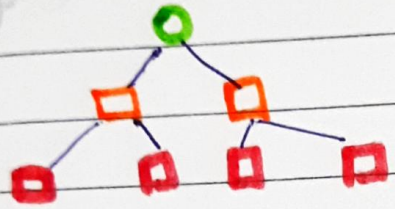
Mo Tu We Th Fr Sa Su

Date

Decision Tree Algorithm -
(Regression & classifier)

ID3 → Iterative Dichotomiser 3 Algorithm



⇒ Information Gain
⇒ Entropy

Entropy - Randomness, uncertainity, disorder, in your data.

$$\text{Entropy} = -\sum P(x) \log_2 P(x)$$

P(x): fraction of data instances of particular type.

Mo Tu We Th Fr Sa Su

Date / /

Example-

Gender
Male
M
M
F

Entropy(Gender) = $\frac{-3}{4}\log_2(\frac{3}{4}) - \frac{1}{4}\log_2(\frac{1}{4})$

if M=2, F=2 Entropy = $-0.5\log_2(0.5) - 0.5\log_2(0.5)$

$= 1$

if Dataset is balanced, Entropy will be 1.

Information Gain - how much info a feature/ input parameter gives us about the output / target parameter.

Info Gain (P_i) = Entropy(target) - Entropy$(P_i, target)$