Learning Decision Trees

What is the best model in the form of a decision tree?

Top-Down Induction of Decision Tree Algorithm on learning set L:
IF all samples from L belong to the same class C:
    Make a terminal node (i.e. a leaf) and label it C;
ELSE:
    Pick the most informative attribute A.
    Split L with respect to the values of A.
    Recursively run this algorithm on all subsets from the previous step.

How can we measure the information score of attribute A?

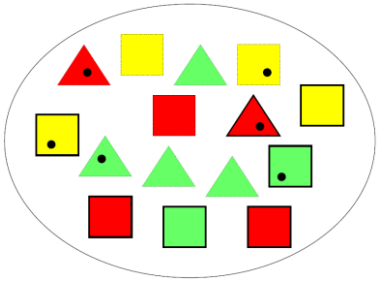
Information Gain

$IG(Y|X) = H(Y) - H(Y|X)$

$H(Y|X) = \sum_{v} p(X=v) * H(Y \mid X=v)$

$H(Y \mid X=v) = - \sum_{r} p(Y=r \mid X=v) * \log2(Y=r \mid X=v)$

Shapes



Given is a set of 15 shapes, described by 3 features: Color, Dot, and Edge.
The task is to learn the concept of a shape based on these features.
The class (or outcome) is a binary variable with 2 values: ▲, ■

		color			
shape		red	yellow	green	
	▲	2	0	4	**6**
	■	3	4	2	**9**
		5	**4**	**6**	**15**

H(shape) = -6/15 log2(6/15) -9/15 log2(9/15) = 0.971

IG(color) =H(shape) –Ires(color) = 0.971 – 0.690 = 0.281

Ires(color) = 5/15 H(red) + 4/15 H(yellow) + 6/15 H(green) =
= 1/3 * 0.971 + 4/15 *0 + 6/15 *0.918 = 0.690

H(**red**) = -2/5 log2(2/5) -3/5 log2(3/5) = 0.971
H(**yellow**) = 0 log2(0) –4/4 log2(4/4) = 0
H(**green**) = -4/6 log2(4/6) -2/6 log2(2/6) = 0.918

H(color) = -5/15 log2(5/15) -4/15 log2(4/15) -6/15 log2(6/15) = 1.565

RIG(color) = IG(color)/H(color) = 0.281/1.565 = 0.179

		Dot		
Shape		**y**	**n**	
	▲	3	3	**6**
	■	3	6	**9**
		6	**9**	**15**

H(Shape) = -6/15 log2(6/15) -9/15 log2(9/15) = 0.971

IG(Dot) =H(Shape) –Ires(Dot) = 0.971 – 0.951 = 0.02

Ires(Dot) = 6/15 H(y) + 9/15 H(n) =
= 2/5 * 1 + 3/5 *0.918 = 0.951

H(**y**) = -2* 1/2 log2(1/2) = 1
H(**n**) = -1/3 log2(1/3) –2/3 log2(2/3) = 0.918

H(Dot) = -6/15 log2(6/15) -9/15 log2(9/15) = 0.971

RIG(Dot) = IG(Dot)/H(Dot) = 0.02/0.971 = 0.021

		Edge		
Shape		**y**	**n**	
	▲	1	5	**6**
	■	6	3	**9**
		7	**8**	**15**

H(Shape) = -6/15 log2(6/15) -9/15 log2(9/15) = 0.971

IG(Edge) =H(Shape) –Ires(Edge) = 0.971 – 0.784 = 0.187

Ires(Edge) = 7/15 H(y) + 8/15 H(n) =
= 7/15 * 0.591 + 8/15 *0.954 = 0.784

H(**y**) = -1/7 log2(1/7) -6/7 log2(6/7) = 0.591
H(**n**) = -5/8 log2(5/8) -3/8 log2(3/8) = 0.954

H(Edge) = -7/15 log2(7/15) -8/15 log2(8/15) = 0.996

RIG(Edge) = IG(Edge)/H(Edge) = 0.187/0.996 = 0.188