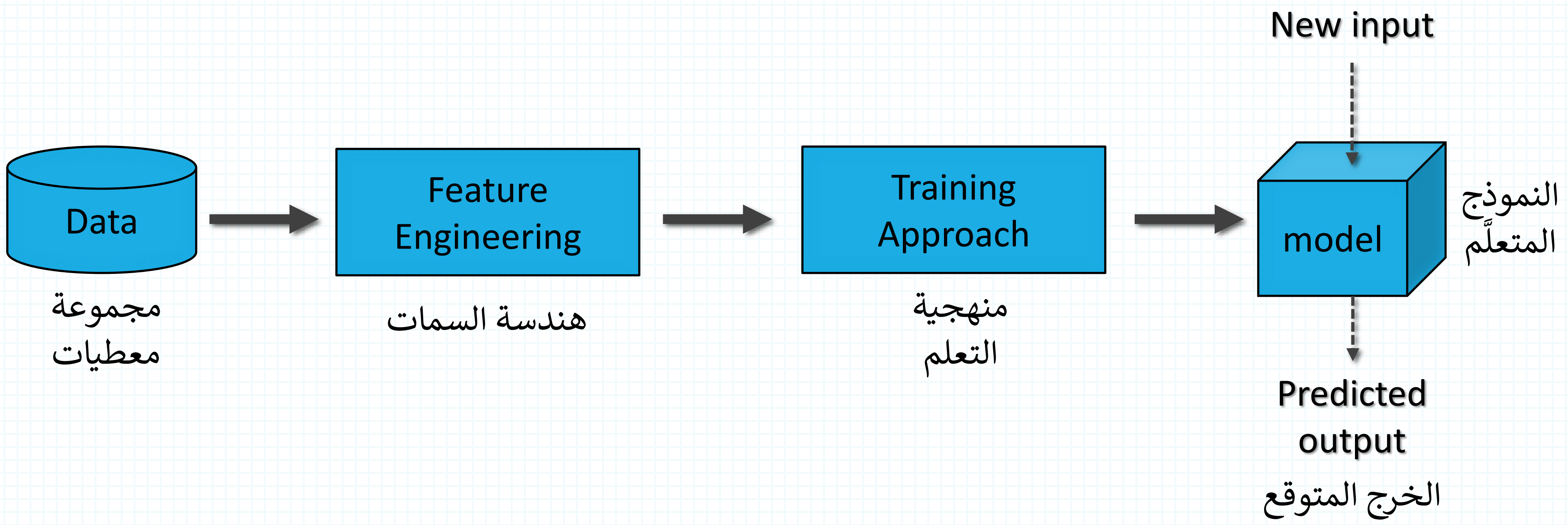
الجامعة السورية الخاصة
SYRIAN PRIVATE UNIVERSITY

المحاضرة الثالثة	كلية الهندسة المعلوماتية	مقرر تعلم الآلة

أشجار القرار 2
Decision Tree 2

د. رياض سنبل

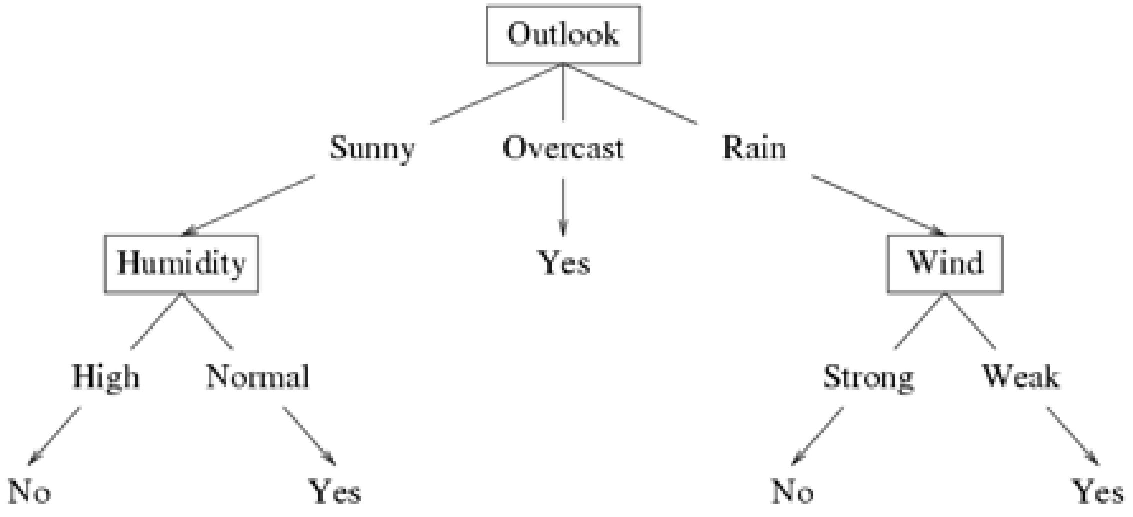
Traditional ML Pipeline



quick overview

Decision Tree Hypothesis Space

- Internal nodes test the value of particular features xi and branch according to the results of the test.
- Leaf nodes specify the class.

