

Gradient boosting tree

1





Prérequis

Pour comprendre la théorie du gradient boosting tree il faut bien comprendre :

- L'ensemble learning
- L'algorithme des arbres de décision
- L'algorithme du random forest
- L'algorithme de AdaBoost

Adaboost



Gradient boosting



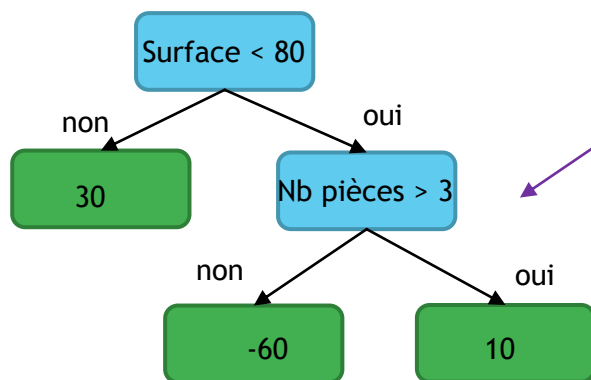
Exemple de construction d'arbres

$$\text{erreur} = y - \hat{y}$$

$\overline{\text{prix}} = 175$

Année de construction	Surface	Nb pièces	Prix (k€)	Résidus
1980	60	3	150	-25
2005	70	4	180	5
2010	90	4	220	45
2000	29	1	70	-105
⋮	⋮	⋮	⋮	⋮

$$\begin{array}{c} \bar{y} \\ 175 \\ 175 \\ 175 \\ 175 \\ \vdots \end{array} + \left[\text{LR} * \begin{array}{c} \text{arbre}_1 \\ -60 \\ 10 \\ 30 \\ -60 \\ \vdots \end{array} \right] = \begin{array}{c} \hat{y} \\ 115 \\ 185 \\ 205 \\ 115 \\ \vdots \end{array}$$



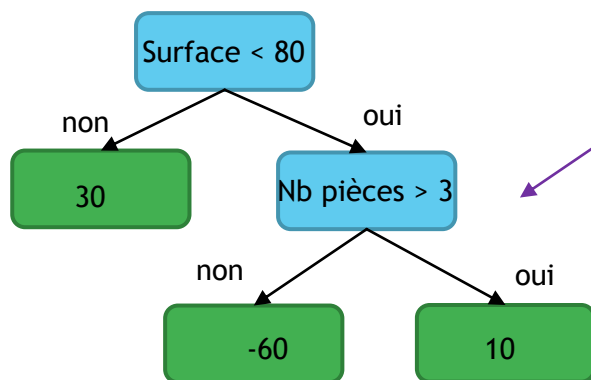
Exemple de construction d'arbres

$$\text{erreur} = y - \hat{y}$$

$\overline{\text{prix}} = 175$

Année de construction	Surface	Nb pièces	Prix (k€)	Résidus
1980	60	3	150	-25
2005	70	4	180	5
2010	90	4	220	45
2000	29	1	70	-105
⋮	⋮	⋮	⋮	⋮

$$\begin{array}{c} \bar{y} \\ 175 \\ 175 \\ 175 \\ 175 \\ \vdots \end{array} + 0.5 * \begin{array}{c} \text{arbre}_1 \\ -60 \\ 10 \\ 30 \\ -60 \\ \vdots \end{array} = \begin{array}{c} \hat{y} \\ 135 \\ 180 \\ 190 \\ 135 \\ \vdots \end{array}$$





