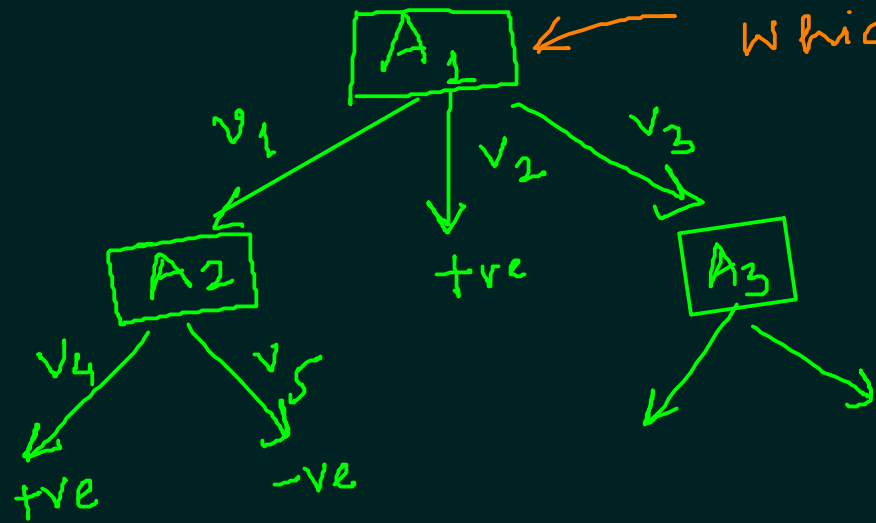


# Summary of Previous Class:

## Decision Tree Learning



Which Attribute? → Information Gain (Entropy)

↳ Gini Impurity (Gini Index)

