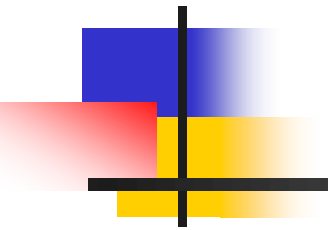


Decision Tree Advanced Topics

Dr. Goutam Chakraborty

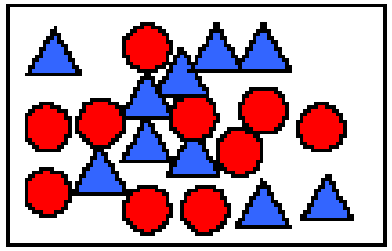
More on Decision Tree

- So far, we have used Chi-square (default) option to grow tree.
  - This maximizes the degree of separation for selecting splits on a given input using logworth
  - Other common options for evaluating goodness of a split are:
    - Gini
    - Entropy
- How many splits are possible for a nominal, ordinal or Interval variable in a Tree?

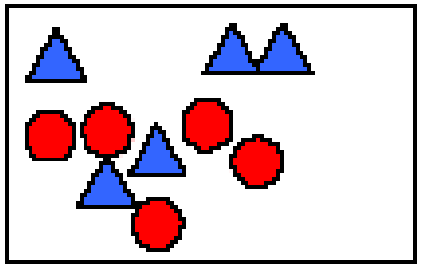
Typical Split Criteria

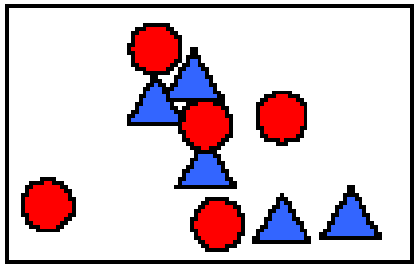
- The best split is defined as one that does the best job of separating the data into groups where a single class predominates in each group
- Measure used to evaluate a potential split is **purity**
  - The best split is one that increases purity of the sub-sets by the greatest amount
  - A good split also creates nodes of similar size or at least does not create very small nodes

Which of these splits is the best .. A,B or C?

