- Select the attribute with the highest information gain (used in typical decision tree induction algorithm: ID3/C4.5)
- Let p_i be the probability that an arbitrary tuple in D belongs to class C_i, estimated by $|C_{i,D}|/|D|$
- Expected information (entropy) needed to classify a tuple in D:

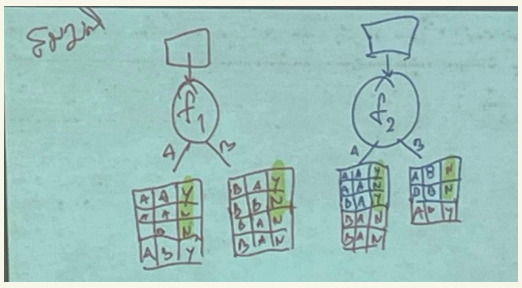
$$Info(D) = -\sum_{i=1}^{m} p_i \log_2(p_i)$$

- Information needed (after using A to split D into v partitions) to classify D:

$$Info_A(D) = \sum_{j=1}^{v} \frac{|D_j|}{|D|} \times Info(D_j)$$

- Information gained by branching on attribute A

$$Gain(A) = Info(D) - Info_A(D)$$



f_1	f_2	Target
A	A	Y
A	A	N
A	B	N
A	B	Y
B	A	Y
B	B	N
B	A	N
B	A	N

STEP 1

$$Info(Target) = I(3,5) = -\frac{3}{8}\log_2\left(\frac{3}{8}\right) - \frac{5}{8}\log_2\left(\frac{5}{8}\right) = 0.954$$

STEP 2

$$Info_{f_1}(Target) = \frac{4}{8} I(2,2) + \frac{4}{8} I(1,3)$$

$$= 0.5 + 0.379$$

$$= 0.879$$

$$Info_{f_2}(Target) = \frac{5}{8} I(2,3) + \frac{3}{8} I(1,2)$$

$$= 0.644 + 0.328$$

$$= 0.972$$

ดังนั้น $Gain(f_1) = 0.954 - 0.879 = 0.075$ ← เลือก Root node

$Gain(f_2) = 0.954 - 0.972 = -0.018$