

# Artificial Intelligence Fundamentals

Learning: Identification Trees,  
Disorder

# Egg fight problem

Champion ?	Multiple layers of painting ?	Big ?	Origin ?	Top?
No	?	No	Chicken	Round
No	Yes	Yes	Duck	Pointed
Yes	?	No	Guinea hen	Round
Yes	Yes	No	Guinea hen	Pointed
Yes	No	No	Chicken	Pointed
No	No	Yes	Duck	Flat
No	?	No	Chicken	Flat
No	?	Yes	Chicken	Flat

Why we cannot use a nearest neighbor algorithm?

# Data set

- Non numeric
- Some characteristics don't matter
- Some characteristics do matter, but only part of the time
- Cost – some of the tests can be more expensive than others

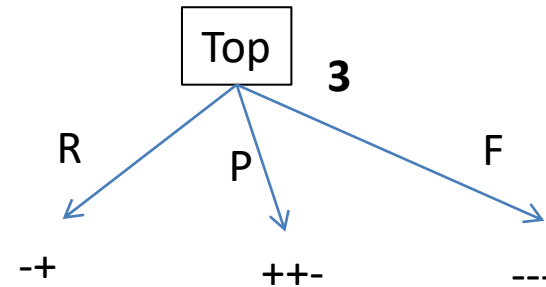
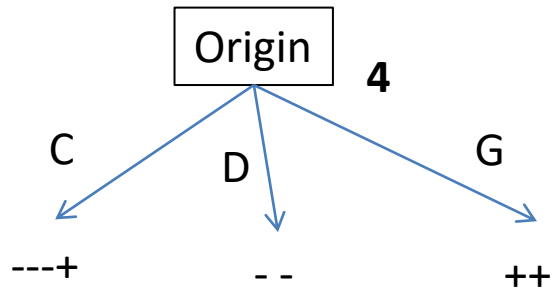
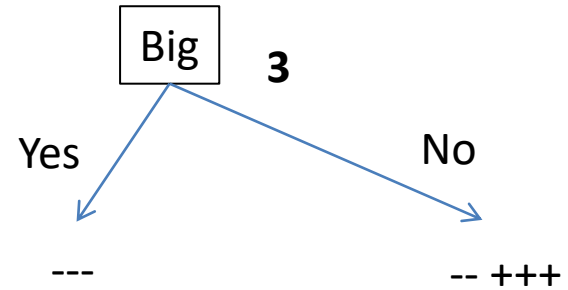
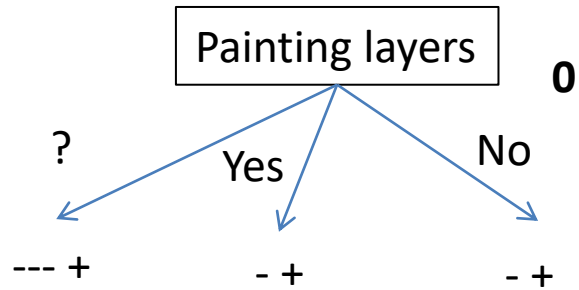
We need a method that enables computers to learn by assembling tests into an *identification tree*.

How we build that tree ? What will be a good characteristic of that tree?

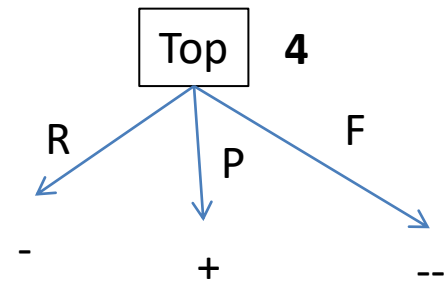
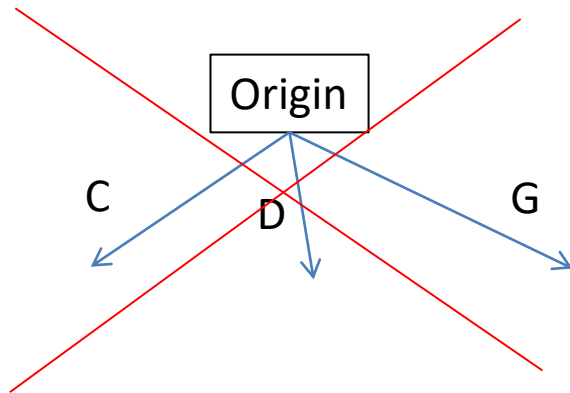
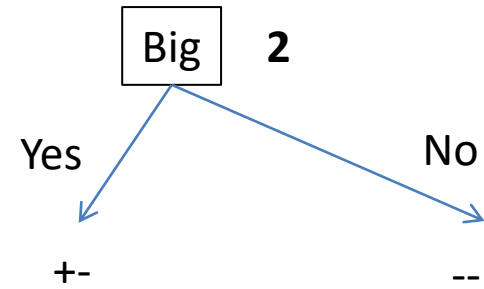
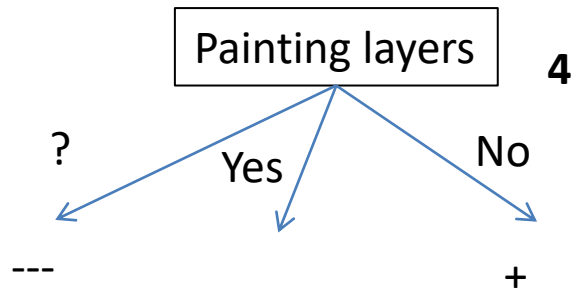
# Occam's razor