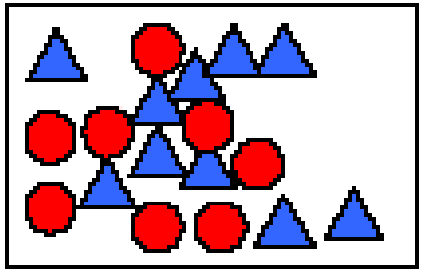


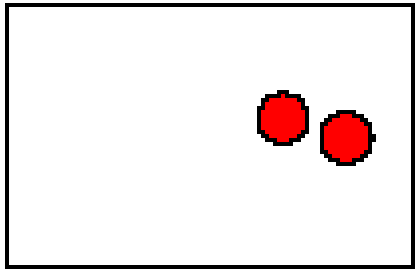
Original Data



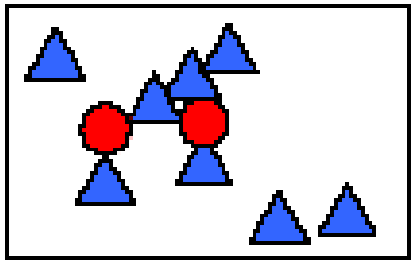


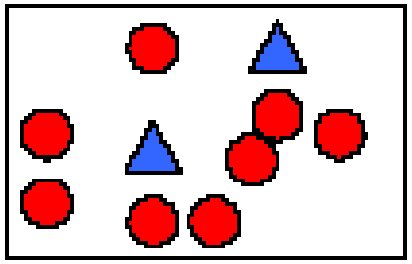
A





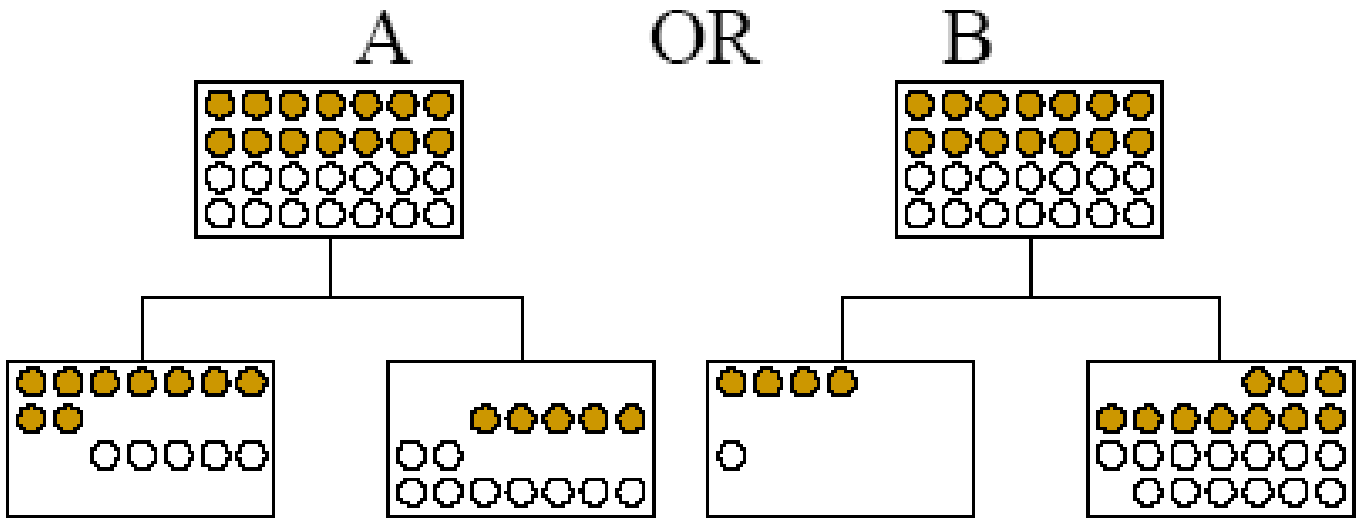
B





C

Which Is the Better Split?

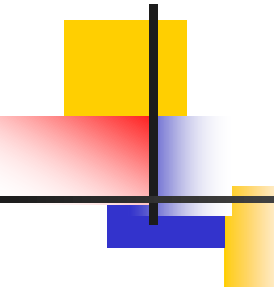


Two children, both are 64% pure.

The sizes are exactly the same.

Two children, one is 80% pure and the other 57%.

However, the sizes are quite different.



Gini Is an Easy Measure to Explain

Gini is used in the social sciences and economics. It is the probability that two things chosen at random from a population will be the same (a measure of purity).

A pure population has a Gini index of 1.

If there are two groups equally represented, then the Gini index is 0.5.

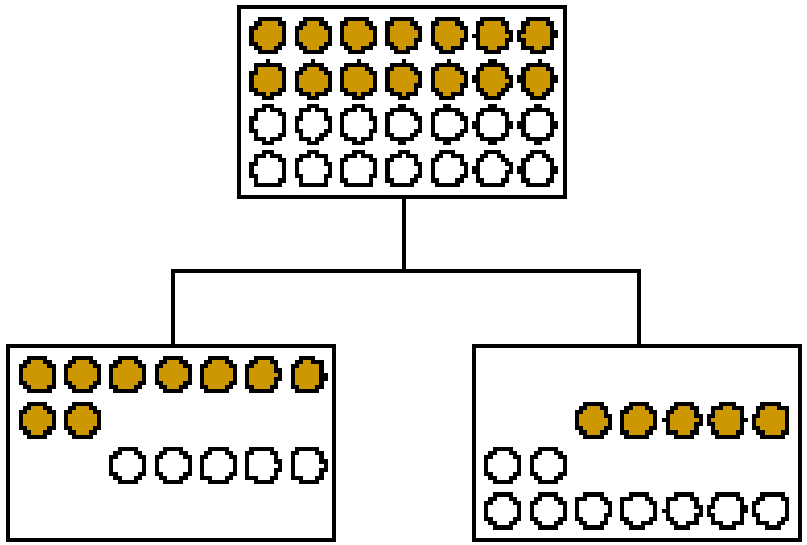
The Gini index is the sum of the square of the proportions:

$$p_1^2 + p_2^2$$

The goal is to maximize the Gini index.

