
H2O.ai

Decision Tree Algorithms

TreeGrowing $(S, A, y, SplitCriterion, StoppingCriterion)$

Where:

S - Training Set

A - Input Feature Set

y - Target Feature

$SplitCriterion$ - the method for evaluating a certain split

$StoppingCriterion$ - the criteria to stop the growing process

Create a new tree T with a single root node.

IF $StoppingCriterion(S)$ THEN

    Mark T as a leaf with the most
    common value of y in S as a label.

ELSE

    $\forall a_i \in A$ find a that obtain the best $SplitCriterion(a_i, S)$.

    Label t with a

    FOR each outcome v_i of a:

        Set $Subtree_i$= TreeGrowing $(\sigma_{a=v_i} S, A, y)$.

        Connect the root node of t_T to $Subtree_i$ with
                an edge that is labelled as v_i

    END FOR

END IF

RETURN TreePruning (S, T, y)

TreePruning (S, T, y)

Where:

S - Training Set

y - Target Feature

T - The tree to be pruned

DO

    Select a node t in T such that pruning it
        maximally improve some evaluation criteria

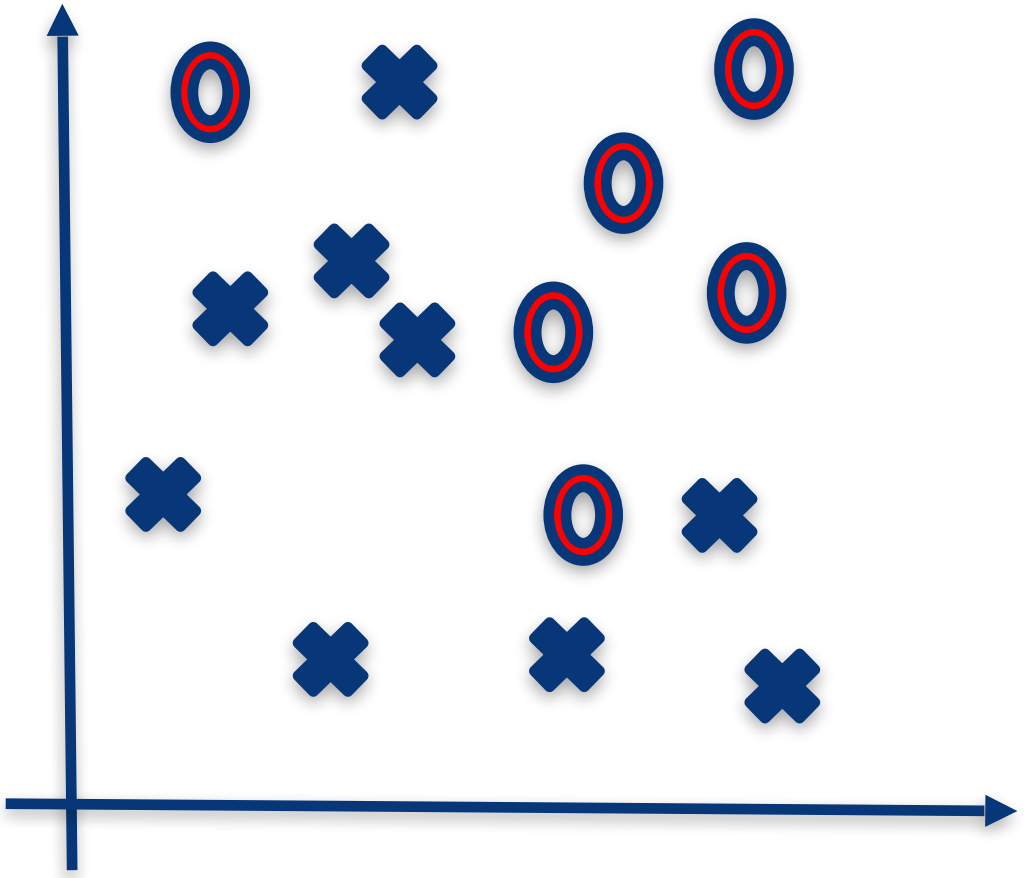
    IF $t \neq \emptyset$ THEN $T = pruned(T, t)$

UNTIL $t = \emptyset$

RETURN T

- Breaks sample data into homogenous pieces
  - Sample means
- Handles categorical and continuous data
- Identifies interaction effects and important variables
- Has multiple algorithms
  - ID3
  - C4.5
  - CHAID

Growing a Tree: Splitting



Decision Tree Algorithms

- Breaks sample data into homogenous pieces
  - Sample means
- Handles categorical and continuous data
- Identifies interaction effects and important variables
- Has multiple algorithms
  - ID3
  - C4.5
  - CHAID

TreeGrowing $(S, A, y, SplitCriterion, StoppingCriterion)$
Where:
S - Training Set
A - Input Feature Set
y - Target Feature
$SplitCriterion$ - the method for evaluating a certain split
$StoppingCriterion$ - the criteria to stop the growing process
Create a new tree T with a single root node.
IF $StoppingCriterion(S)$ THEN
    Mark T as a leaf with the most
    common value of y in S as a label.
ELSE
    $\forall a_i \in A$ find a that obtain the best $SplitCriterion(a_i, S)$.
    Label t with a
    FOR each outcome v_i of a:
        Set $Subtree_i$= TreeGrowing $(\sigma_{a=v_i} S, A, y)$.
        Connect the root node of t_T to $Subtree_i$ with
            an edge that is labelled as v_i
    END FOR
END IF
RETURN TreePruning (S, T, y)

TreePruning (S, T, y)
Where:
S - Training Set
y - Target Feature
T - The tree to be pruned
DO
    Select a node t in T such that pruning it
        maximally improve some evaluation criteria
    IF $t \neq \emptyset$ THEN $T = pruned(T, t)$
UNTIL $t = \emptyset$
RETURN T