Decision Trees
for Regression

How to grow a decision tree

- The tree is built **greedily** from top to bottom
- Each split is selected to **maximize information gain (IG)**

$$IG = \underbrace{\mathbf{Impurity}(Z)}_{\text{Error before split}} - \underbrace{\left(\frac{|Z_L|}{|Z|}\mathbf{Impurity}(Z_L) + \frac{|Z_R|}{|Z|}\mathbf{Impurity}(Z_R)\right)}_{\text{Error after split}}$$

Decision Tree for Regression

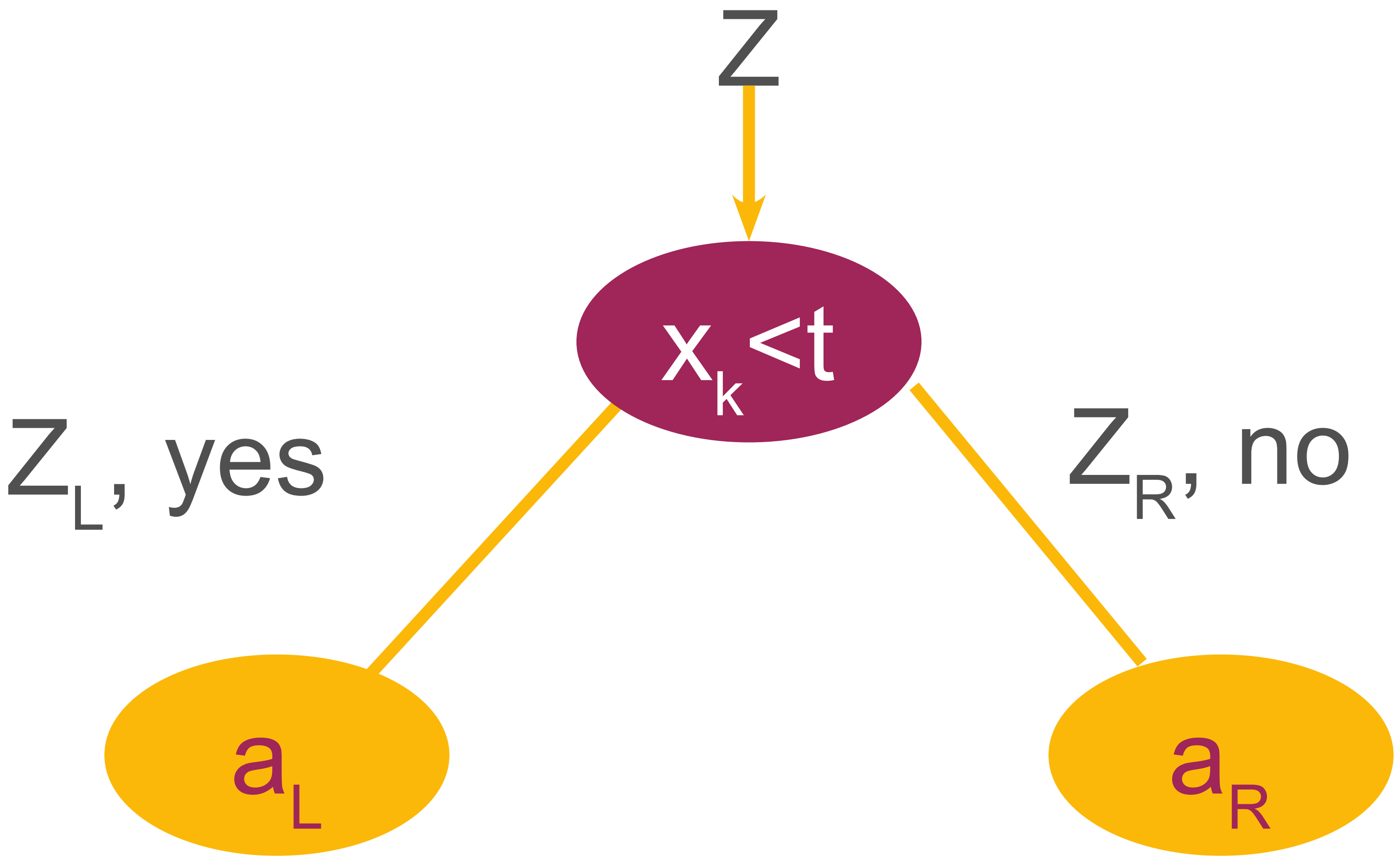
Given a training set: $\boldsymbol{Z}=\{(\boldsymbol{x}_1,y_1),\ldots,(\boldsymbol{x}_n,y_n)\}$
y_i- real values

Goal is to find $f(\boldsymbol{x})$ (a tree) such that

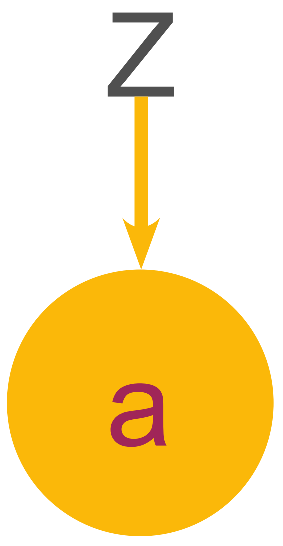
$$\min \sum_{i=1}^{n} (f(\boldsymbol{x}_i) - y_i)^2$$

How to grow a decision tree for regression?

How to find the best split



Without split: predict one number ***a***



$$\hat{a} = \min_{a} \sum_{i \in Z} (a - y_i)^2$$

$$\hat{a} = \frac{1}{|Z|} \sum_{i \in Z} y_i$$

|Z| - number of elements in Z

$$\text{Impurity(Z)} = \frac{1}{|Z|} \sum_{i \in Z} (\hat{a} - y_i)^2$$

Find the best split (x_k<t):

$$\min_{k,\, t,\, a_L, a_R} \sum_{i \in Z_L} (a_L - y_i)^2 + \sum_{i \in Z_R} (a_R - y_i)^2$$

Values in leaves

$$\widehat{a_L} = \frac{1}{|Z_L|} \sum_{i \in Z_L} y_i\,, \qquad \widehat{a_R} = \frac{1}{|Z_R|} \sum_{i \in Z_R} y_i$$

$|Z_L|$ - number of elements in Z_L,
$|Z_R|$ - number of elements in Z_R

$$\min_{k,\, t} \sum_{i \in Z_L} (\widehat{a_L} - y_i)^2 + \sum_{i \in Z_R} (\widehat{a_R} - y_i)^2$$

Find the best split:

$$\text{Impurity}(Z_L) = \frac{1}{|Z_L|} \sum_{i \in Z_L} (\widehat{a_L} - y_i)^2$$

$$\text{Impurity}(Z_R) = \frac{1}{|Z_R|} \sum_{i \in Z_R} (\widehat{a_R} - y_i)^2$$

Maximize the information gain (IG):

$$\boldsymbol{IG} = \textbf{Impurity}(\boldsymbol{Z}) - \left(\frac{|\boldsymbol{Z}_L|}{|\boldsymbol{Z}|} \textbf{Impurity}(\boldsymbol{Z}_L) + \frac{|\boldsymbol{Z}_R|}{|\boldsymbol{Z}|} \textbf{Impurity}(\boldsymbol{Z}_R) \right)$$

with respect to k, t (splitting criteria x_k<t)

Stopping rule

- The node depth is equal to the **maxDepth** training parameter.

- No split candidate leads to an information gain greater than **minInfoGain**.

- No split candidate produces child nodes which have at least **minInstancesPerNode** training instances ($|Z_L|, |Z_R| <$ **minInstancesPerNode**) each