Gini for Tree A

Gini score for the root node is = $0.5^2 + 0.5^2 = 0.5$

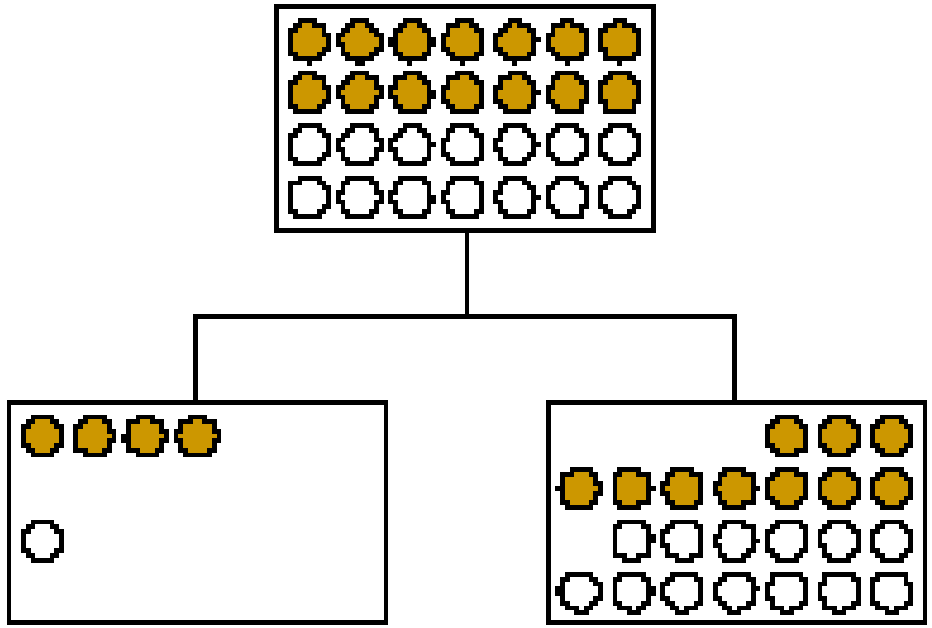


Gini score for either child is ?

$(5/14)^2 + (9/14)^2 = 0.128 + 0.413 = 0.541.$

Gini score for the split is the weighted sum of Gini scores for each child in the split (weighted by the size of the child).
So, Gini for this split is= $0.5*0.541+0.5*0.541= 0.541$

Gini for Tree B



Gini for Left Child?

$(1/5)^2 + (4/5)^2 =$

$0.04 + 0.64 = 0.68.$

Gini for Right Child?