▶ Entropy:  $E = - \sum_{c=1}^m p_c \log_2 p_c$

$$IG = E(S) - \sum \frac{|S_v|}{|S|} E(S_v) \uparrow$$

▶  $GI = 1 - \sum_{c=1}^m p_c^2$  ← computationally easy

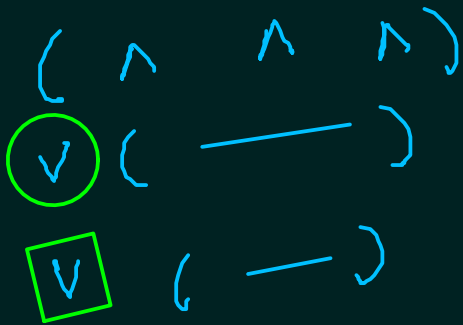
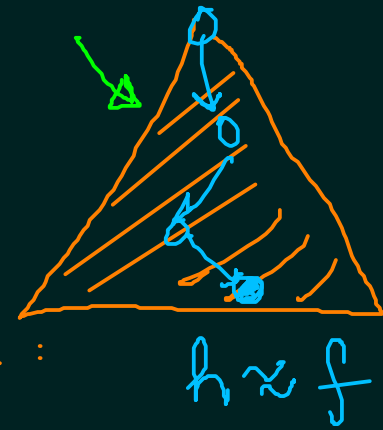
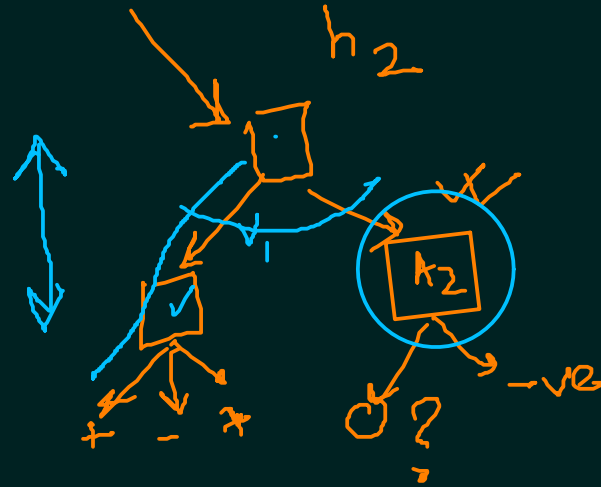
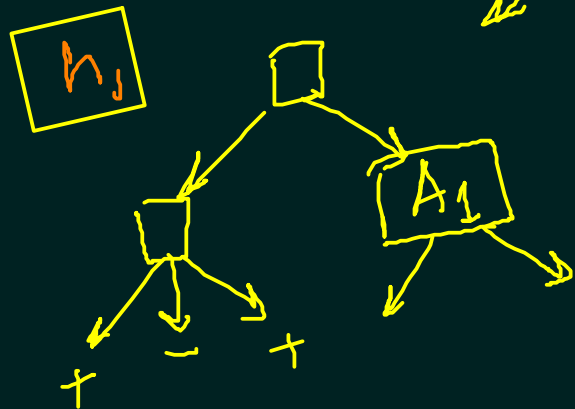
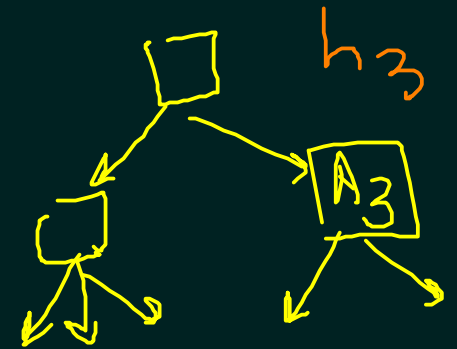
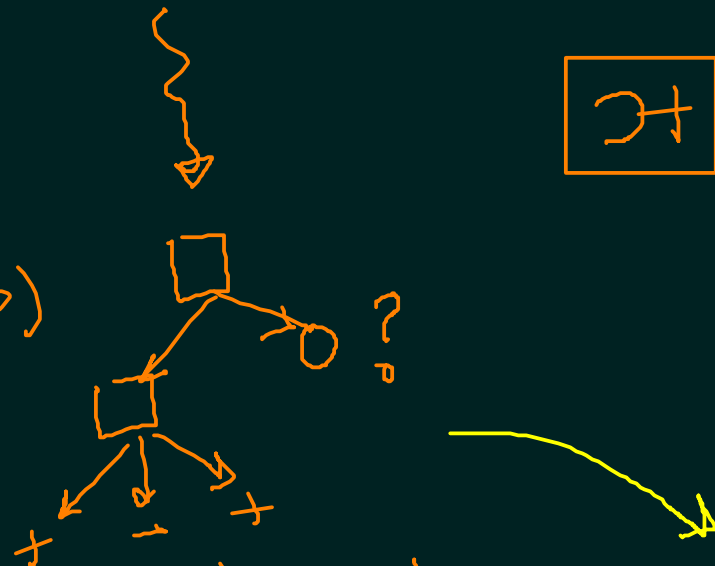
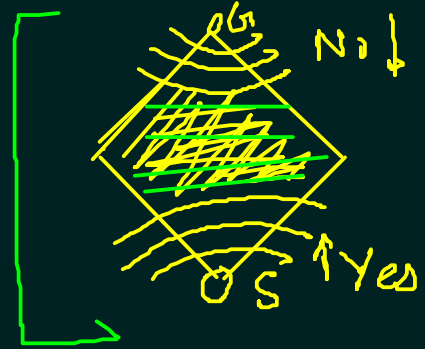
$$\left. \begin{array}{l} \mathcal{H} = \{ \text{Decision Trees} \} \\ \text{LA} = \{ \text{DTLA} \equiv \text{ID}_3 \} \end{array} \right\} \leftarrow \text{Represent } h \approx \underline{f}$$

# Searching in H.S. :

↳ Best-first  
(Prefer shorter Hyp)

↳ DONOT  
Backtrack

TC



CE lim  $\rightsquigarrow$  VS  $\rightarrow$   $h_1, h_2 \in VS$   
 Incomplete

Boolean AND

▶ Candidate Elimination

↳ Incomplete H.S.

↳ Searches Completely

Inductive } Restriction Bias  
 Bias }



Yes ↑  $\langle \emptyset \emptyset \emptyset \emptyset \rangle$   
 ↓  $\langle a_1, a_2, ?, ?, a_5, ? \rangle$