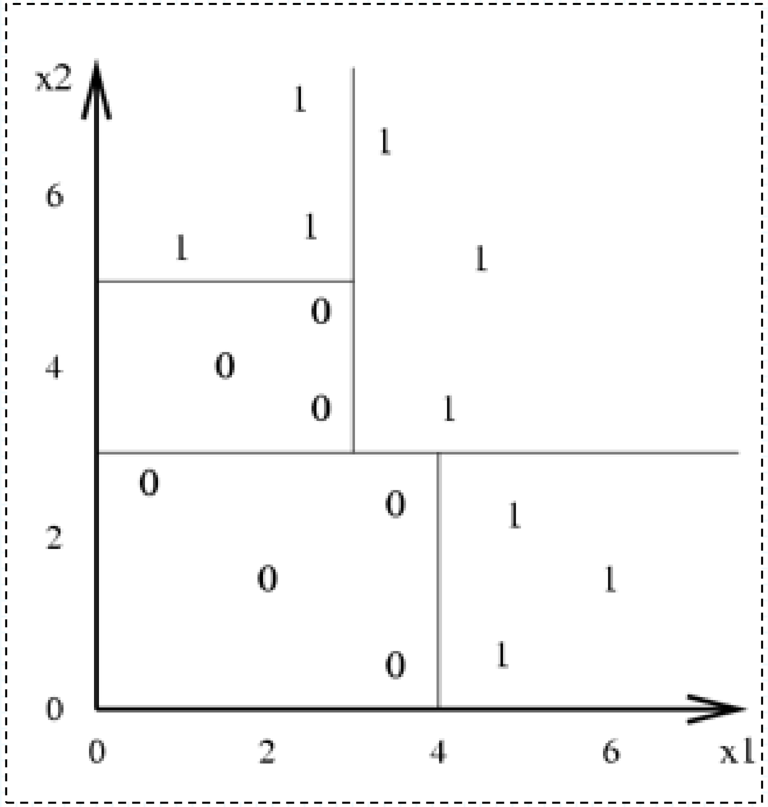


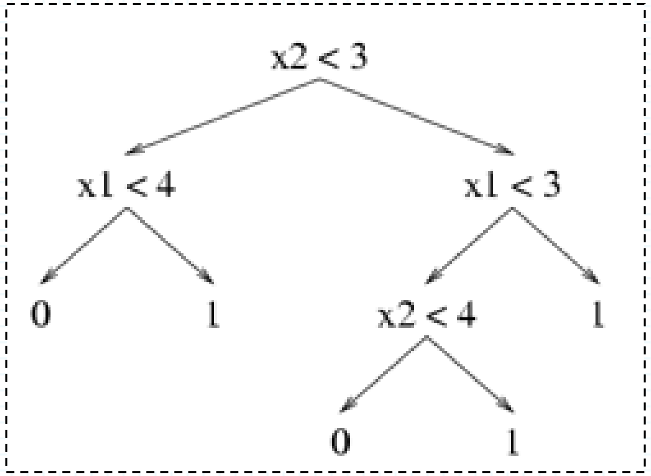
Suppose the features are **Outlook** (x_1), **Temperature** (x_2), **Humidity** (x_3), and **Wind** (x_4). Then the feature vector $\mathbf{x} = (Sunny, Hot, High, Strong)$ will be classified as **No**. The **Temperature** feature is irrelevant.

quick overview

Decision Tree Decision Boundaries

- Decision trees divide the feature space into axis-parallel rectangles, and label each rectangle with one of the K class.





- Advantages:
  - explicit relationship among features
  - human interpretable model
- Limitations?

quick overview

Learning Algorithm for Decision Trees

$$S = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N)\}$$

$$\mathbf{x} = (x_1, \ldots, x_d)$$

$$x_j, y \in \{0,1\}$$

$\textsc{GrowTree}(S)$

if $(y = 0 \text{ for all } \langle \mathbf{x}, y \rangle \in S)$ **return** new leaf(0)

else if $(y = 1 \text{ for all } \langle \mathbf{x}, y \rangle \in S)$ **return** new leaf(1)