- The word is inherently simple. Therefore the smallest identification tree that is consistent with the samples is the one that is most likely to identify unknown objects correctly.
- How we can construct the smallest identification tree?

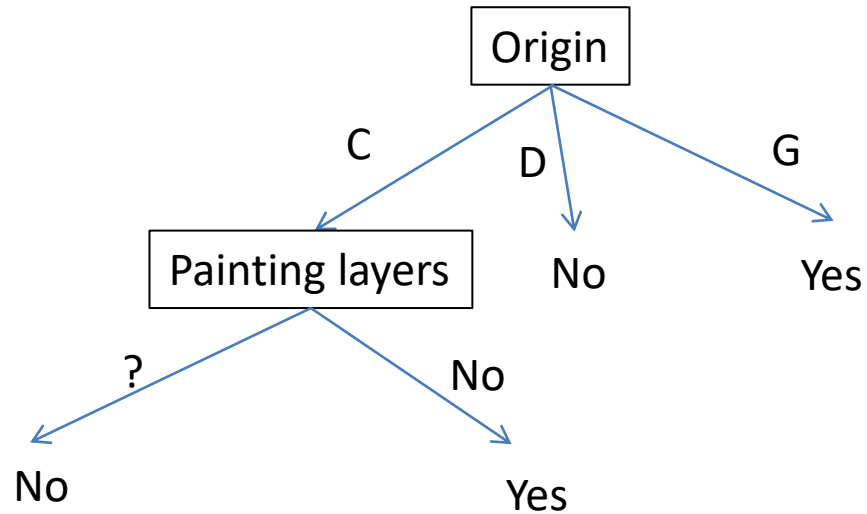
# Tests



Champion ?	Multiple layers of painting ?	Big ?	Top?
No	?	No	Round
Yes	No	No	Pointed
No	?	No	Flat
No	?	Yes	Flat



