Lecture 5:

Using the algorithm ID3, compute (manually) the decision tree corresponding to the task of realizying simple Boolean formulas (AND, OR, XOR, ...) in n variables. Hint: Consider the truth table of the formula as your training sample.

Computing the decision tree by using ID3 for the formula: $(X_1 \vee X_2) \wedge X_3$

Truth Table:

X_1	X_2	X_3	$(X_1 \vee X_2) \wedge X_3$
T	T	T	T
T	T	F	F
T	F	T	T
T	F	F	F
F	T	T	T
F	T	F	F
F	F	T	F
F	F	F	F

Classes: T, F

Attributes: X_1, X_2, X_3

STEP 1:

$$E(S) = -P_T \log P_T - P_F \log P_F = -\frac{3}{8}\log\frac{3}{8} - \frac{5}{8}\log\frac{5}{8} = 0.95$$

$$E(S_{X_1=T}) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{2}\log\frac{1}{2} = 1$$

$$E(S_{X_1=F}) = -\frac{1}{4}\log\frac{1}{4} - \frac{3}{4}\log\frac{3}{4} = 0.81$$

$$G(S, X_1) = E(S) - \frac{|S_{X_1=T}|}{|S|} \cdot E(S_{X_1=T}) - \frac{|S_{X_1=F}|}{|S|} \cdot E(S_{X_1=F}) =$$

$$0.95 - \frac{1}{2} \cdot 1 - \frac{1}{2} \cdot 0.81 = \boxed{0.045}$$

$$E(S_{X_2=T}) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{2}\log\frac{1}{2} = 1$$

$$E(S_{X_2=F}) = -\frac{1}{4}\log\frac{1}{4} - \frac{3}{4}\log\frac{3}{4} = 0.81$$

$$G(S, X_2) = E(S) - \frac{|S_{X_2=T}|}{|S|} \cdot E(S_{X_2=T}) - \frac{|S_{X_2=F}|}{|S|} \cdot E(S_{X_2=F}) =$$

$$0.95 - \frac{1}{2} \cdot 1 - \frac{1}{2} \cdot 0.81 = \boxed{0.045}$$

$E(S_{x_3=T}) = -\frac{3}{4}\log\frac{3}{4} - \frac{1}{4}\log\frac{1}{4} = 0.81$

$E(S_{x_3=F}) = -1 \cdot \log 1 = 0$

$G(S, x_3) = E(S) - \frac{|S_{x_3=T}|}{|S|} \cdot E(S_{x_3=T}) - \frac{|S_{x_3=F})}{|S|} \cdot E(S_{x_3=F}) =$

$0.95 - \frac{1}{2} \cdot 0.81 - \frac{1}{2} \cdot 0 = \boxed{0.545}$

We choose x_3 as the root.

STEP 2:

Sub-Tree 1 $(S = S_{x_3=T})$

$E(S) = -P_T \log P_T - P_F \log P_F = -\frac{3}{4}\log\frac{3}{4} - \frac{1}{4}\log\frac{1}{4} = 0.81$

$E(S_{x_1=T}) = -1 \cdot \log(-1) = 0$

$E(S_{x_1=F}) = -\frac{1}{2} \cdot \log\frac{1}{2} - \frac{1}{2} \cdot \log\frac{1}{2} = 1$

$G(S, x_1) = E(S) - \frac{|S_{x_1=T}|}{|S|} \cdot E(S_{x_1=T}) - \frac{|S_{x_1=F})}{|S|} \cdot E(S_{x_1=F}) =$

$0.81 - \frac{1}{2} \cdot 0 - \frac{1}{2} \cdot 1 = \boxed{0.31}$

$E(S_{x_2=T}) = -1 \cdot \log(-1) = 0$

$E(S_{x_2=F}) = -\frac{1}{2} \cdot \log\frac{1}{2} - \frac{1}{2} \cdot \log\frac{1}{2} = 1$

$G(S, x_2) = E(S) - \frac{|S_{x_2=T}|}{|S|} \cdot E(S_{x_2=T}) - \frac{|S_{x_2=F})}{|S|} \cdot E(S_{x_2=F}) =$

$0.81 - \frac{1}{2} \cdot 0 - \frac{1}{2} \cdot 1 = \boxed{0.31}$

We choose x_1 (same to choose x_2) as the ROOT.

Sub-Tree 2 ($S = S_{X_3=F}$)

All the samples in S are the same class F -
return leaf node with label F.

$\boxed{\text{STEP 3:}}$

Sub-Tree 1 ($S = S_{X_3=T, X_1=T}$)

All the samples in S are the same class T -
return leaf node with label T.

Sub-Tree 2 ($S = S_{X_3=T, X_1=F}$)

We only have $X_2 \in A$, so we choose X_2.

$\boxed{\text{STEP 4:}}$

Sub-Tree 1 ($S = S_{X_3=T, X_1=F, X_2=T}$)

All the samples in S are the same class T -
return leaf node with label T.

Sub-Tree 2 ($S = S_{X_3=T, X_1=F, X_2=F}$)

All the samples in S are the same class F -
return leaf node with label F.

FINAL TREE: