Laplace Smoothing

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

$$P(C=1|w_i) \propto P(w_i|C=1)\,P(C=1)$$

"let's get together for drinks!"

$P(w_i \mid C=1) =$

$P(\text{lets, get, together} \ldots \mid C=1) =$

$\Rightarrow P(\text{lets} \mid C=1) \cdot P(\text{get} \mid C=1)$

$\cdot P(\text{together} \mid C=1)$

(C=1)

$P(\text{lets} \mid C=1) = \frac{0}{1000}$ In Spam emails there are 1000 words

$P(\text{get} \mid C=1) = \frac{10}{1000}$ $P(\text{together} \mid C=1) = \frac{2}{1000}$

$\left(\frac{0}{1001}\right)\left(\frac{10}{1001}\right)\left(\frac{2}{1001}\right)$ ← $P(w_i|c=1)$

$P(c=1|w_i) = 0$

K

1

$P(w_i|c=1)$ count of word w_i + K

total words in class + (K)(# of C)

K=1

$\frac{0+1}{1000+(1)(2)} = \left(\frac{1}{1002}\right)\left(\frac{12}{1002}\right)\left(\frac{4}{1002}\right)$

: -

"publication pending payment".

$P(C=1 \mid w_i)$

$P(w_i \mid C=1) = P(\text{publication} \mid C=1)$,
$P(\text{pending} \mid C=1) \cdot P(\text{payment} \mid C=1)$

$\left(\frac{30}{1000}\right)\left(\frac{100}{1000}\right)\left(\frac{0}{1000}\right)$

Decision Trees

- divide & conquer
- with an attack be a bombing or an assassination?

Region = 1, Pay = 1

Weapon = Fire arm

Attach type ?

Training

Region	Dog	Weapon	Attack
1	1	F	Asn
2	2	B	Asn
1	2	F	Bom
1	1	F	Asn
2	2	F	Asn
1	1	B	Bom
2	2	B	Asn

5 Asn / 2

S_1 (Regm, Dog)

S_2 (Dog, Weapon)

S_3 (Region, Dog)

5 As / 2 Bom

Resim

"pure subset"

2/2 (1) (2) 3/0 Ash

Day

Decision Tree Classifier

2/1 0/1

1 3 Bom

Regim = 1

Day = 1

Weapon = F

Weapon

F 2/0 Ash

B 0/1 Bom

predicted Attach

Ash

Problems w/ decision trees

① overfitting - saved by pruning

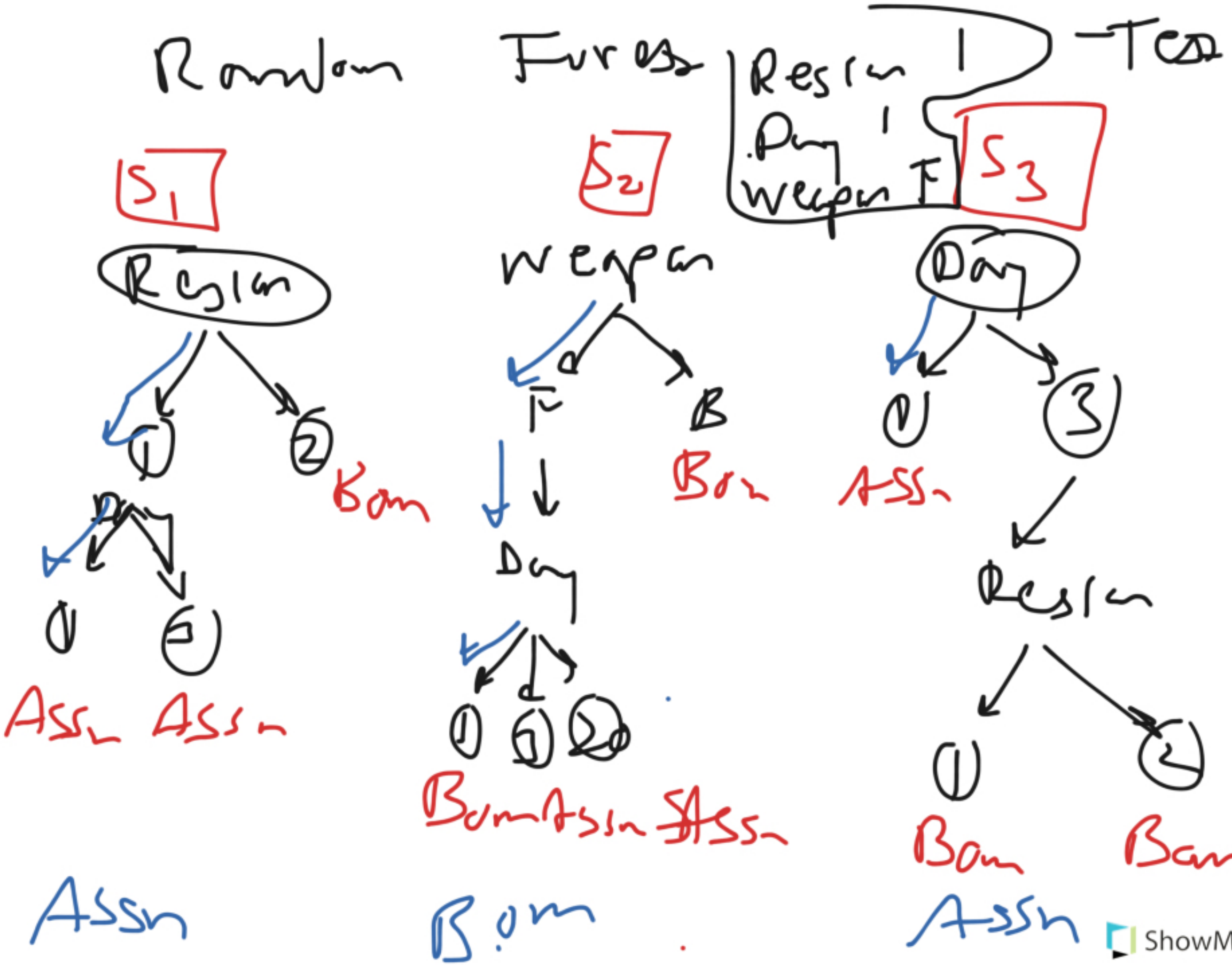
Random Forest Algorithm

→ randomly select subsets of the training data and...

→ subsets of the features,

→ grow trees for each subset

→ use the majority vote to make class


S1
S2
S3
Assn
Assn
Assn
Bom
Bom
Bom
Bom
Assn
Assn
Assn
Bom
Bom
Assn
Bom
Bom
Assn
ShowMe.com

Vote? $P(\text{Assn} \mid \text{Trees}) = \frac{2}{3}$ ✓

$P(\text{Bon} \mid \text{Trees}) = \frac{1}{3}$

Assn wins majority vote

→ Parameters

① Depth

② num trees