$(10/23)^2 + (13/23)^2 =$

$0.189 + 0.319 = 0.508.$

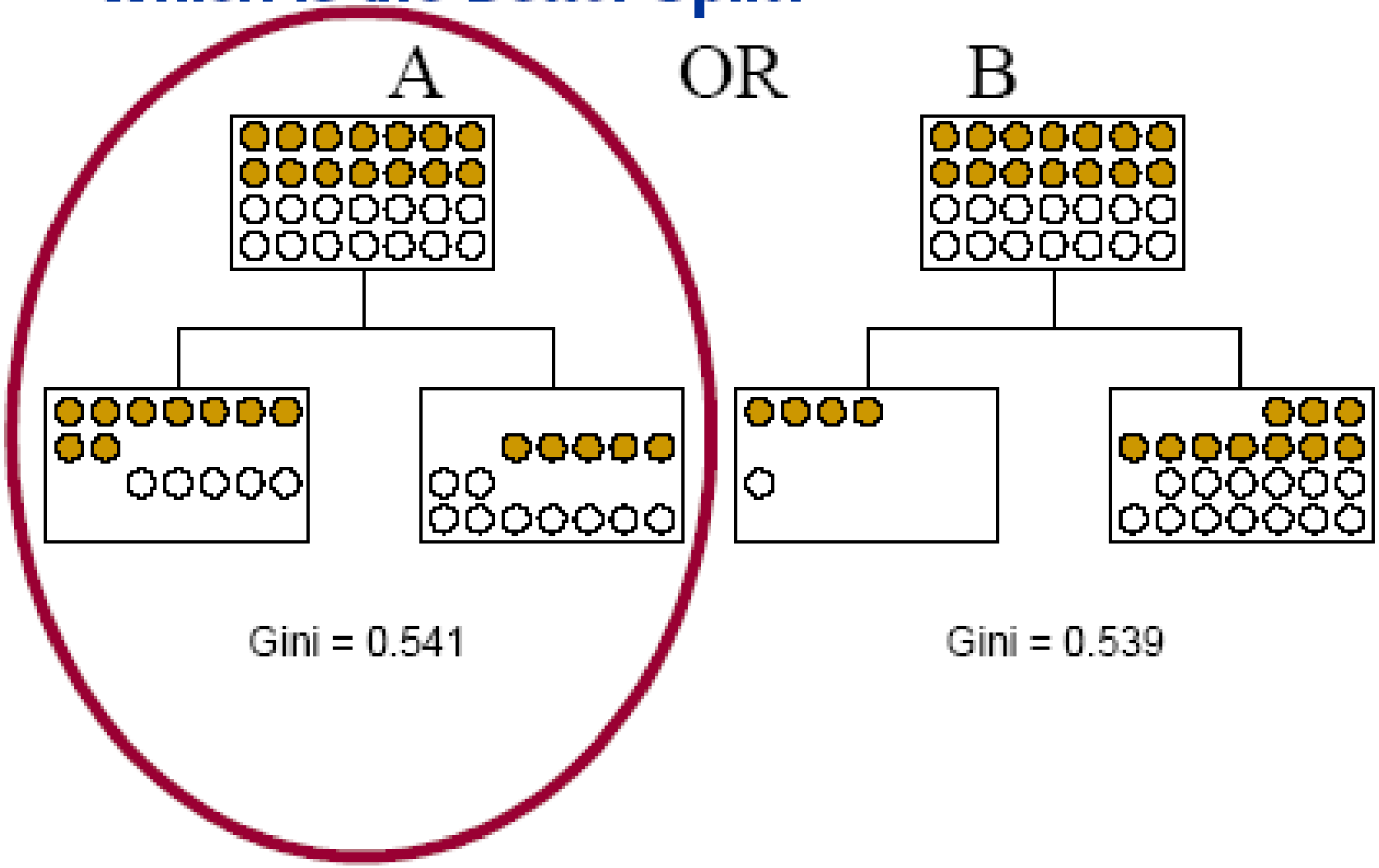
Gini for the split?

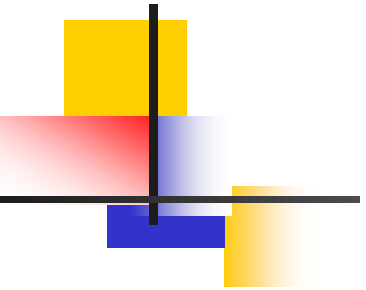
$(5/28)*Gini_{left} +$

$(23/28)*Gini_{right} =$

$0.539.$

Which Is the Better Split?





Entropy Is a Harder Measure to Explain

Entropy is used in information theory to measure the amount of information stored in a given number of bits.

A pure population has an entropy of 0.

If there are two groups equally represented, then the entropy is 1.

