Laplace smoothing

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)}$$

spam / ham

$$P(s=1 \mid w_i) = \frac{\overbrace{P(w_i \mid s=1)}^{\text{likelihood}}\, P(s=1)}{P(w_i)}$$

"today is a great day"

$$P(w_i \mid S=1) = P(\text{today}, \text{is}, \text{a}, \text{good}, \text{day} \ldots \mid S=1)$$

$$= \boxed{P(\text{today} \mid S=1)}\, P(\text{is} \mid S=1) \ldots P(\text{day} \mid S=1)$$

$$P(\text{today} \mid S=1) = \frac{0}{1000\ (\text{spam})}$$

$$\Rightarrow P(S=1 \mid w_i) = 0$$

$$K = 1$$

$$P(w_1 \mid c) = \frac{\#\text{words} + K}{|F|}$$

$$K = 1$$

$$P(w_1 \mid c = s) = \frac{\#\text{ of times } w_1 \text{ appears in class } (s) + K}{\#\text{ of words in class } (s) + (Kn)}$$

$n \equiv$ number of classes

$k \equiv$ smoothing value

$$P(\text{today} \mid s = 1) = \frac{0+1}{1000 + (1)(2)}$$

$$= \boxed{1/1002}$$

$$P(\text{day} \mid S=1) = \frac{20}{1000}$$ (orignl)

$$\frac{20+1}{1000+(1)(2)} = \frac{21}{1002}$$

(smoothed)

$k = 0.01$

$$P(\text{today} \mid S=1) = \frac{0 + 0.01}{1000 + (0.01)(2)}$$

Decision Trees

- divide & conquer

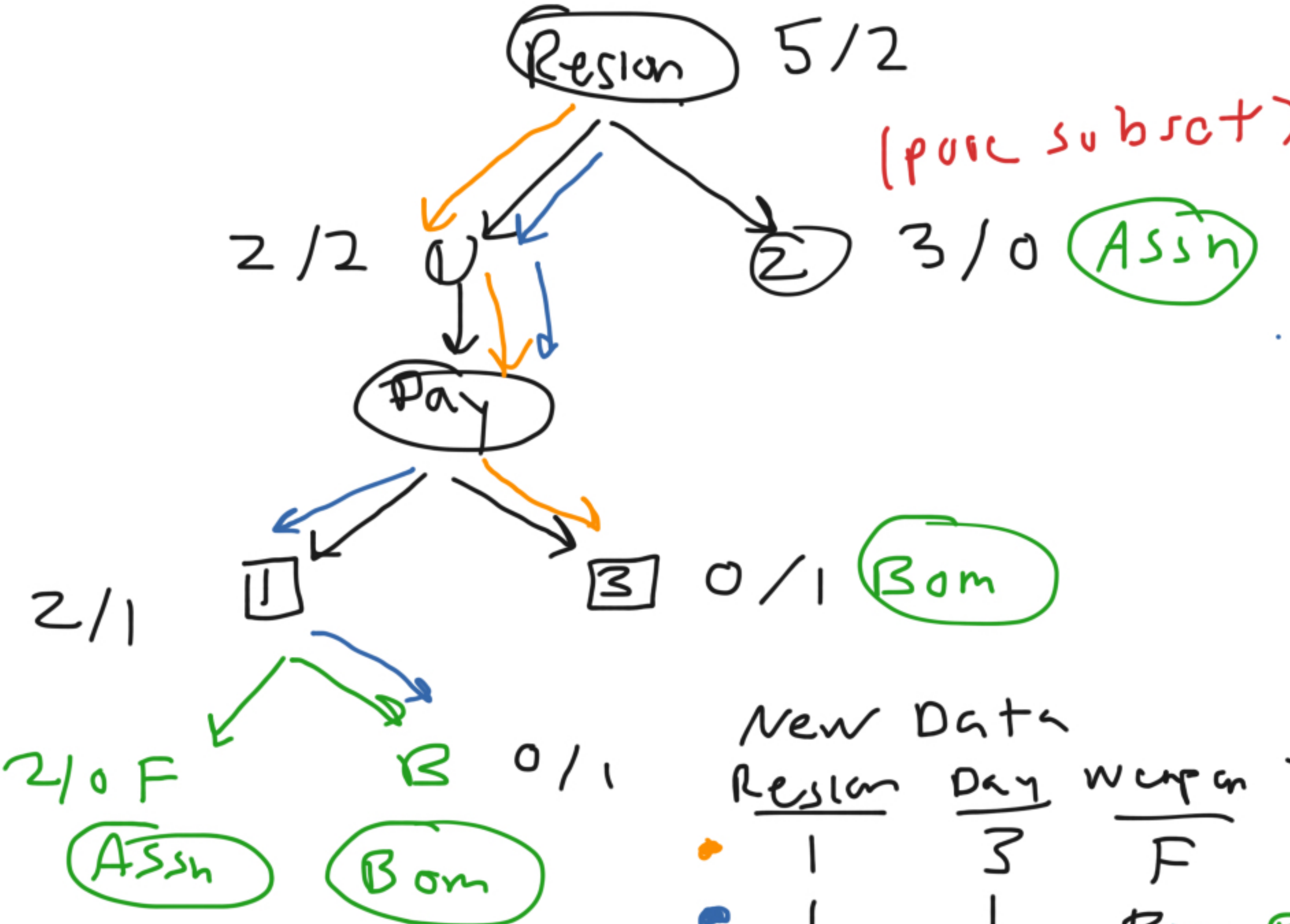
- Build a classifier that will predict attack type