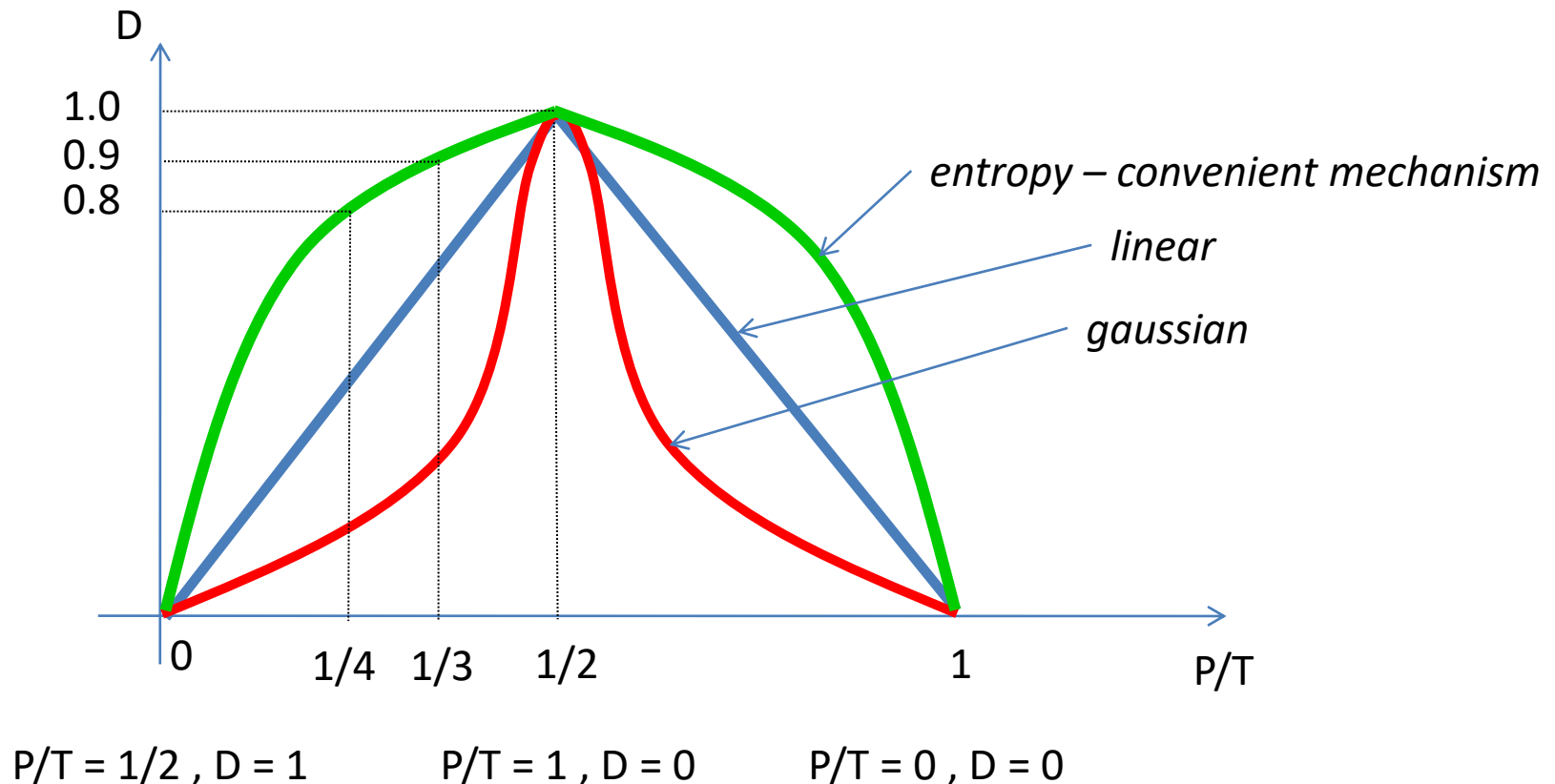
# Final identification tree





# Measuring the disorder - entropy

$$D(\text{set}) = -\frac{P}{T} \log_2 \frac{P}{T} - \frac{N}{T} \log_2 \frac{N}{T}$$



# Measuring the error of the test

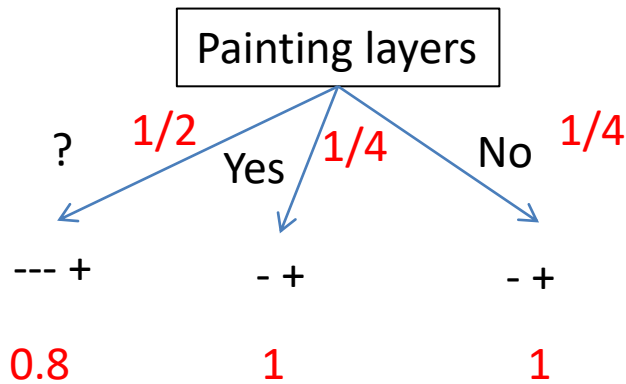
$$E(\text{Test}) = \sum_{\substack{\text{SETS} \\ \text{PRODUCED}}} D(\text{set}) * \frac{\# \text{ of samples in set}}{\# \text{ of samples handled by test}}$$

$$E(\text{Test Painting}) = 0.8 * \frac{1}{2} + 1 * \frac{1}{4} + 1 * \frac{1}{4} = 0.9$$

$$E(\text{Test Big}) = 0 * \frac{3}{8} + 0.9 * \frac{5}{8} \approx 0.56$$

$$E(\text{Test Origin}) = 0.8 * \frac{1}{2} + 0 * \frac{1}{4} + 0 * \frac{1}{4} = 0.4$$

$$E(\text{Test Top}) = 1 * \frac{1}{4} + 0.9 * \frac{3}{8} + 0 * \frac{3}{8} = 0.5875$$



# Numeric data