The calculation for entropy is

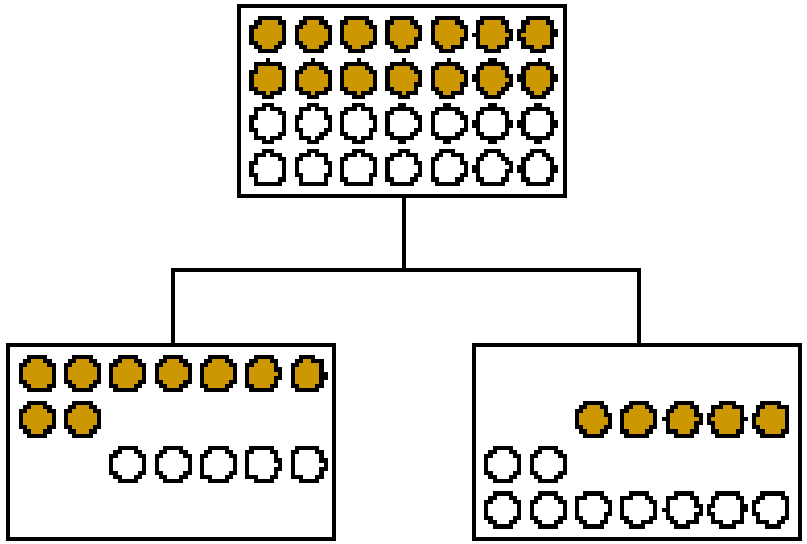
$$-1*(p_1 \log_2 (p_1) + p_2 \log_2 (p_2)).$$

(The -1 just keeps the entropy positive.)

The goal is to minimize entropy.

Entropy for Tree A

Entropy for the root node is? 1 (-(0.5*log(0.5) + 0.5*log(0.5))).



Entropy for either child is ?

-((5/14)log(5/14) + (9/14) log(9/14)) =
-(-0.5305 + -0.4098) = 0.9403.

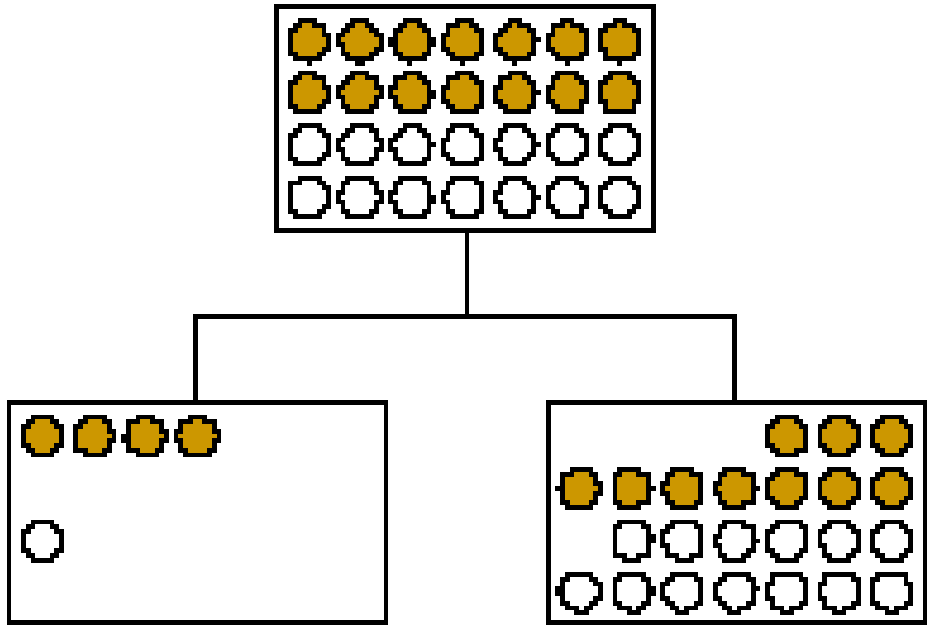
Entropy for the split is the weighted sum of entropy
for each child in the split (weighted by the size of the child).
So, entropy for this split is= 0.5*0.9403+0.5*0.9403= 0.9403

Information gain due to split = Entropy at root node – Entropy of split
= 1 –0.9403 = 0.0597

Entropy for Tree B

Entropy for Left Child?