else

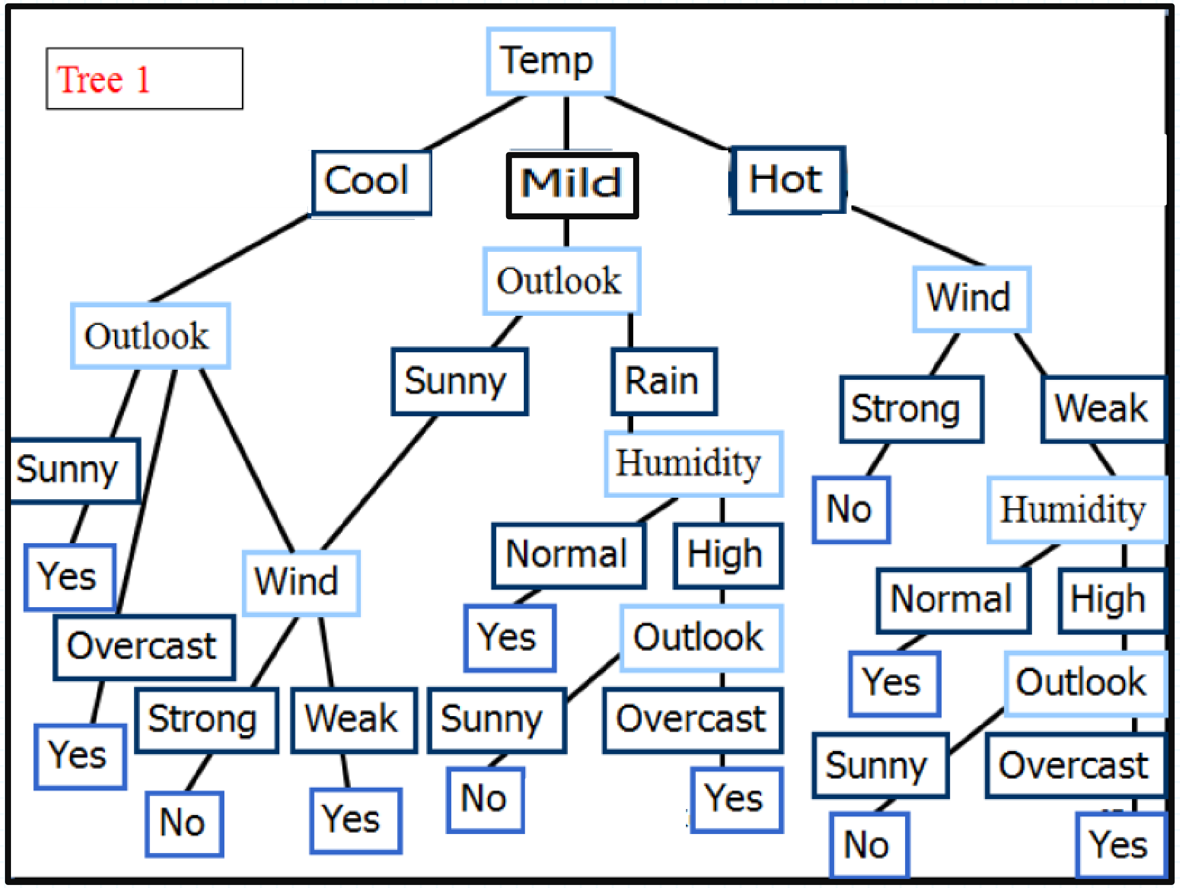
$\boxed{\text{choose best attribute } x_j}$

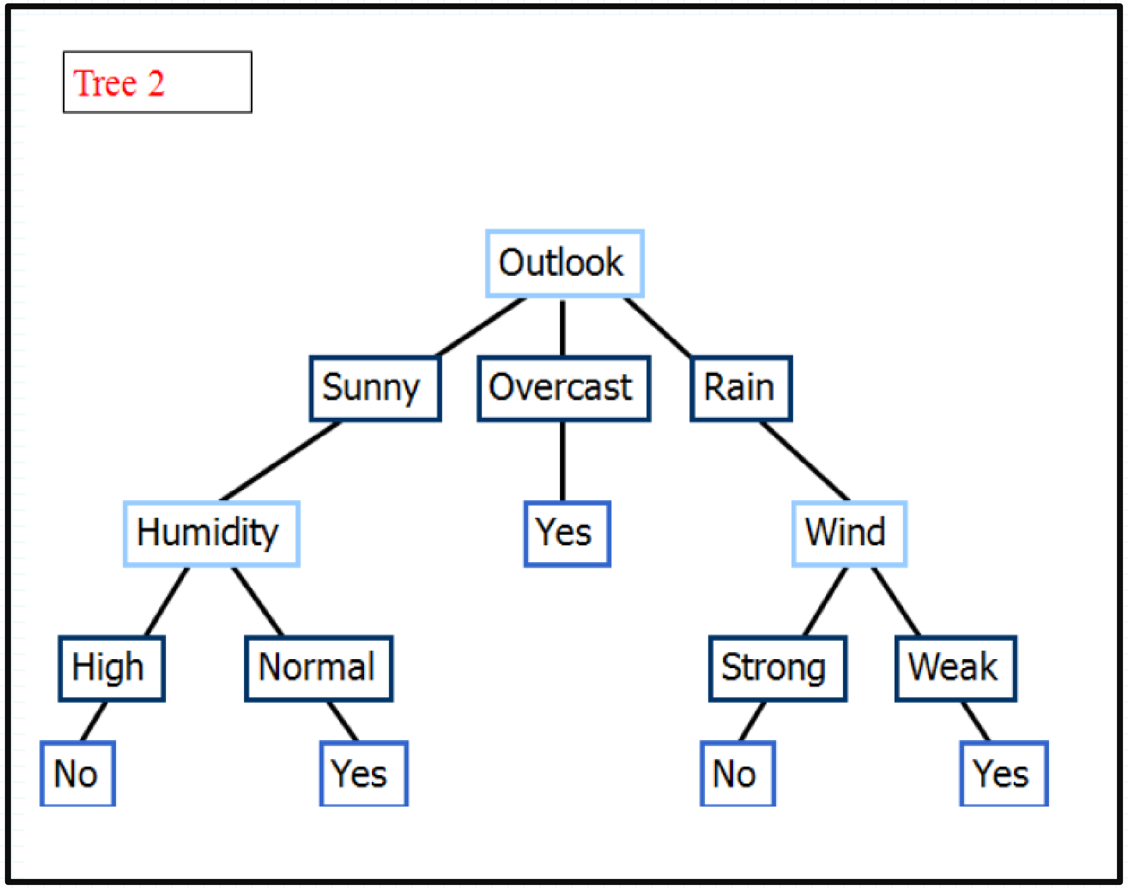
$S_0 = \text{all } \langle \mathbf{x}, y \rangle \in S \text{ with } x_j = 0;$

