

'Decision Tree'

Decision Tree

- Rule Based algorithm.
- Interpretable

Black Box.

→ 'Neural Nets'.

→ Financial transactions

Dataset

SL	PL	species.
----	----	----------

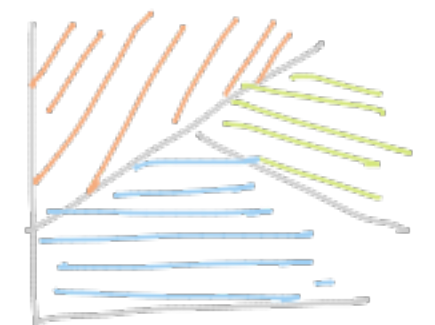
—	—	versicolor
—	—	virginica
—	—	setosa

'Partitioning the feature space into pure regions, each belonging to a specific class'.

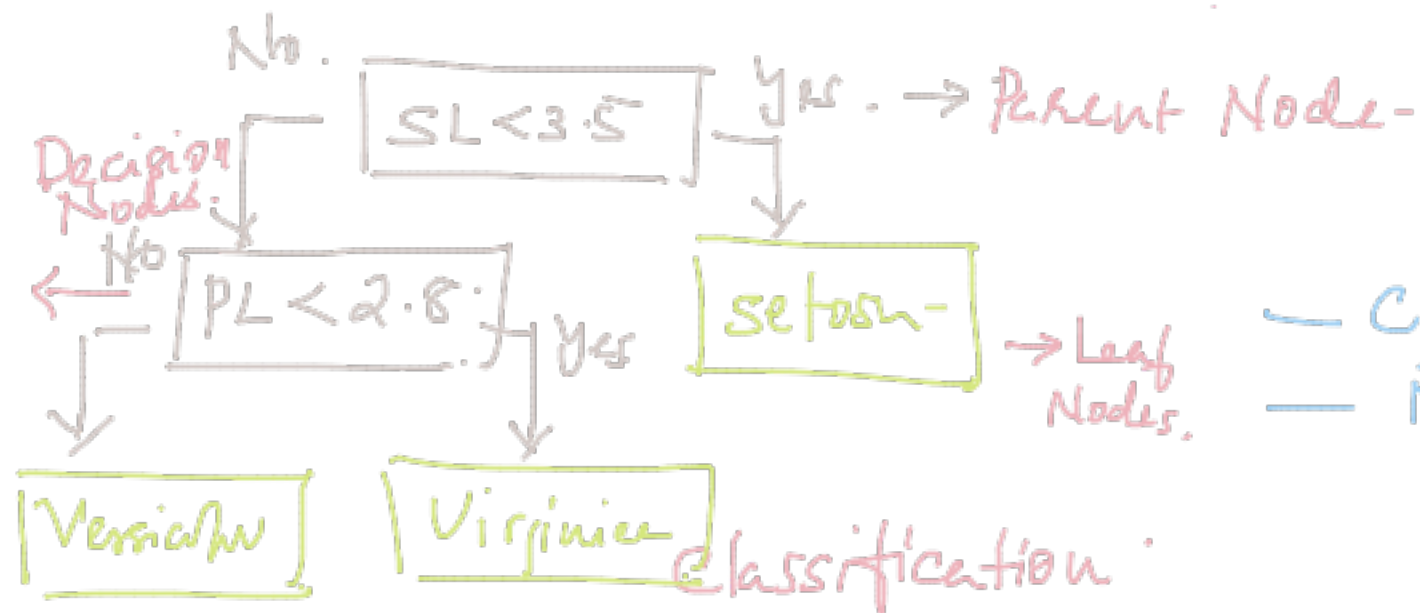
if $SL < 3.5$
setosa.

else.

