Decision Tree

① Decision Tree classification [classification]

② Decision Tree Regression [Regression]

① Decision Tree Classification =)

① ↑ ID3 [Iterate Dichotomiser 3]

② ↓ CART [classification And Regression Tree]
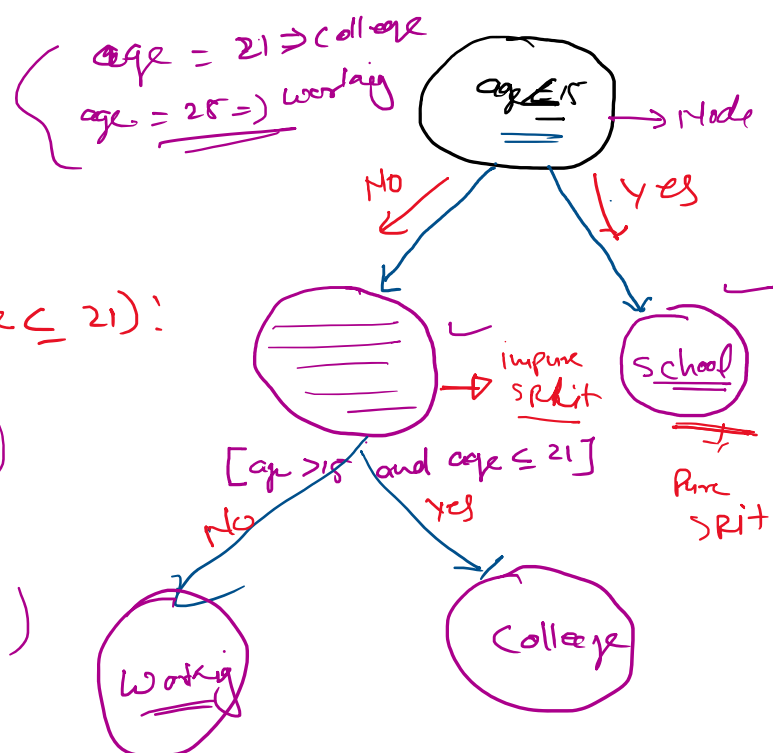
scikit-learn

age = 14

```
if (age ≤ 15):
    Print ("school")

elif (age > 15 and age ≤ 21):
    Print ("college")

else:
    Print ("working")
```
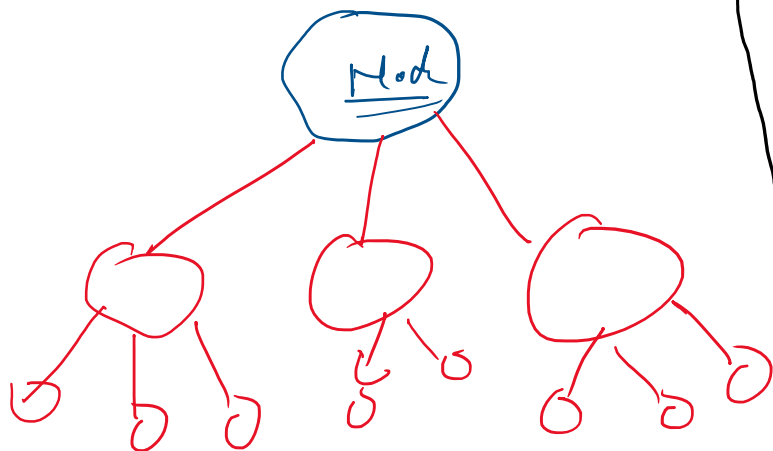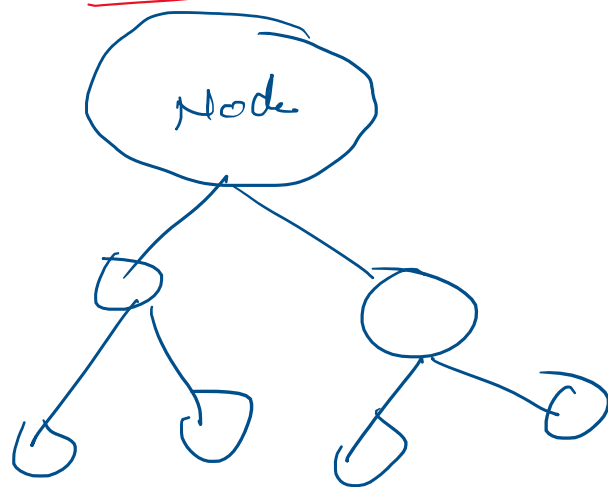
age = 21 ⇒ college
age = 25 ⇒ working

age ≤ 15 → Node

No / Yes

impure split

School → Pure split

[age > 15 and age ≤ 21]