$S_1 = \text{all } \langle \mathbf{x}, y \rangle \in S \text{ with } x_j = 1;$

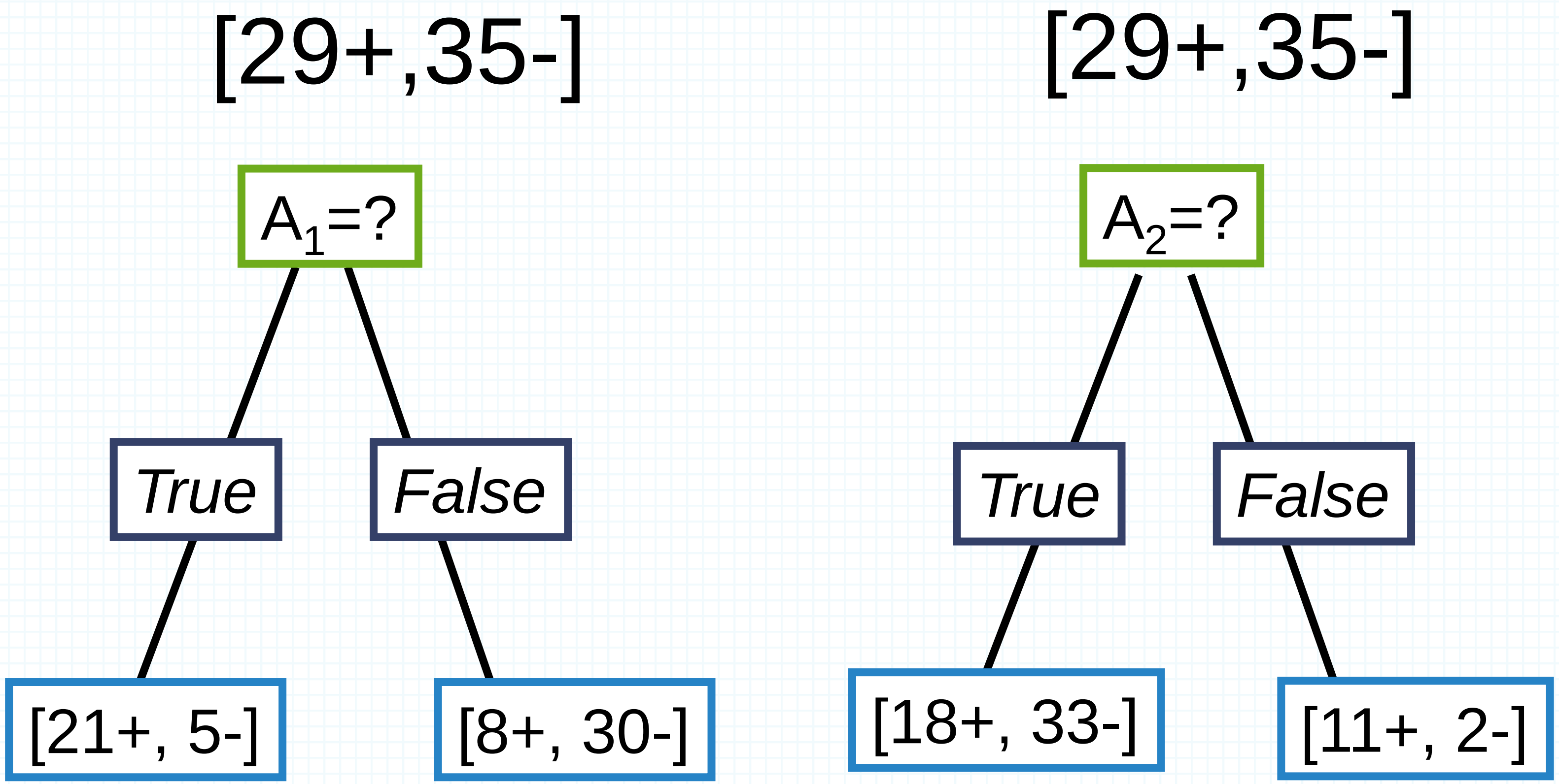
return new node$(x_j, \textsc{GrowTree}(S_0), \textsc{GrowTree}(S_1))$

Which Attribute is "best"?





Which Attribute is "best"?



Using Absolute error

CHOOSEBESTATTRIBUTE(S)

choose j to minimize J_j, computed as follows:

$S_0 =$ all $\langle \mathbf{x}, y \rangle \in S$ with $x_j = 0$;

$S_1 =$ all $\langle \mathbf{x}, y \rangle \in S$ with $x_j = 1$;