Forêt du gradient descent

$$Y_3 = Y_1 + Lr * R_2 + Lr * R_3$$

$$X \rightarrow Y - Y_2 = R_3$$

$$X \rightarrow Y - Y_1 = R_2$$

$$Y_2 = Y_1 + Lr * R_2$$

$$X \rightarrow Y - Y_4 = R_5$$

$$X \rightarrow Y - Y_3 = R_4$$

$$Y_5 = Y_1 + Lr * R_2 + Lr * R_3 + Lr * R_4 + Lr * R_5$$

$$Y_4 = Y_1 + Lr * R_2 + Lr * R_3 + Lr * R_4$$

L'algorithme du gradient boosting tree

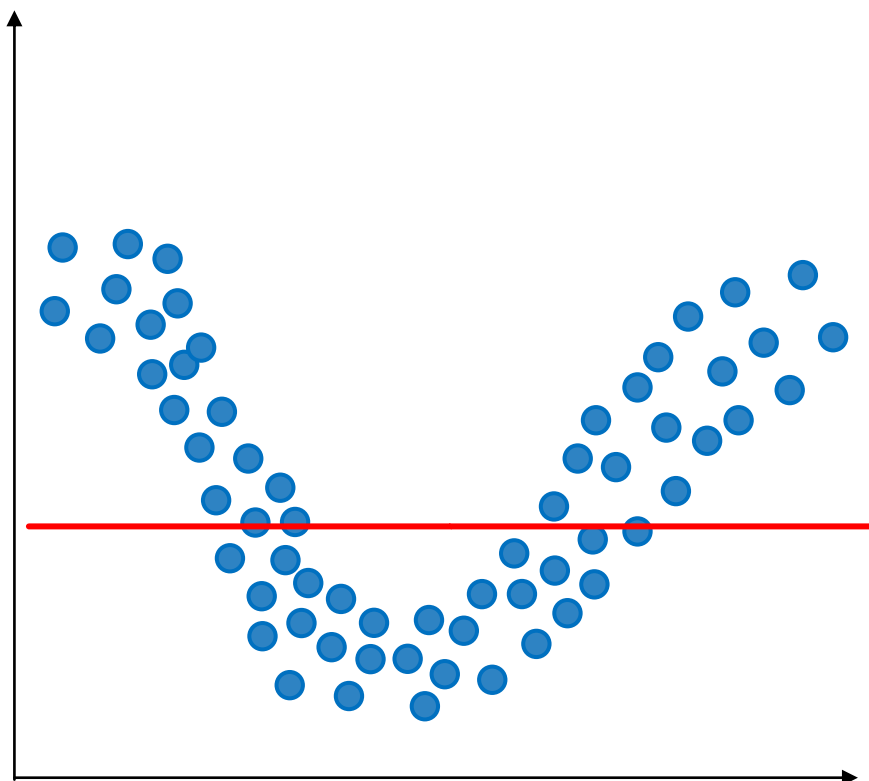
$$L(y, \hat{y}) = \frac{1}{2}(y - f(x))^2$$

Step 1 : Initialiser la première feuille par \bar{y} : $F(x) = f_0(x) = \bar{y}$

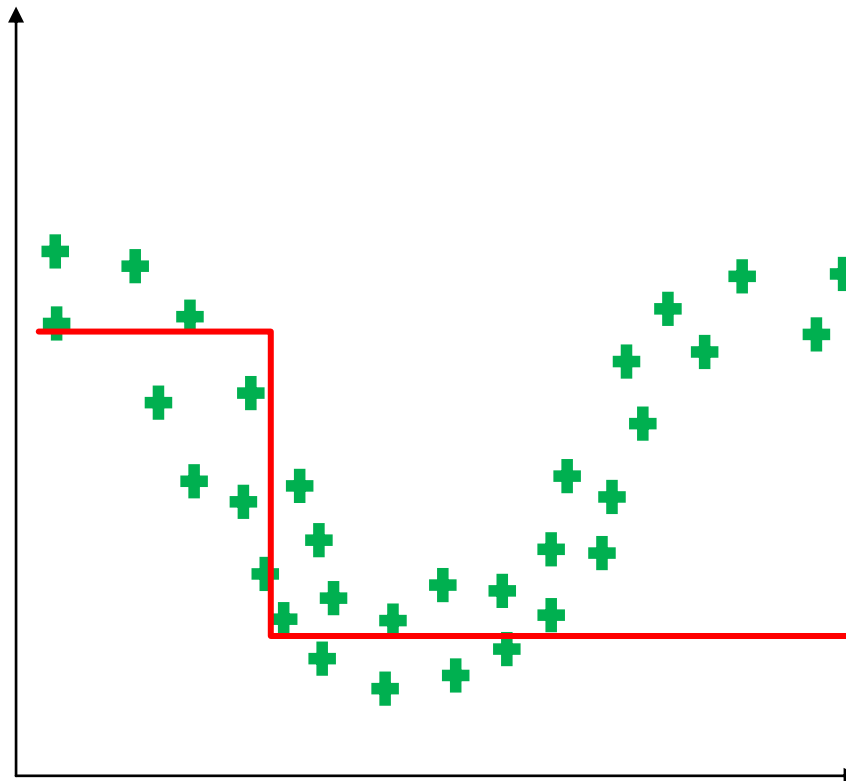
Step 2 : for t = 1 to nb_tree :