→ Features: Region, Day, Weapon

→ Target: Attack type $\hat{y} = \hat{f}(x)$
(Assn, Bom)

Region	Day	Weapon	Type
1	1	F	Assn
2	2	B	Assn
1	3	F	Bom
1	1	F	Assn
2	2	F	Assn
1	1	B	Bom
2	20	B	Assn

1. Region Day
2. Day Weapon
3. Region Weapon



New Data

Region	Day	Weapon	Type
1	3	F	Bom
1	1	B	Bom

CART / ID3 Algorithm

→ choose the best feature on the basis of "information gain".

→ create a new child node for each feature

→ for each child node

if (subset is pure) {stop)

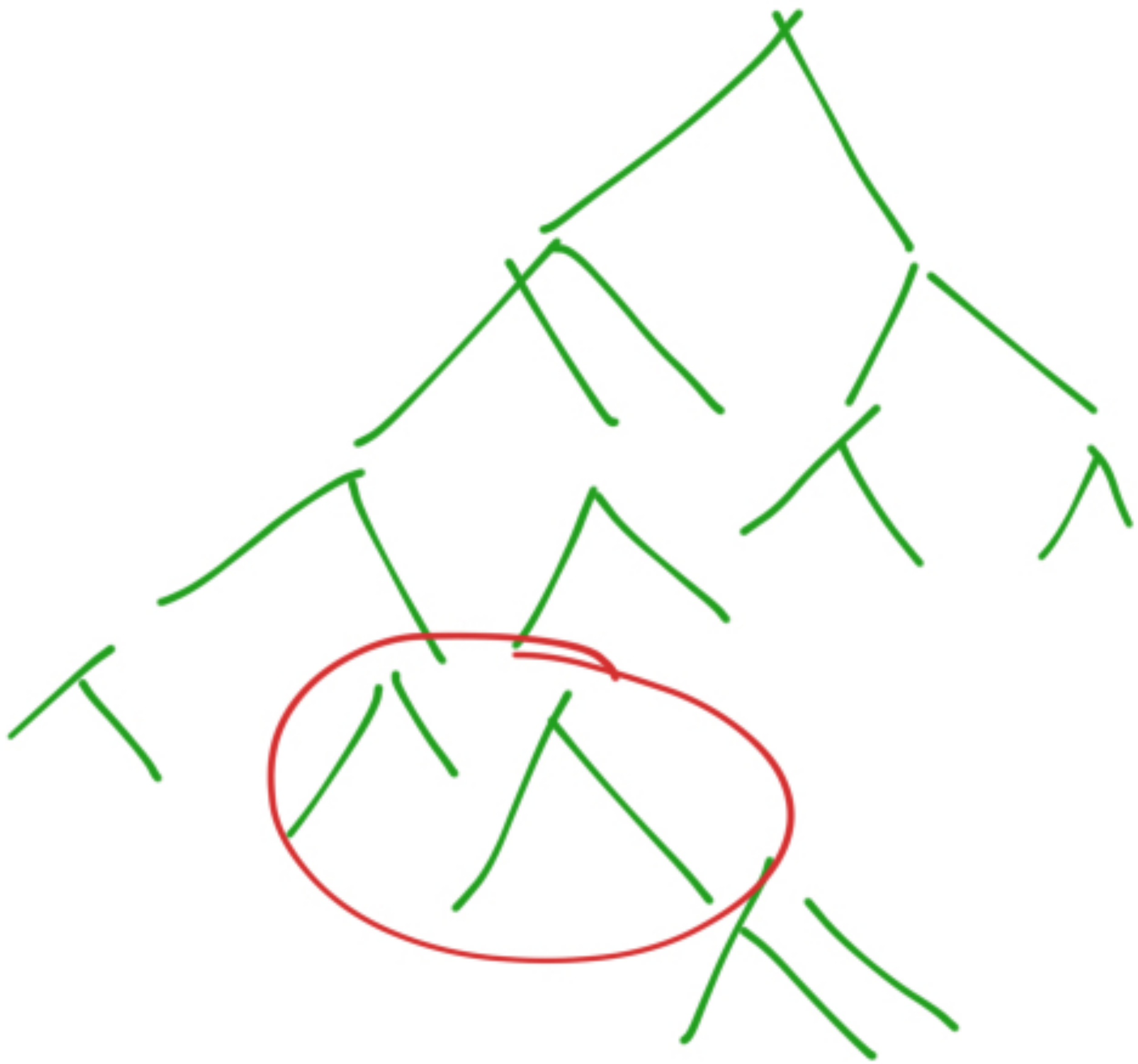
else split(node, examples)

why DTs are great

1. high-performance
2. interpretable / transparent

why DT's stink

1. overfit (pruning) (random forests)
2. computationally intensive
(gradient boosting)