No / Yes

Working   College

ID3

Node

CART

Node

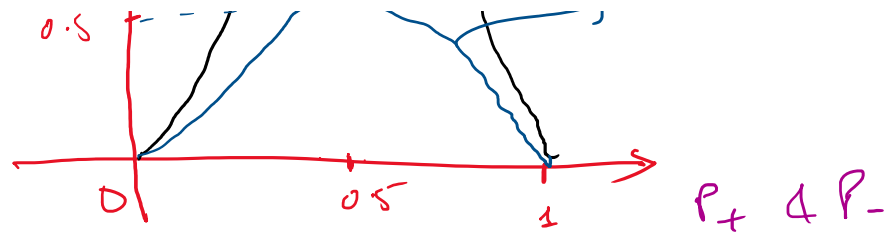① Purity check : Pure split or Impure split

↳ Entropy
↳ Gini Impurity

① Entropy

$$H(S) = -P_+ \log_2 P_+ - P_- \log_2 P_-$$

② Gini Impurity

$$G.I = 1 - \sum_{i=1}^{N} (P_i)^2$$

H(S)

1

0.5

→ Entropy

→ Gini Impurity

$$0.5$$

$$0 \qquad 0.5 \qquad 1 \qquad P_+ \; \& \; P_-$$

$$\Big\{ \quad \underline{\text{Entropy}} \quad \text{Vs} \quad \underline{\text{Gini Impurity}} \quad \Big\}$$

$$\downarrow \qquad\qquad\qquad \downarrow$$

small data set   large data set

Decision Tree Post Pruning & Pre Pruning

(1) Post Pruning
(2) Pre Pruning

(1) Post Pruning

small dataset

$\rightarrow 0$

$\boxed{1}$

$\boxed{1}$

$\boxed{3}$

$\boxed{9 \mid 21}$

$\rightarrow u$
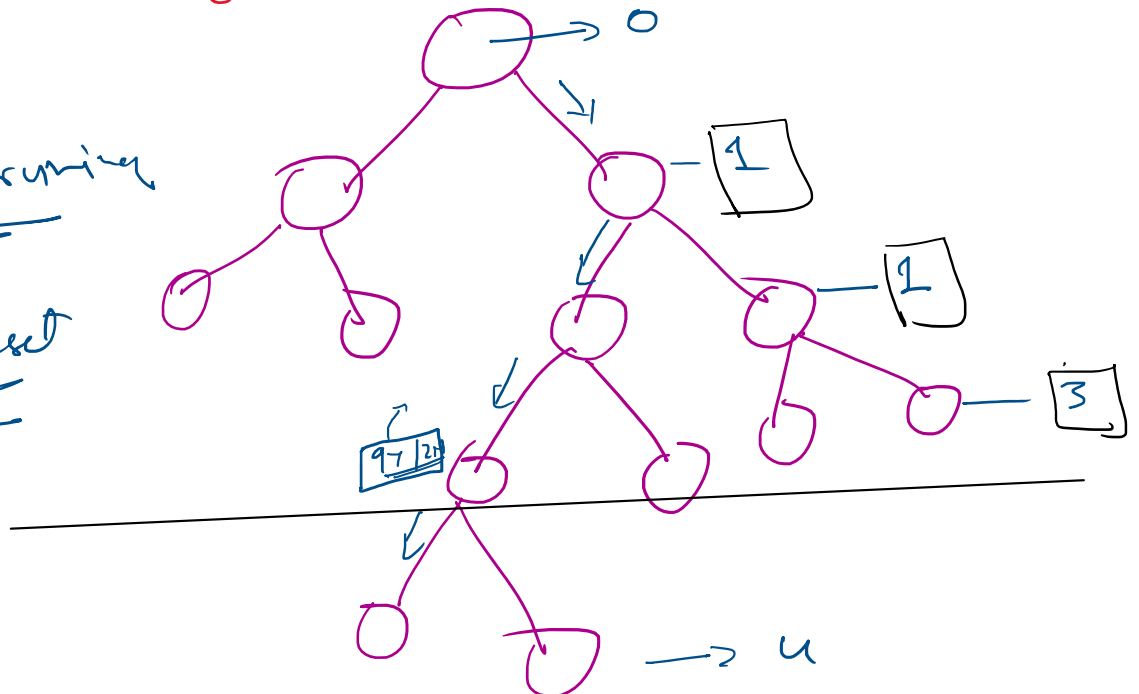
(1) Pre Pruning → Hyper parameter tuning
$\underline{||}$

Large data set