$y_0 =$ the most common value of y in S_0

$y_1 =$ the most common value of y in S_1

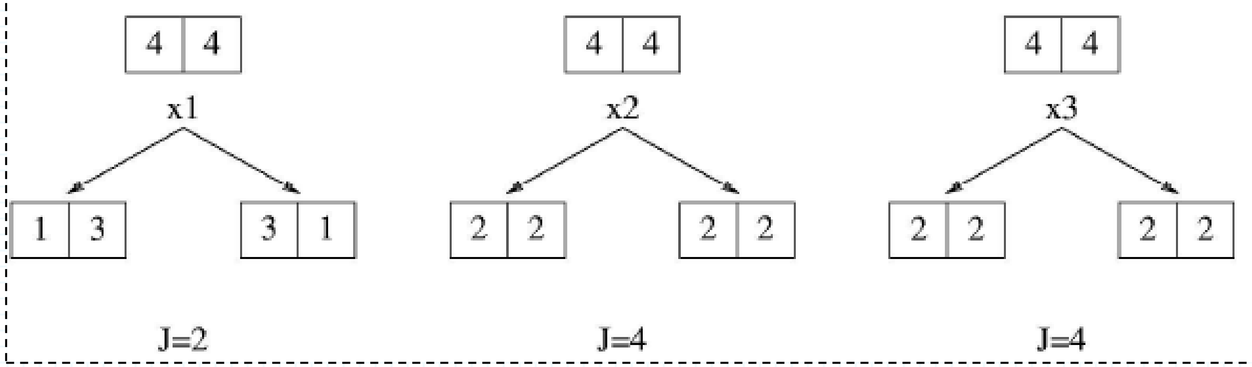
$J_0 =$ number of examples $\langle \mathbf{x}, y \rangle \in S_0$ with $y \neq y_0$

$J_1 =$ number of examples $\langle \mathbf{x}, y \rangle \in S_1$ with $y \neq y_1$

$J_j = J_0 + J_1$ (total errors if we split on this feature)

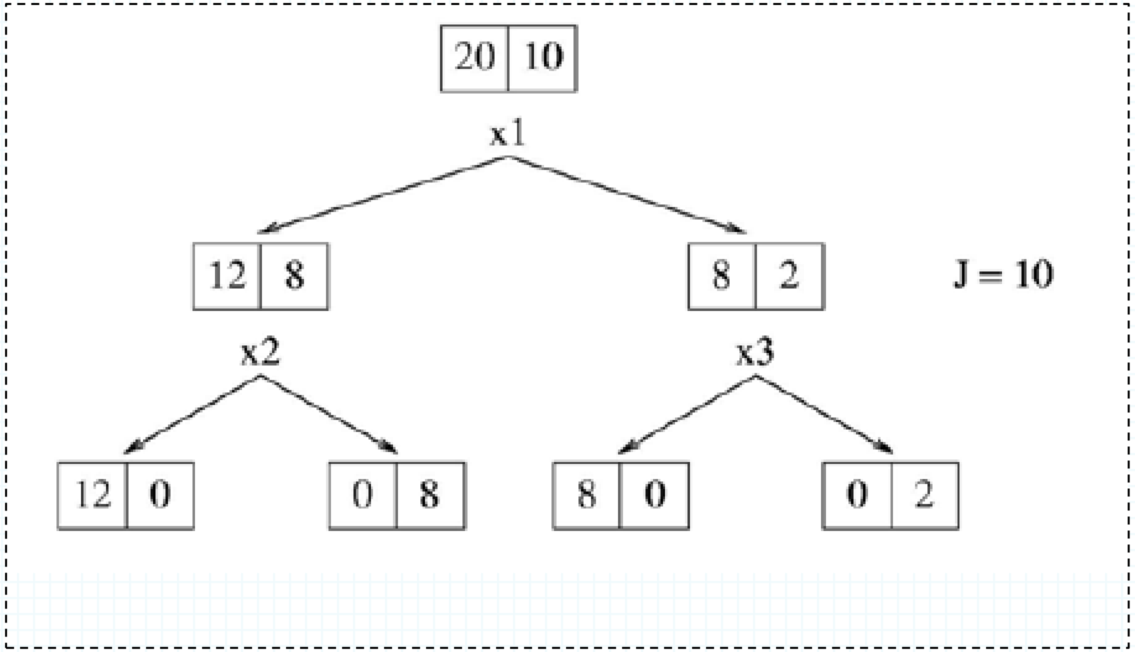
return j

x_1	x_2	x_3	y
0	0	0	1
0	0	1	0
0	1	0	1
0	1	1	1
1	0	0	0
1	0	1	1
1	1	0	0
1	1	1	0

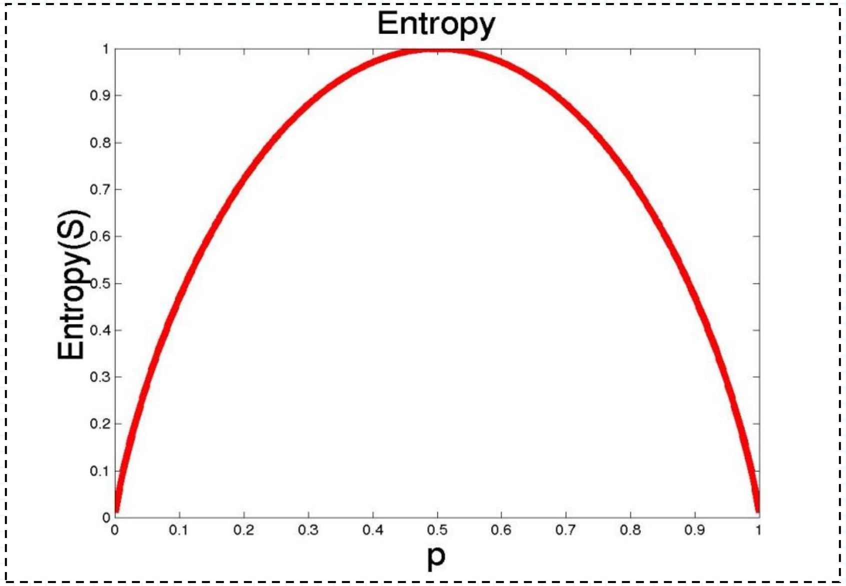


Using Absolute error

BUT…



Entropy



- S is a sample of training examples
- p_+ is the proportion of positive examples
- p_- is the proportion of negative examples
- Entropy measures the impurity of S