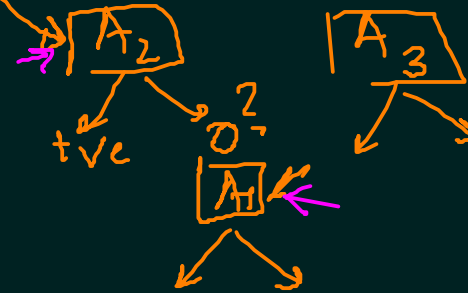
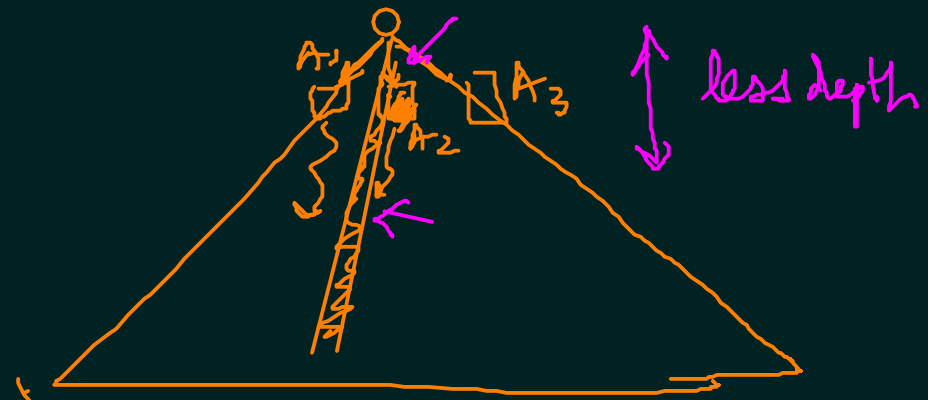
$\langle \text{wavy line} \rangle \rightarrow \text{Yes}$   
 $\langle \quad \quad \quad \rangle \rightarrow \text{NO}$

▶ Decision Tree Algo

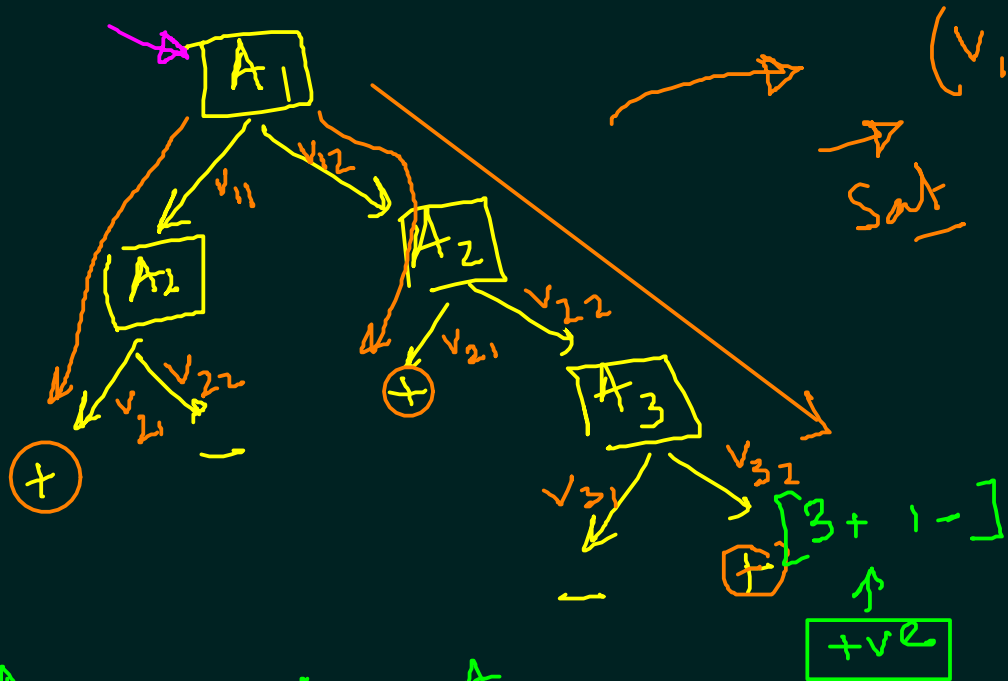
↳ Complete HS

↳ Searches Incompletely

Inductive } Preference Bias  
 Bias }



Classify



$$(v_{11} \wedge v_{21}) \vee (v_{12} \wedge v_{21})$$

$$\xrightarrow{\text{Sat}} \vee (v_{12} \wedge v_{22} \wedge v_{32})$$

Best Attribute Selection  
 ↳ DT shorter  
 ↳ Hyp. shorter

$$h = (v_{11} \wedge v_{21} \wedge v_{31})$$

Test

$$h' = (v_{11} \wedge v_{21})$$

Prefer

$$\left. \begin{array}{l} v_{11} \wedge v_{21} \wedge v_{31} \quad \checkmark \\ v_{11} \wedge v_{21} \wedge v_{32} \quad \checkmark \end{array} \right\} \text{Generalize}$$