-1*((1/5)*log(1/5) + (4/5)log(4/5)) =
-1*(-0.4644 + -0.2575) = 0.7219.



Entropy for Right Child?

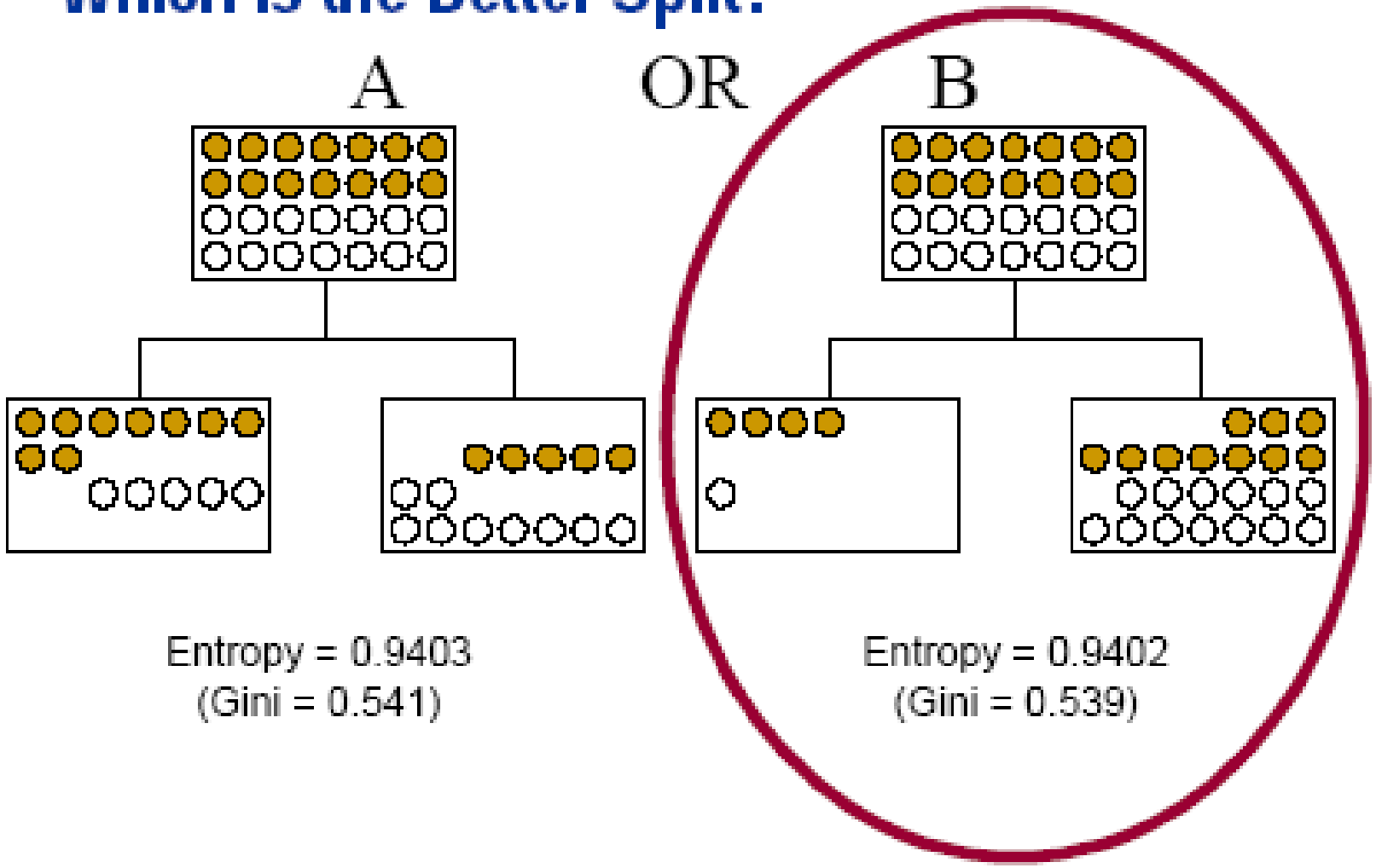
-1*((10/23)*log(10/23) +
(13/23)log(13/23)) =
-1*(-0.5225 + -0.4652) = 0.9877.

