COL 774
Machine Learning
Nov 25, 2020

Last class:- Finished SVMs
↳ Primal formulation — Hard Margin → Soft Margin
↳ Lagrangian
Dual formulation
→ Kernels (Non-linear SVMs)

SVR:- Support Vector Regression
Supervised Model

Decision Trees:-

$\{x^{(i)}, y^{(i)}\}_{i=1}^{m}$

$y^{(i)} \in \{0,1\}$
OR
$y^{(i)} \in \{1 \ldots r\}$

Till now:-
$x^{(i)} \in \mathbb{R}^n$
↳ $\forall j$: $x_j^{(i)}$ is continuous (Real valued features).

1 of L one-hot encoding

$x_j^{(i)} \in \{v_1 \ldots v_k\}$

Explicitly Deal with discrete valued attributes (or a mix of discrete & continuous attributes)

$x_j^{(i)} \in \{v_1^j, v_2^j, v_3^j \ldots v_k^j\}$ (Discrete Domain)

Earlier, $x_j^{(i)} \in \mathbb{R}$

if domains are identical for $x_j^{(i)}$'s

$x_j^{(i)} \in \{v_1 \ldots v_k\}$

Decision Trees:- The hypothesis space is organized in the form of a decision tree

Play Tennis:- Yes/No

X_1: Outlook $\in$ {Sunny, Overcast, Rain}

could also be modeled as real valued.
X_2: Humidity $\in$ {High, Normal}
X_3:- Wind $\in$ {Strong, Weak}
X_4: Temperature $\in$ {Hot, Moderate, Cold}

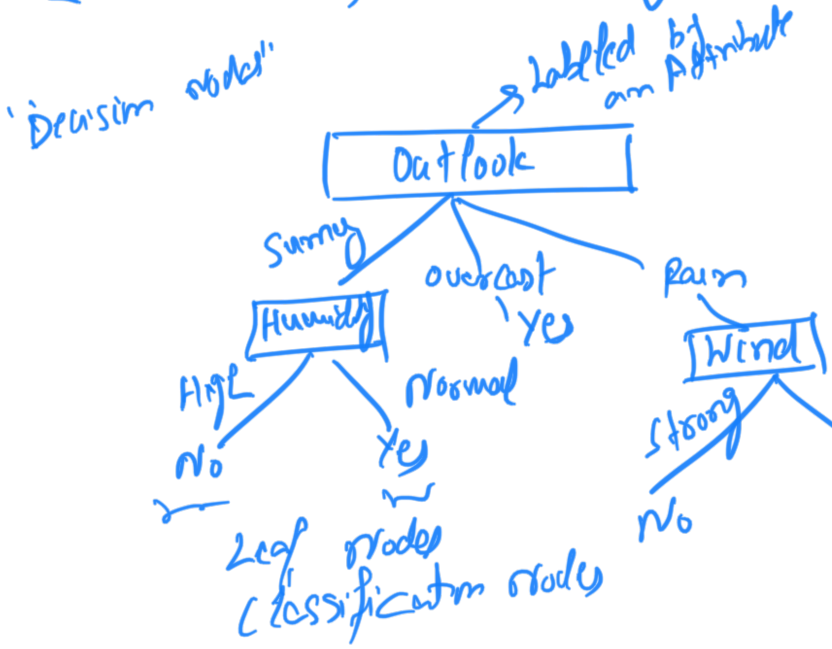
$[\quad x_4$: Temperature: $\{$Hot, ...

$\{x^{(i)}, y^{(i)}\}_{i=1}^{m}$ $\quad \{ \langle$ Sunny, Normal, Weak, Moderate $\rangle$, $\langle$ Yes $\rangle \}$ ($x^{(1)}$, $y^{(1)}$)

⋮

$x^{(m)}$ $\{ \quad \}$

$\langle x_1, x_2, x_3, x_4 \rangle \rightarrow y$? Classification Problem.



Decision Tree for example considered above.

↳ Internal Node

↳ Leaf Node

⇒ organized in the form of a set of "Rules"

Space of all possible Decision Trees:

How many such possible decision trees are there?

(How many such possible distinct functions can they represent)

$\forall j: \; x_j^{(i)} \in \{0, 1\} \quad y^{(i)} \in \{0, 1\}$

How? $\quad 2^{2^n}$

2^n

$\forall j: \; x_j^{(i)} \in \{v_1, \ldots, v_k\} \quad y^{(i)} \in \{0, 1\}$

↳ How many functions?