↑ NO

✓ Training

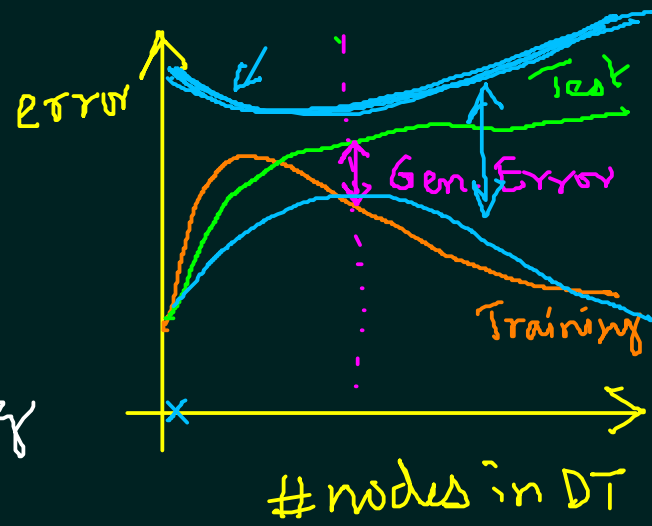
Occam's Razor

► Overfitting :  $h$

less error  
in training

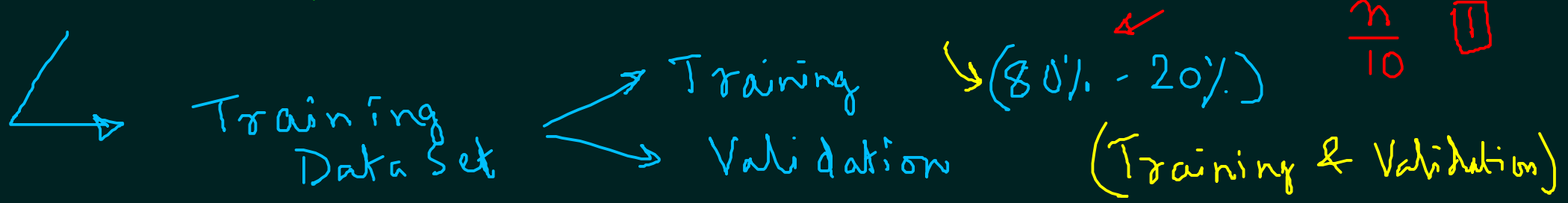
$h'$   
more error  
in training  
but less error  
in testing

Gen X | but more error  
in testing



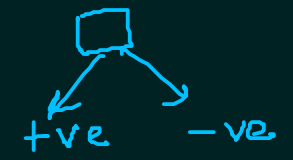
— What are the remedies?

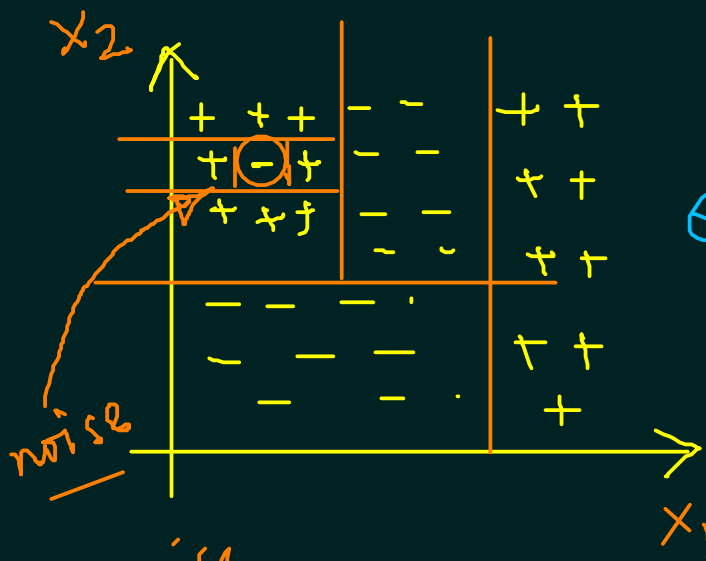
- ↳ Do not grow the DT after certain depth ✓
- ↳ Grow DT but Prune it ✓