Entropy for the split?

$(5/28)*Entropy_{left}+(23/28)*Entropy_{right}$ =
0.9402.

Information gain due to split = Entropy at root node – Entropy of split
= 1 –0.9402 = 0.0598

Which Is the Better Split?



Partitioning Using An Ordinal Input With L Categories

Splits

1—234
12—34
123—4

$$\binom{3}{1} = 3$$

$$\binom{L-1}{B-1} = \frac{(L-1)!}{(B-1)!\,(L-B)!}$$

1—2—34
1—23—4
12—3—4

$$\binom{3}{2} = 3$$

$$\sum_{l=2}^{L}\binom{L-1}{l-1} = 2^{L-1} - 1$$

1—2—3—4

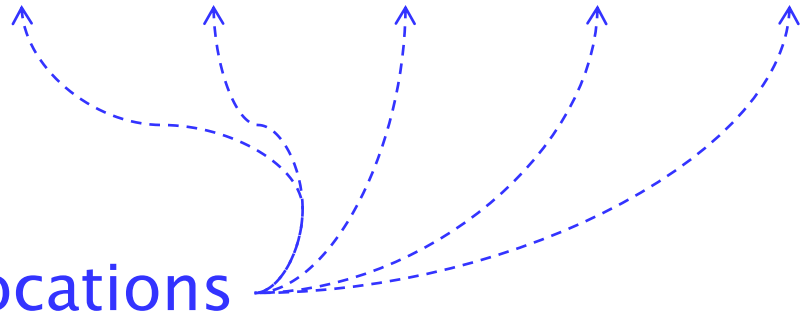
$$\binom{3}{3} = 1$$

Splits on ordinal inputs are restricted to preserve the ordering. Only adjacent values are grouped. For an ordinal input with *L* distinct levels, there are $\binom{L-1}{B-1}$ partitions into *B* branches.

There are (2^{L-1}-1) possible splits on a single ordinal input.

Interval = At Least Ordinal

X	.20	1.7	3.3	3.5	14	2515
ln(X)	-1.6	.53	1.2	1.3	2.6	7.8
rank(X)	1	2	3	4	5	6



potential split locations

For interval or ordinal inputs, splits in a decision tree depend only on the ordering of the levels, making tree models robust to outliers in input space. The application of a rank or any monotonic transformation to an interval variable will not change the fitted tree. ¶

Partitioning on a Nominal Input

Trees treat splits on inputs with nominal and ordinal measurement scales differently. Splits on a nominal input are not restricted. For a nominal input with L distinct levels, there are $S(L,B)$ partitions into B branches, where $S(L,B)$ is a *Stirling number of the second kind*. The total number of partitions is one less than the *Bell number* for L levels, $B_L = \sum_{i=0}^{L} S(L,i)$

1—234
2—134
3—124
4—123
12—34
13—24
14—23

1—2—34
1—3—24
1—4—23
2—3—14
2—4—13
3—4—12

1—2—3—4

$$S(L,B) = B \cdot S(L-1,B) + S(L-1,B-1)$$

L \ B:	2	3	4	total
2	1			1
3	3	1		4
4	7	6	1	14
5	15	25	10	51
6	31	90	65	202
7	63	301	350	876
8	127	966	1701	4139
9	255	3025	7770	21146