- calculer $r_{t-1} = - \left[\frac{\partial L(y - F(x))}{\partial F(x)} \right] = F(x) - y = \hat{y} - y$
- Entraîner un arbre avec les données (x, r_{t-1}) : $f_t(x) = \hat{r}_{t-1}$
- Mettre à jour $F(x) = f_0(x) + \nu \sum_{j=1}^m f_j(x)$

Gradient boosting

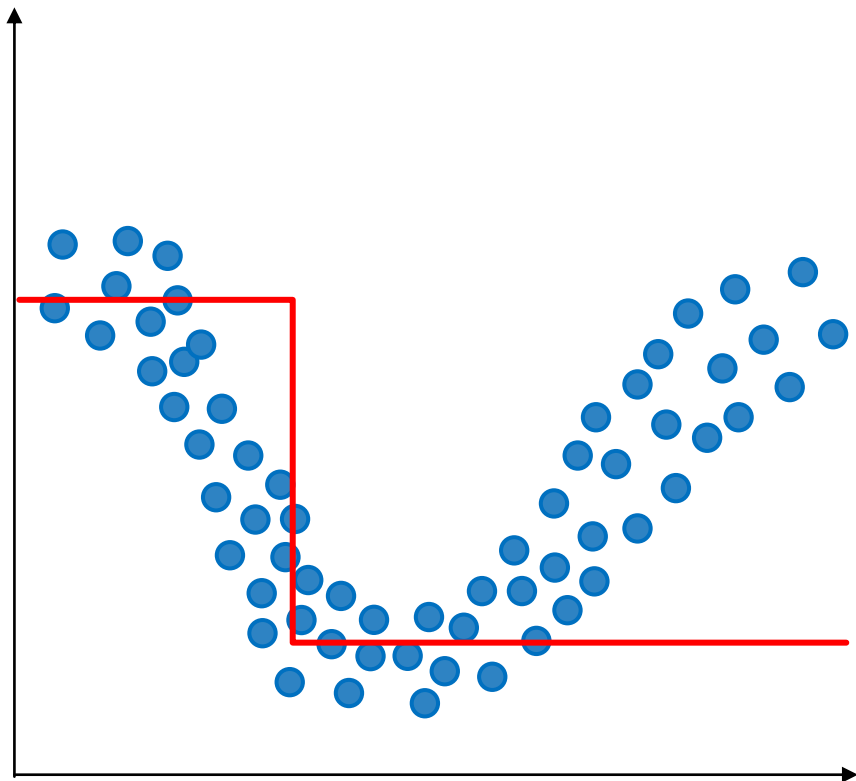


● Training set

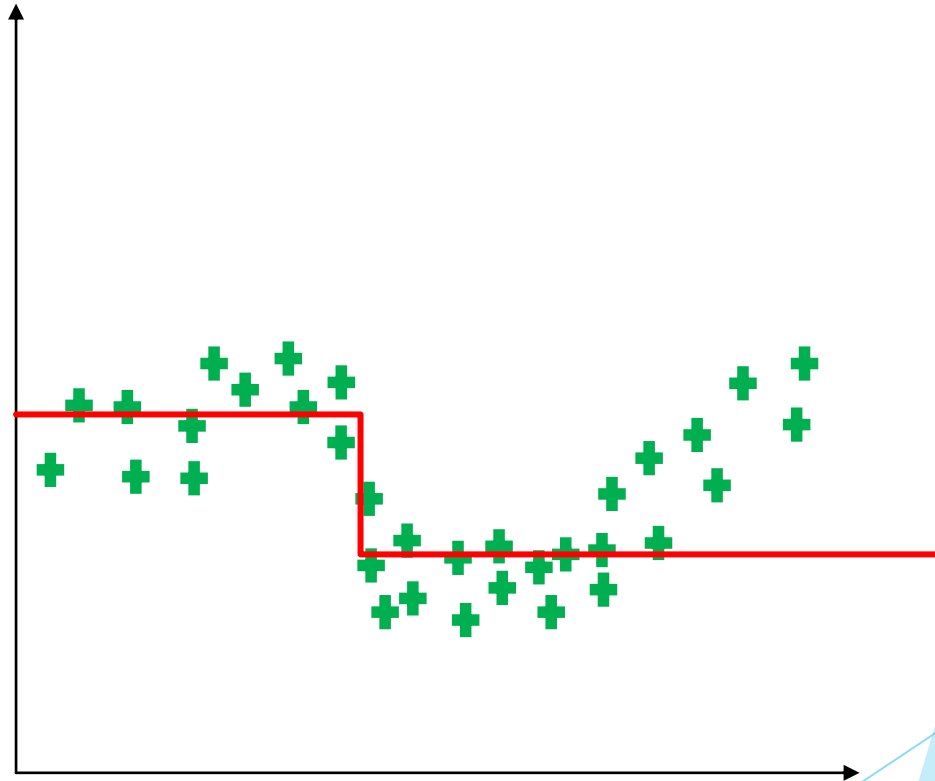


+ Residual

Gradient boosting

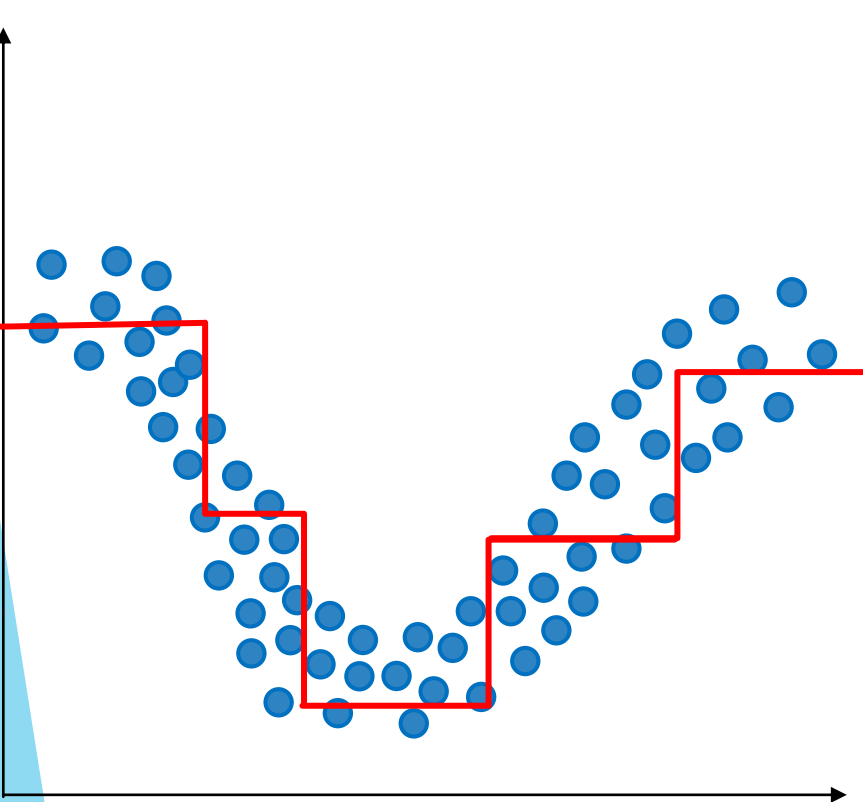


● Training set