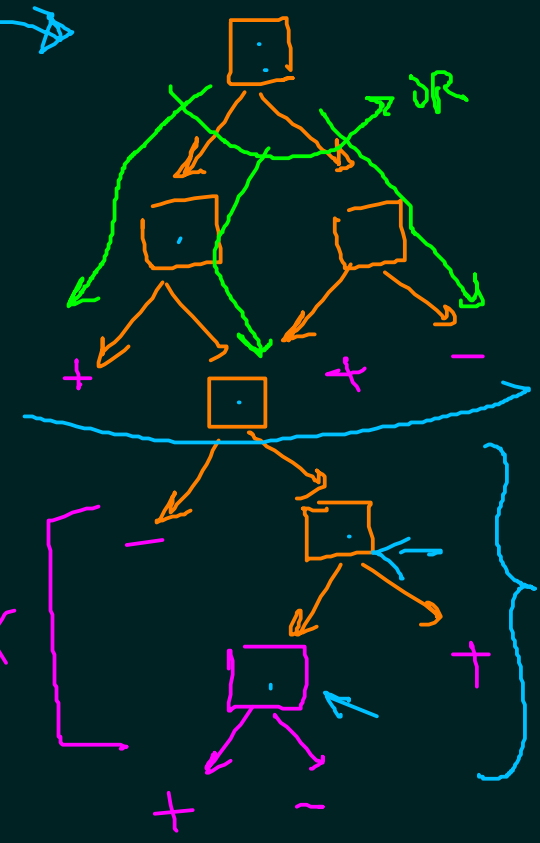
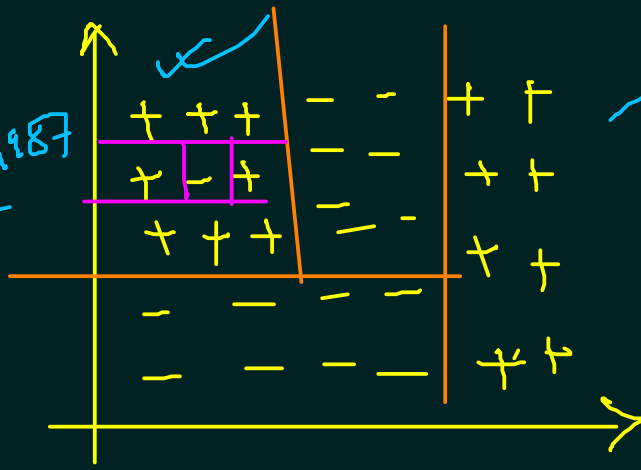
10 times → CROSS - VALIDATION

► PRUNNING → Reduced order Pruning  
→ Rule Post Pruning





$ID_3++$   
Quinlan 1987



noise  
overfit  
Reduced Order Pruning

▷ Rule Post Pruning:

$$\left( (A_1 = v_{11}) \wedge (A_2 = v_{21}) \right)$$

BF

$$V(A_1 = v_{13}) \quad V \left( (A_1 = v_{12}) \wedge (A_3 = v_{31}) \right)$$

← Pref. Bias (Play)

↳ C4.5 (Quinlan 1993)

$A_1$ :  
 $\sqrt{\text{Temp}}$ : 10 12 16 : 22 28 30 36 : 40 48  
 $\rightarrow$  Play: N N N : Y Y Y : Y : N N

$$T \geq \frac{16+22}{2} = 19$$

$$T \geq 38$$

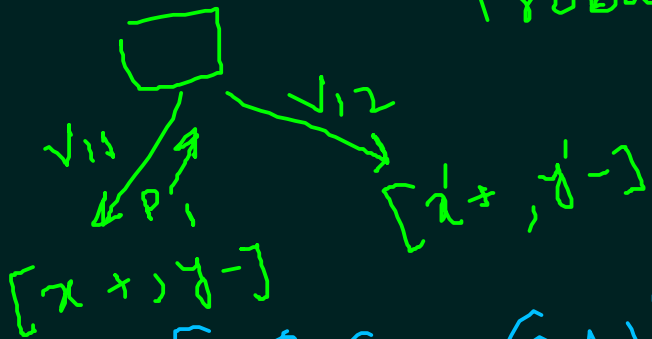
$T \geq 19$  vs.  $T \geq 38$

Majority value

Missing Attribute val

Probabilistic

$A_1$	$A_2$	$A_3$	Ans
$v_{11}$	$v_{21}$	$v_{31}$	Y
$v_{11}$ <input checked="" type="checkbox"/>	$v_{21}$	$v_{32}$	Y
$v_{12}$	$v_{22}$	$v_{32}$	N
$v_{13}$	...	...	Y
...	...	...	N



DTL

$\rightarrow$  Noise

$\rightarrow$  Missing Attr

$\rightarrow$  Continuous Value

$v_{11} \rightarrow$  Prob. Yes ( $P_1$ )

$v_{12} \rightarrow$  Prob Yes

$$\left[ \frac{I. \text{Gain}(S, A)}{ws + (A)} \right]^2$$