$PL < 2.8$
virginica
else
versicolor

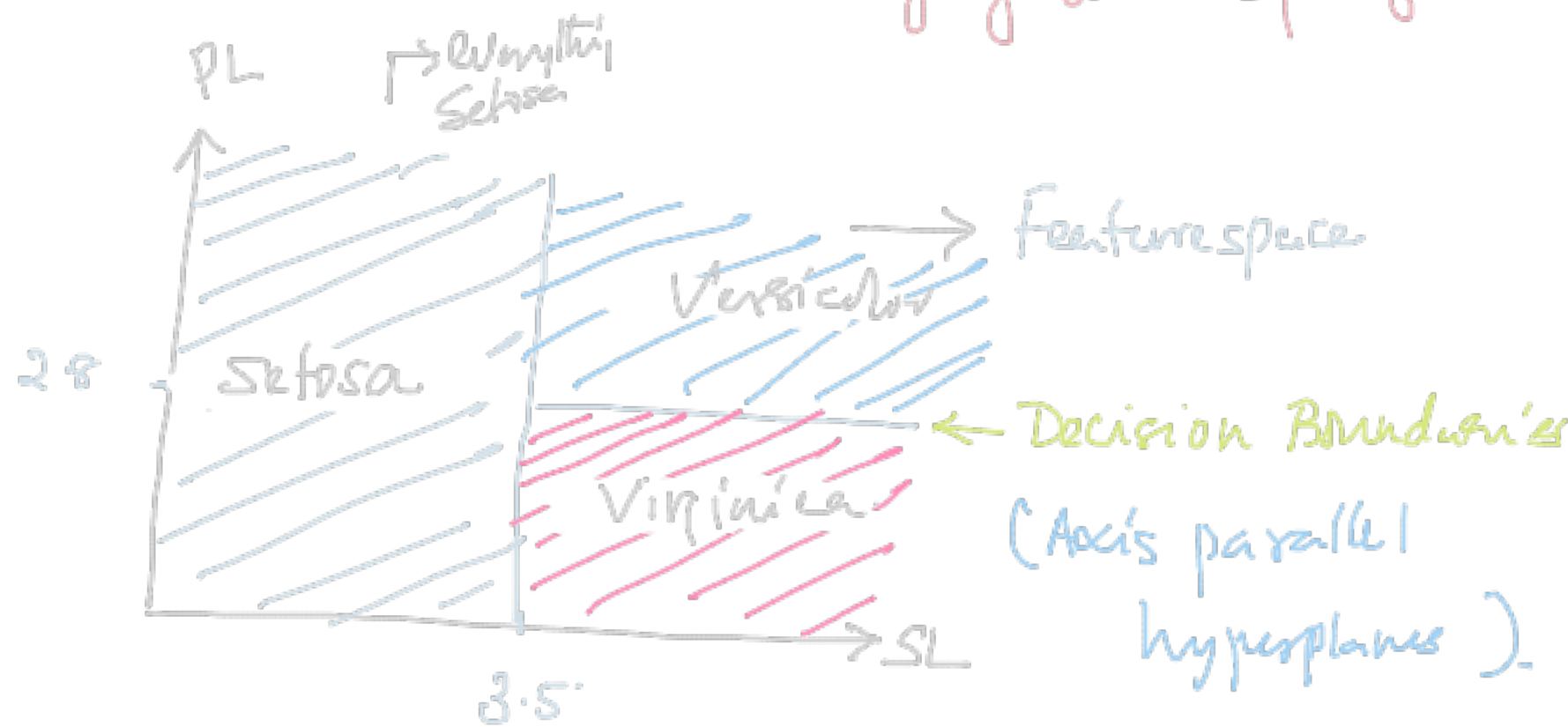


Not Axis parallel.



- Classification ✓
- Regression

Classification



Entropy & Information Gain

— Degree of uncertainty in the data.

Feature X \longrightarrow		Y \longleftarrow
Size (x)	Color (X)	Fur
3.5	Red	No
4.3	Red	No
5.8	Red	Yes
6.3	Red	Yes