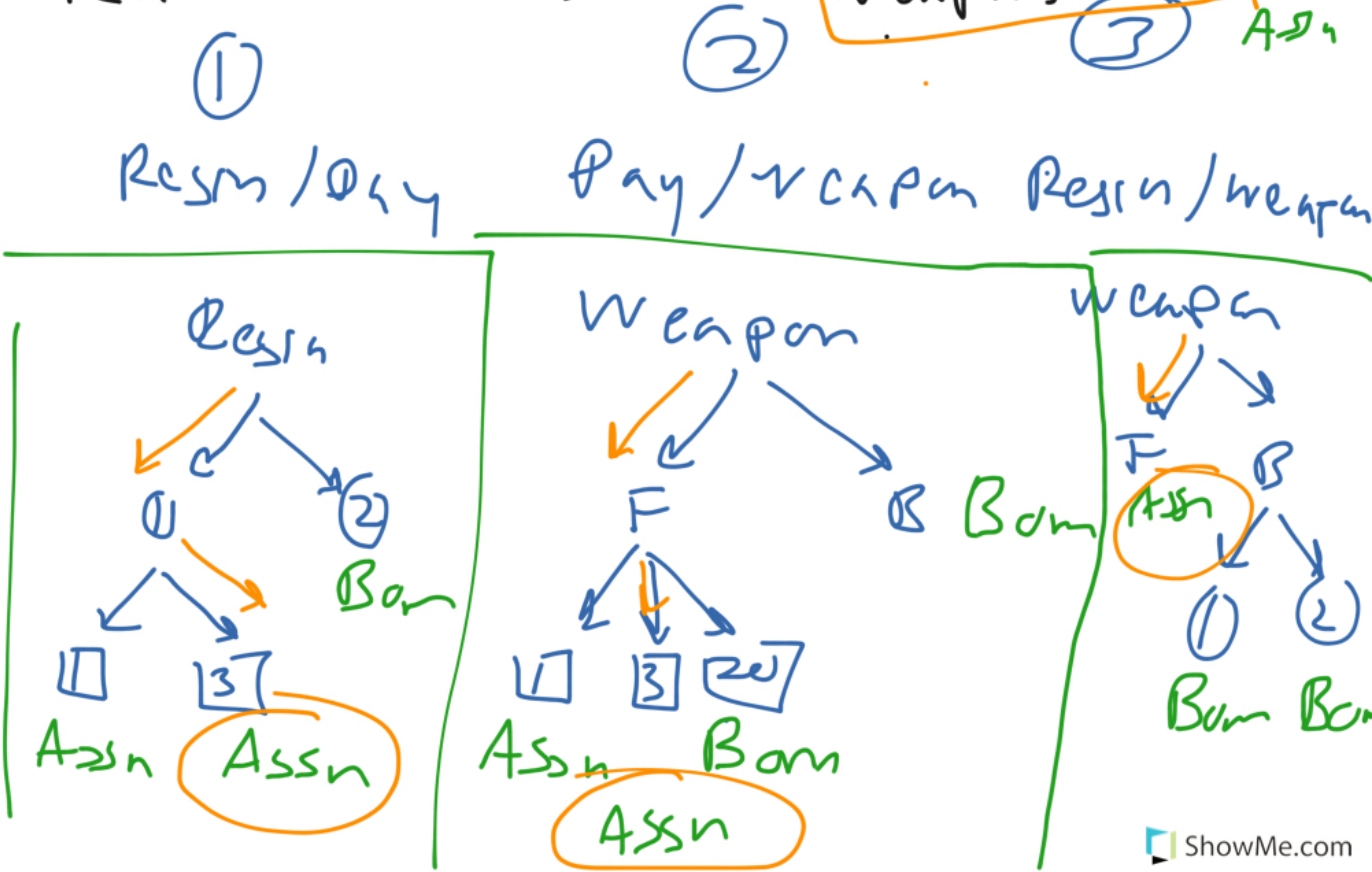


Pruning

Random Forests

Region 1, Day 3,
Weapon F

Asn



1	2	3	Label
Assn	Assn	Bom	Assn
"	Bom	"	Bom