$2^{(k^n)}$ $\quad 2^{k^n}$

$\forall j :- x_j^{(i)} \in \{v_1 - \quad v_k\}$ $\quad y^{(i)} \in \{1 - \cdots 0\}$

$\# :- \quad \gamma k^n$

Goal in decision tree learning :-

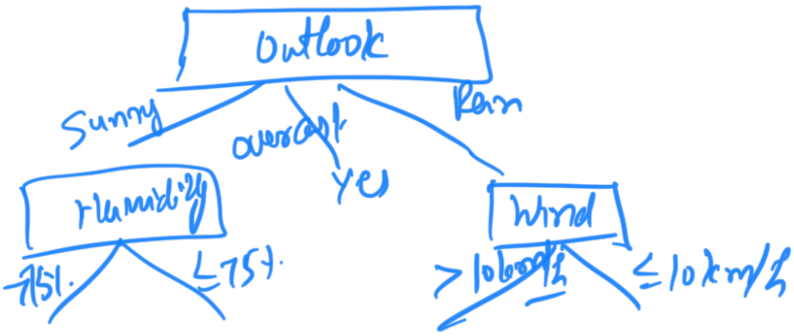
Construct a decision tree "of smallest size" which fits the training data "well"

if not :- overfitting (Inductive Bias)

How do we construct such a decision tree given $\{x^{(i)}, y^{(i)}\}_{i=1}^{m}$

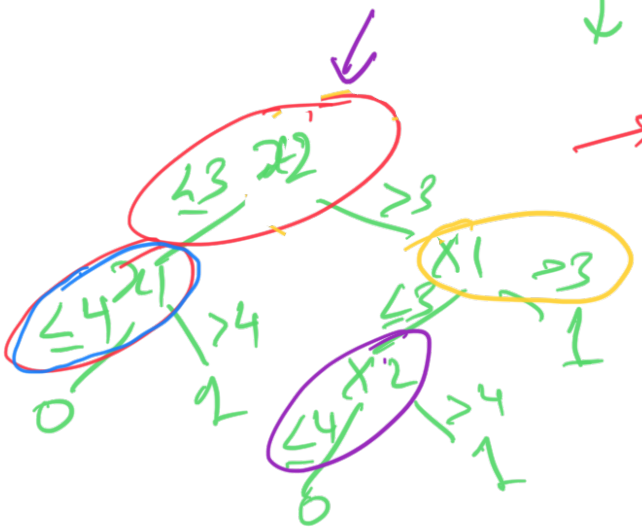
Two Remarks :-

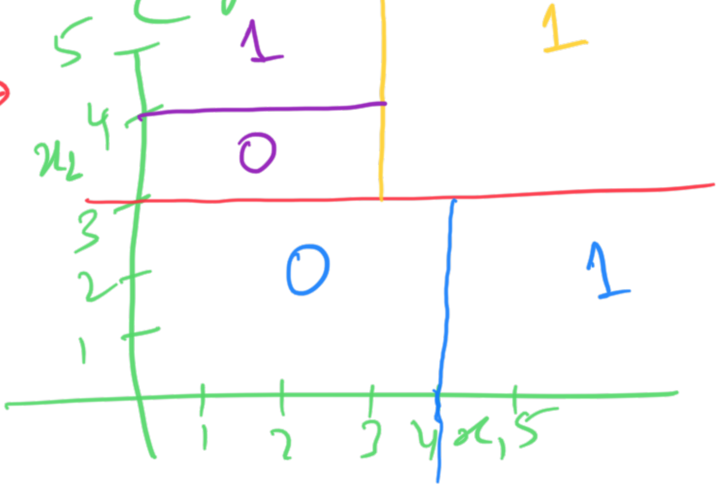
(A) Decision tree can handle cont attribute also



(B) In case of continuous valued attributes, the decision surface consists of axis parallel lines ("Rectangles")

$$\left\{ \begin{array}{l} x_1, x_2 \in R \\ y \in \{0, 1\} \end{array} \right.$$





Assumption of decision model :-

We split one attribute at a time.

$\hookrightarrow$

$I \{x^{(i)}, y^{(i)}\}_{i=1}^{m}$

Learning from data

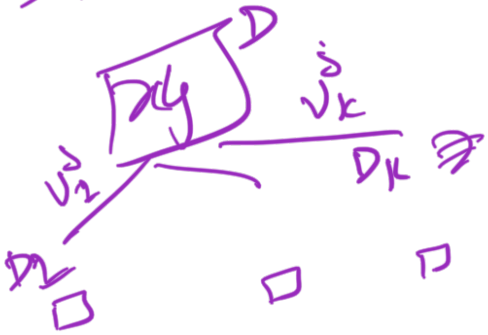
How do we grow decision tree?

$\hookrightarrow$ Data will come to a node (Root (initially))

$\hookrightarrow$ (1) classify the data at the current node.

i) `pyari` chooseAttrToSplit(D)

- Decide Attribute to split on
- Split the data on all possible attr value
- Recurse on each branch



(2) split the data based on an attribute value