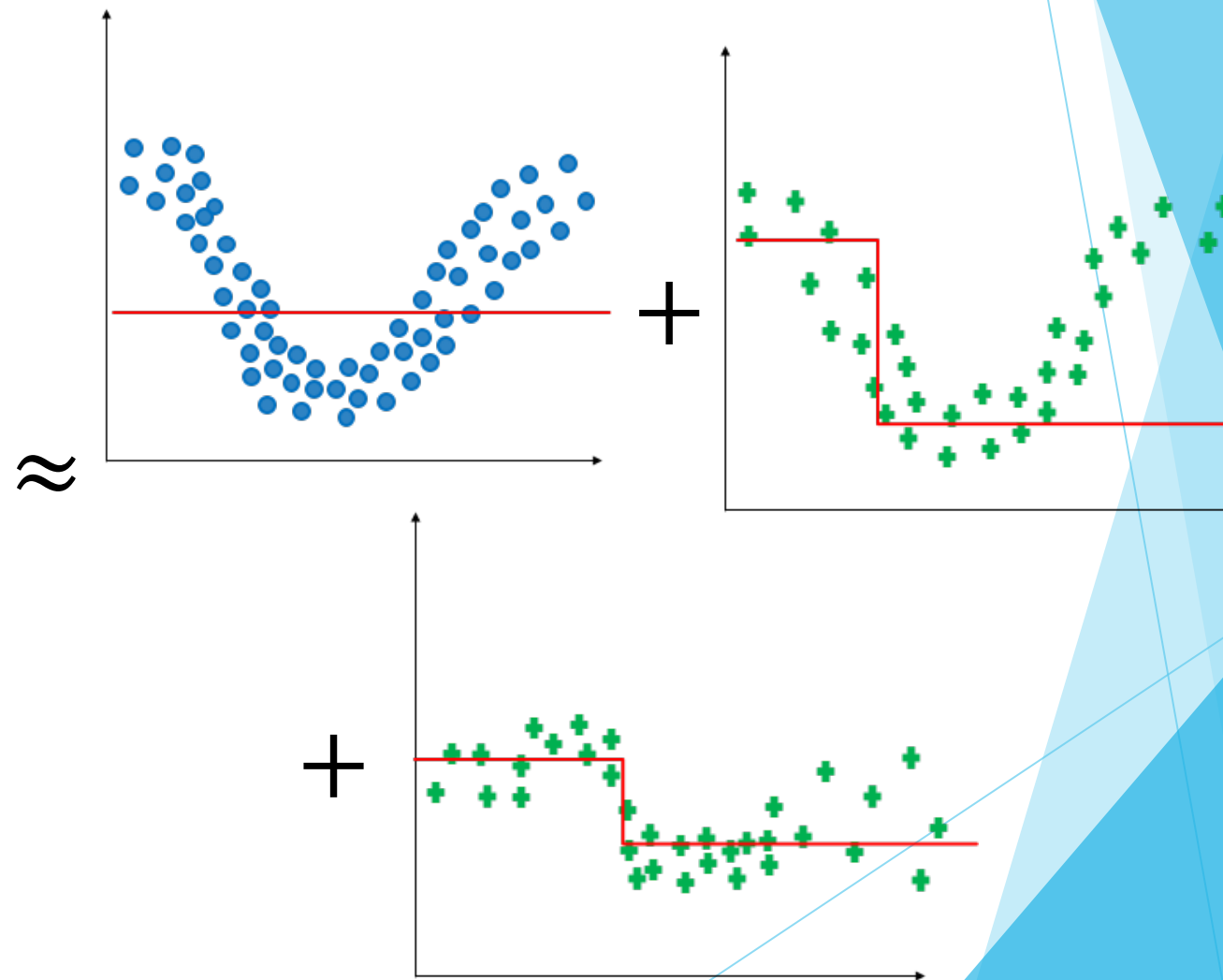


+ Residual

Gradient boosting



● Training set



Implémentation

- ▶ Sklearn gradient boosting tree
- ▶ XGBoost ++
- ▶ CATBoost (données catégorielle)
- ▶ Lightgbm (plus rapide)