Entropy(S) = $-p_+ \log_2 p_+ - p_- \log_2 p_-$

Information Gain

Entropy([21+,5-]) = 0.71

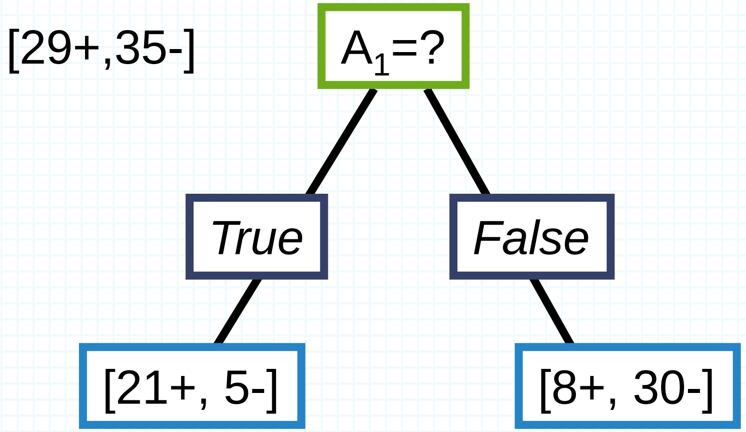
Entropy([8+,30-]) = 0.74

Gain(S,A_1)=Entropy(S)

-26/64*Entropy([21+,5-])

-38/64*Entropy([8+,30-])

=0.27



Entropy([18+,33-]) = 0.94

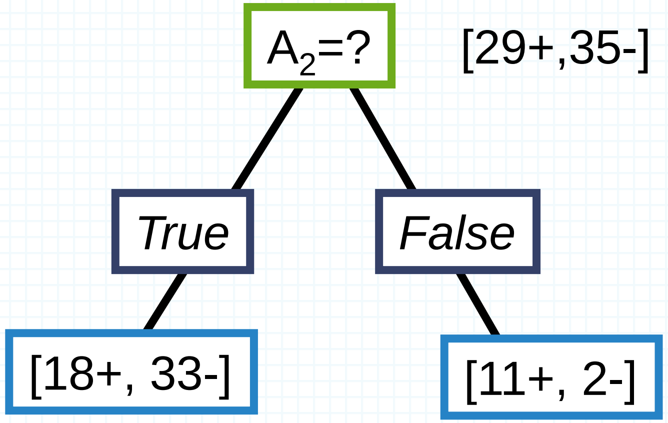
Entropy([8+,30-]) = 0.62

Gain(S,A_2)=Entropy(S)

-51/64*Entropy([18+,33-])

-13/64*Entropy([11+,2-])

