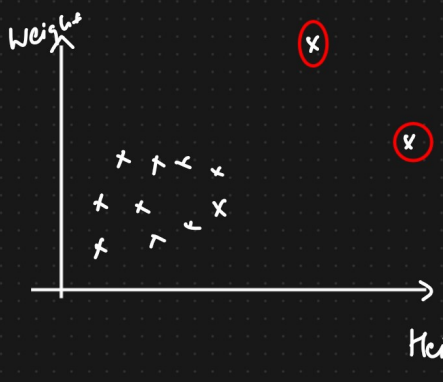
Anomaly Detection [To detect Outliers]

↳ Play a important Role



IPL	
1	15
2	10
3	12
4	100

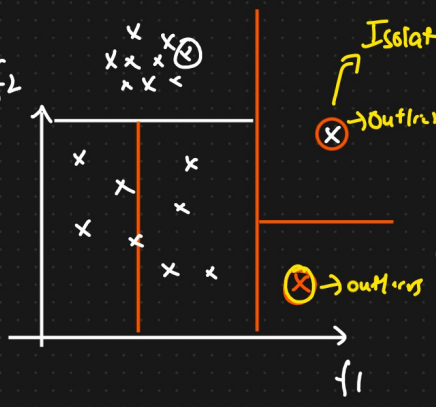
>36

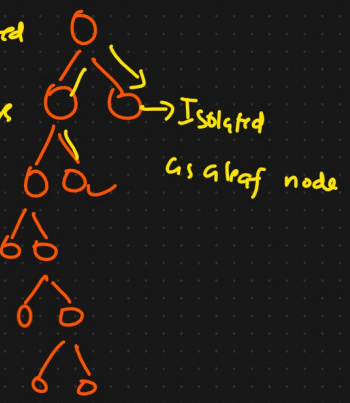
① Isolation Forest [Decision Trees].

Many Trees

Isolated Trees

f_1	f_2	f_3	f_n
-	-	-	-
-	-	-	-
-	-	-	-
-	-	-	-





Anomaly Score

Mathematical Formula : Compute anomaly score for a new point

$$S(x,m) = 2^{-\frac{E(h(x))}{c(m)}}$$

m = no. of data points

x = Data point.

$E(h(x))$ = Average search depth for x from the isolate tree.

$c(m)$ = Average depth of $h(x)$

[Threshold >0.5]

$E(h(x)) << c(m) \Rightarrow S(x,m) \approx 1$. ⇒ Anomaly score ⇒ Outliers

$E(h(x)) >> c(m) \Rightarrow S(x,m) \approx 0.5$ ⇒ Normal data point.