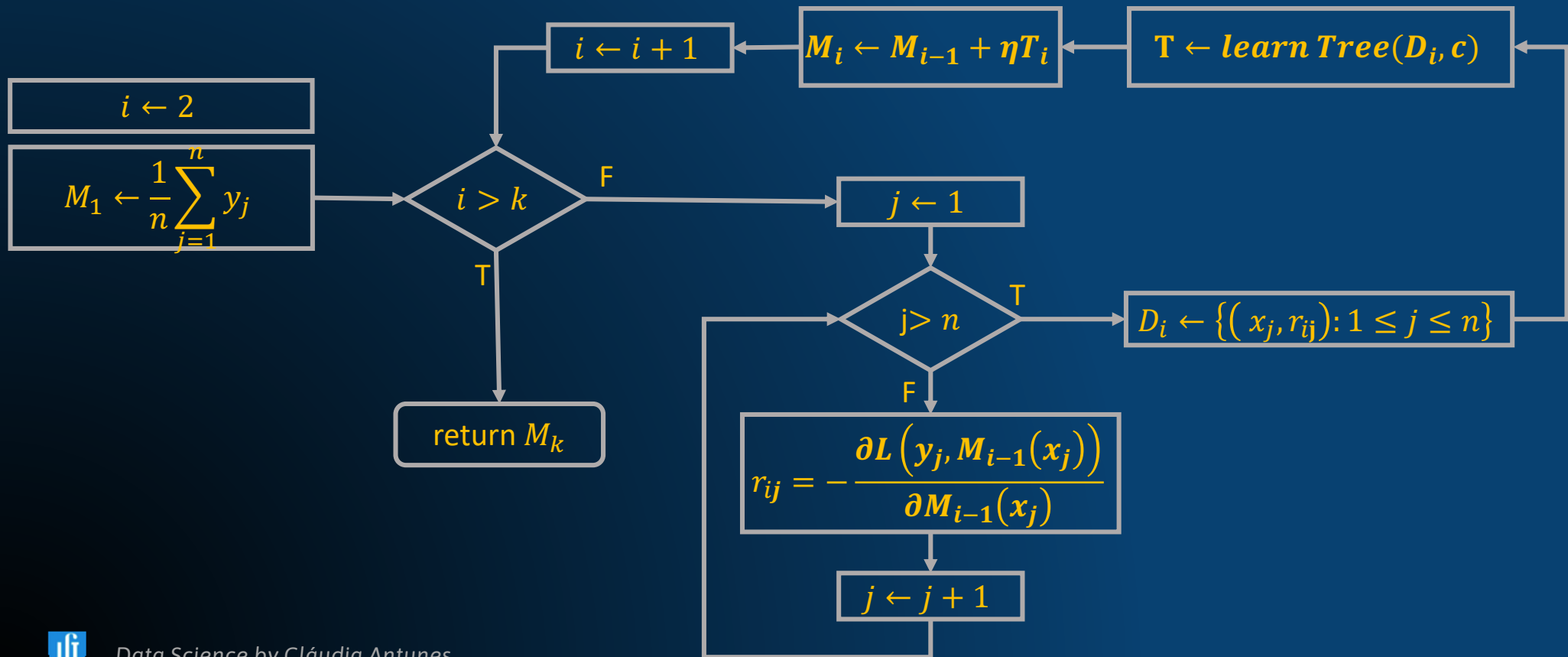
Output M_k



Input

$$D = \{(x_j, y_j) : 1 \leq j \leq n\}, k, c, \eta, L(y, M(x))$$

Output M_k



GRADIENT BOOSTING



Records as **conjunctions of propositions**

Training algorithm

- Learn **k decision trees** from **re-labeling the dataset**, according to the gradient of the loss function
- Create a model from the **composition of the k learnt trees**

Classification algorithm

For each **Z** to be classified













- Classify **Z** according to **the leaf reached by the final model**

Prediction Example

LEARNING ALGORITHM



Dataset

			1	-3.75
			7	2.25
			2	-2.75
			9	4.25

4.75

$$MAE = 3.25$$

LEARNING ALGORITHM



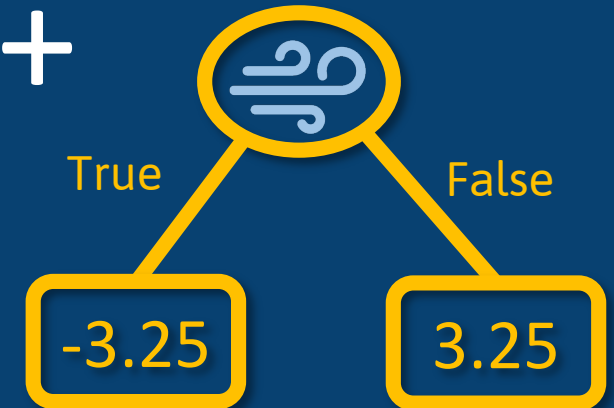
Dataset

			1	-3.75	1.5
			2	-2.75	1.5
			7	2.25	8.0
			9	4.25	8.0

Observed Residuals Prediction

4.75













+



LEARNING ALGORITHM



Dataset

			1	1.5	-0.5
			2	1.5	0.5
			7	8.0	-1
			9	8.0	1

Prediction Residuals



$$MAE = 0.75$$