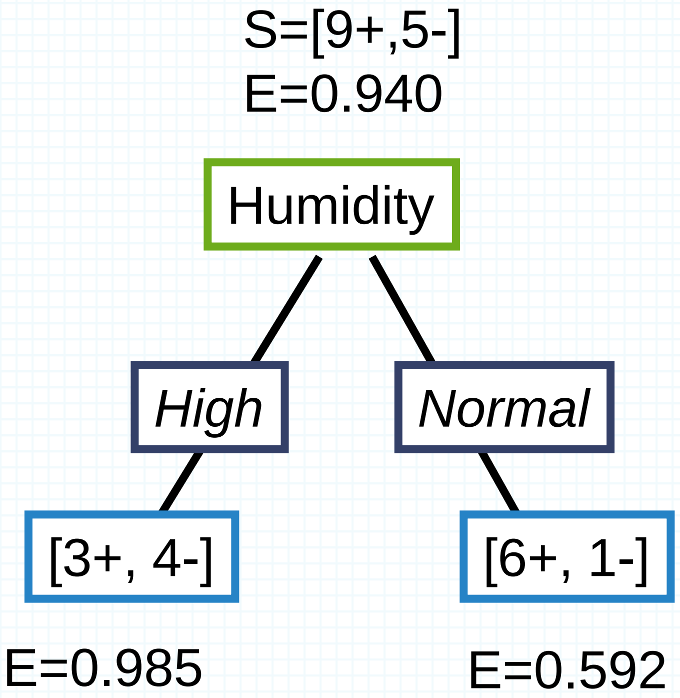
=0.12



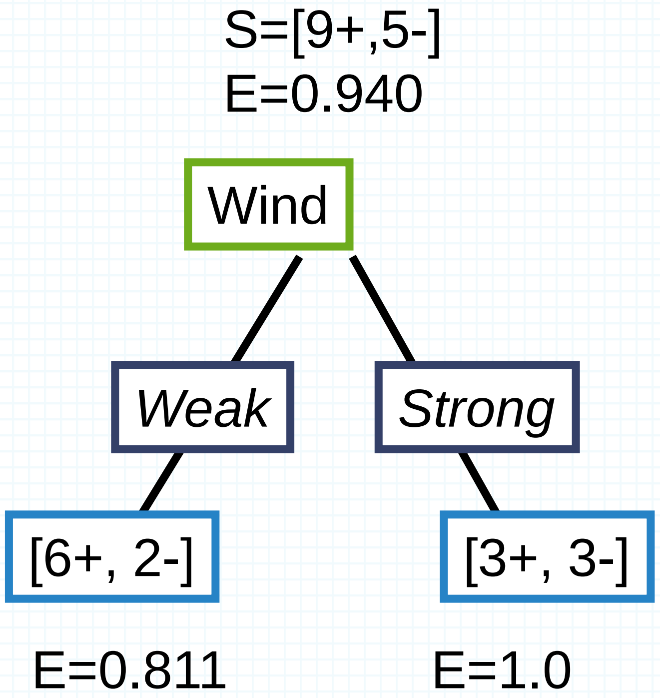
Training Examples

Day	Outlook	Temp.	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Weak	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Strong	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Selecting the Next Attribute



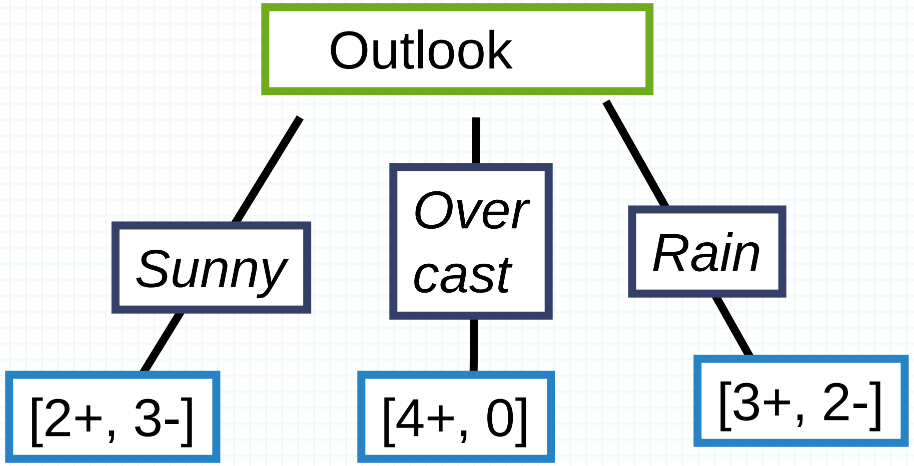
Gain(S,Humidity)
=0.940-(7/14)*0.985
– (7/14)*0.592
=0.151