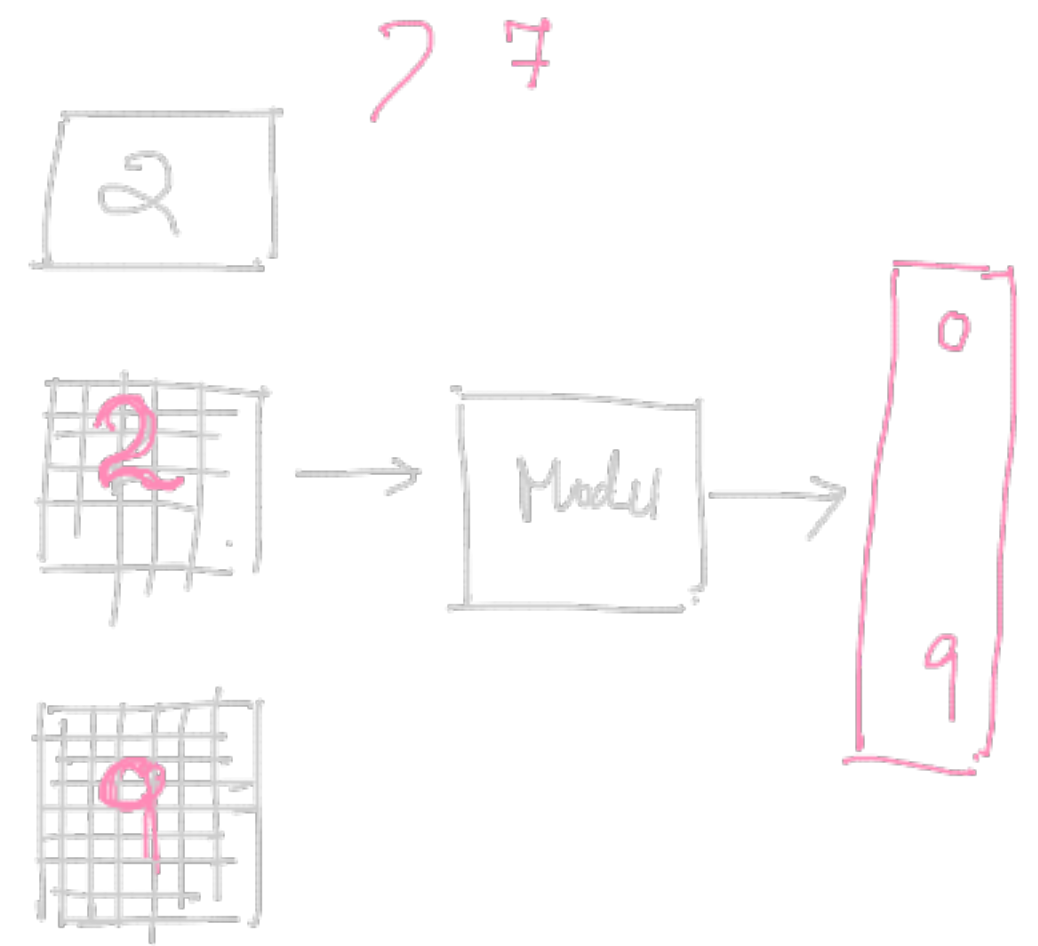
Entropy high Entropy zero

$$\text{Entropy} = - \sum_{i=1}^k p_i \log(p_i).$$

k - No. of classes in 'y'

$$H = - \left[p_{\text{yes}} \log(p_{\text{yes}}) + p_{\text{no}} \log(p_{\text{no}}) \right]$$

$\begin{matrix} (-0.6) & & (-0) \end{matrix}$



$$\text{Information Gain (IG)} = \text{Entropy}_{\text{Parent}} - \text{Weighted Entropy}_{\text{child groups.}}$$

2 7 3

$$= 0.94 - \left[\frac{5}{14} \times 0.97 + \frac{4}{14} \times 0 + \frac{5}{14} \times 0.97 \right]$$

$$= 0.94 - 0.69,$$

$$\text{IG} = 0.25$$

Outlook Temp Humidity Windy

overfitting \leftarrow

10.45 Am

\rightarrow

Yes
Yes
Yes
Yes
Yes

Pure

$$I_G: 0.97 - 0$$

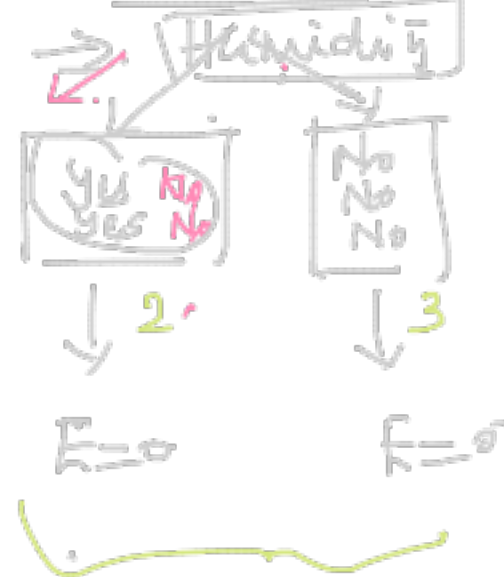
$I_G \rightarrow$

Maximum (I_G)

Minimum (Impurity)

Depth

$$E = 0.97$$



Pure Node

Outlook

Sunny

Rainy

yes
yes
yes
yes

leaf

3 Yes.
2 No.

Windy

3 Yes

2 No

5 leaf nodes

Classification
Model

Ans.

Sehen
Sehen
Verschieden

Verschieden

V_i

Minimum
Impurity
Maximum I_G

$$I_G = 1 - \sum_{i=1}^k (P_i)^2$$

minimize

Cross Impurity

$I_G \Rightarrow$ Information
Gain

\rightarrow (Maximize)