Gain(S,Wind)
=0.940-(8/14)*0.811
– (6/14)*1.0
=0.048

Selecting the Next Attribute

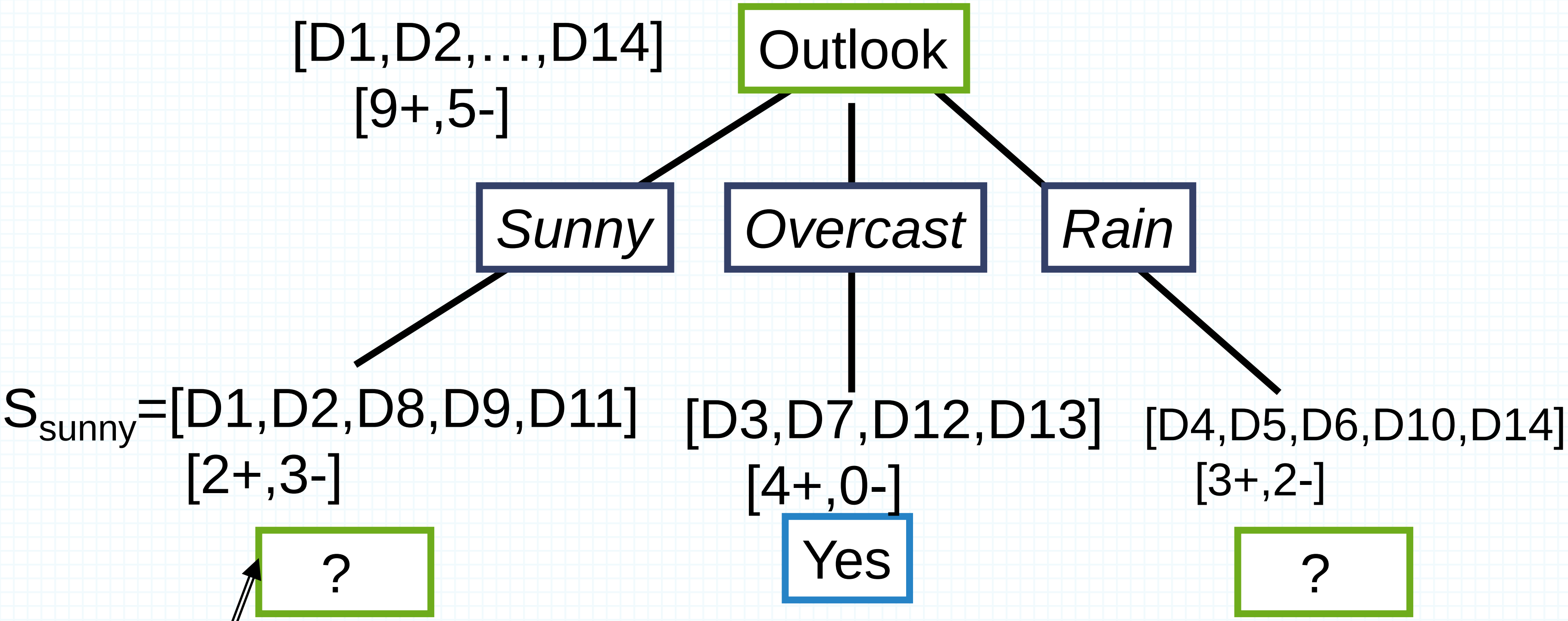
S=[9+,5-]
E=0.940



E=0.971 E=0.0 E=0.971

Gain(S,Outlook)
=0.940-(5/14)*0.971
-(4/14)*0.0 – (5/14)*0.0971
=0.247

ID3 Algorithm



Gain(S_{sunny} , Humidity)=0.970-(3/5)0.0 – 2/5(0.0) = 0.970

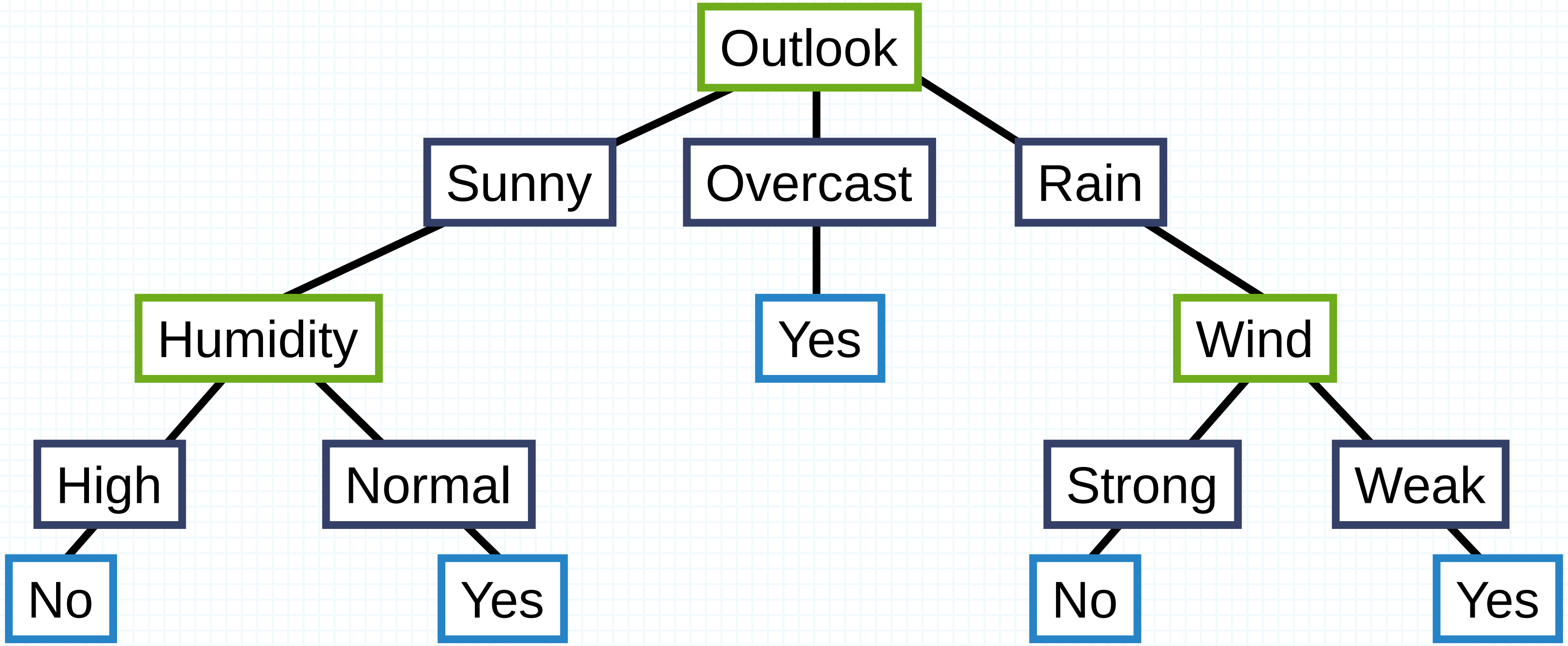
Gain(S_{sunny} , Temp.)=0.970-(2/5)0.0 –2/5(1.0)-(1/5)0.0 = 0.570

Gain(S_{sunny} , Wind)=0.970= -(2/5)1.0 – 3/5(0.918) = 0.019

When should we stop

- All samples for a given node belong to the same class.
- There are no remaining attributes for further partitioning.
- There are no samples left

Converting a Tree to Rules



R_1: If (Outlook=Sunny) $\wedge$ (Humidity=High) Then PlayTennis=No
R_2: If (Outlook=Sunny) $\wedge$ (Humidity=Normal) Then PlayTennis=Yes
R_3: If (Outlook=Overcast) Then PlayTennis=Yes
R_4: If (Outlook=Rain) $\wedge$ (Wind=Strong) Then PlayTennis=No
R_5: If (Outlook=Rain) $\wedge$ (Wind=Weak) Then PlayTennis=Yes