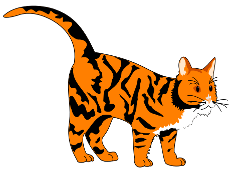
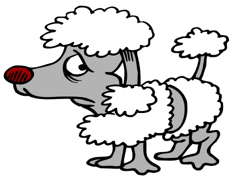
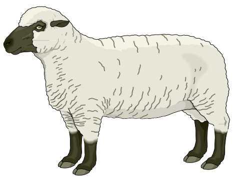
Session 1: Concept Learning and Decision Trees

Concept learning = infer concept from given training examples
E.g., learn concept "dog"

Instance = tuple that assigns values to attributes
Training example = labeled instance

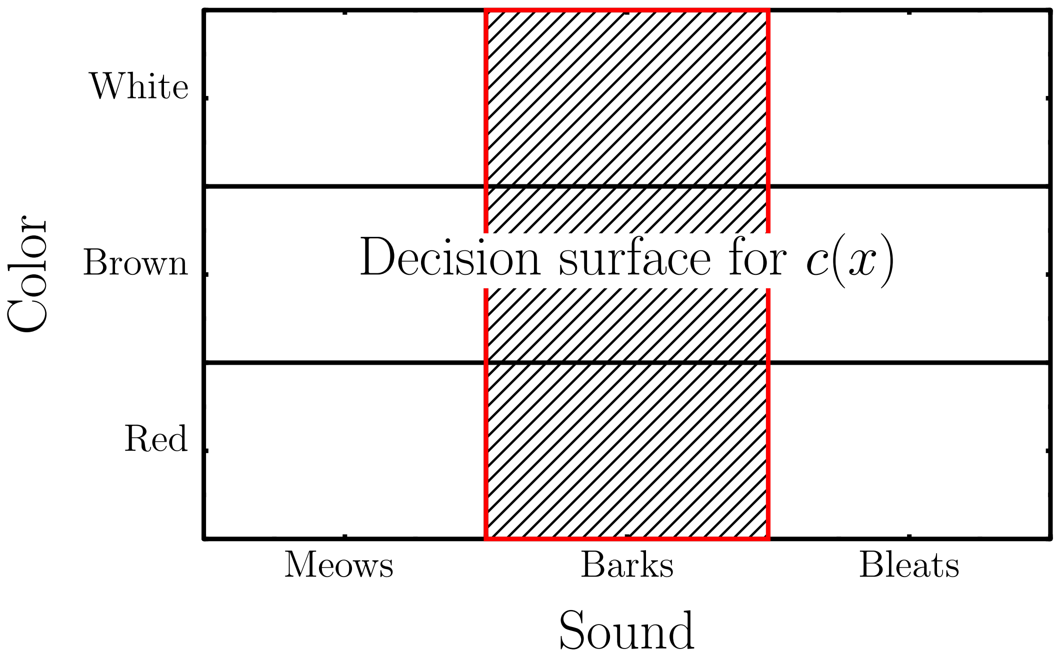
	Sound	Color	Class label
	Barks	Brown	T (true)
	Meows	Red	F (false)
	Barks	White	T (true)
	Bleats	White	F (false)

→ Something is a dog if it barks

Mathematically:

- Given instance space X (e.g., Sound $\times$ Color)
- A concept = boolean function $c(x) : X \to \{T, F\}$
- E.g., $c(x) = \begin{cases} T & \text{if Sound}(x) = \text{Barks} \\ F & \text{otherwise} \end{cases}$

Decision surface for $c(x)$ = boundary of region in instance space for which $c(x) = T$



Generality

- $g(x)$ more general or equal to $s(x)$: $g(x) \geq_g s(x)$
- $\leftrightarrow \forall x \in X\colon\ s(x) = T \rightarrow g(x) = T$
- $\leftrightarrow$ decision surface $s(x)$ is "inside" of that of $g(x)$

Learning Concepts

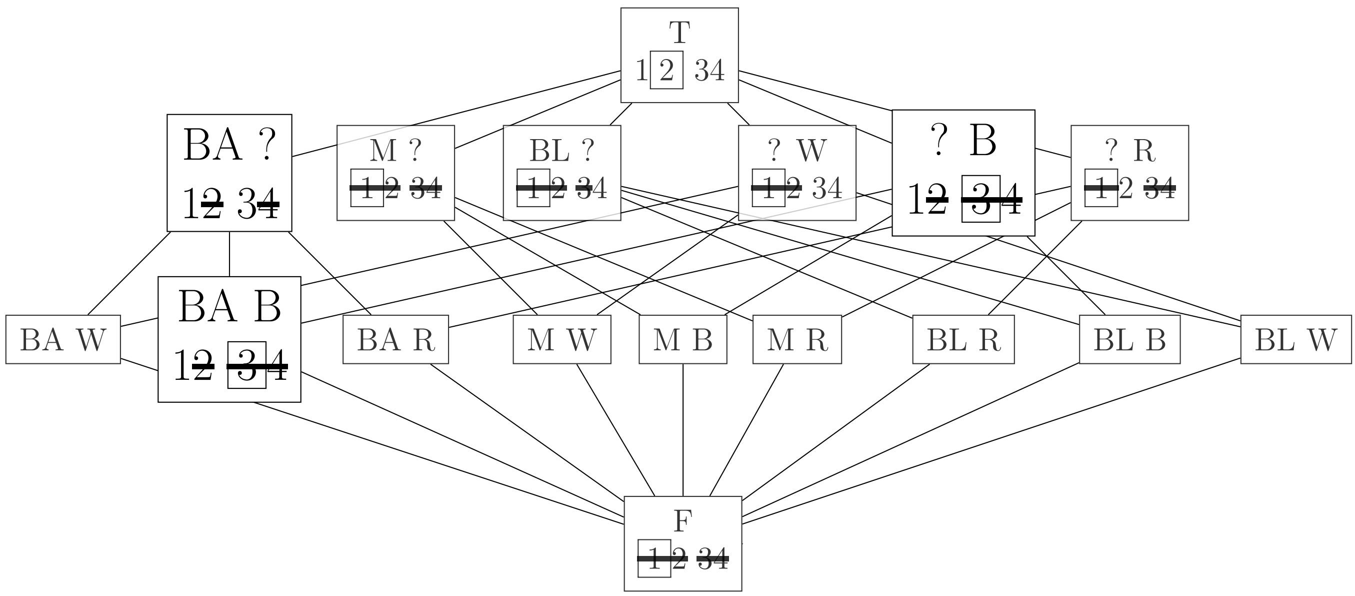
- Version Spaces
- Decision Trees

Version Space?

- Given a hypothesis space H, training data D
- Consistent$(h, D) = \forall x, c(x) \in D : h(x) = c(x)$
- $\mathrm{VS}_{H,D} = \{h \in H \mid \text{Consistent}(h, D)\}$
- Usually defined by:
  - G is maximally general members of $\mathrm{VS}_{H,D}$
  - S is maximally specific members of $\mathrm{VS}_{H,D}$
- $\mathrm{VS}_{H,D} = \{h \in H \mid \exists s \in S, \exists g \in G : g \geq_g h \geq_g s\}$

  More specific than maximally general hypotheses G and more general than maximally specific hypotheses S

VS after seeing (1) $\{Barks, Brown\} \rightarrow T$ and (2) $\{Meows, Red\} \rightarrow F$ before seeing (3) $\{Barks, White\} \rightarrow T$ and (4) $\{Bleats, White\} \rightarrow T$

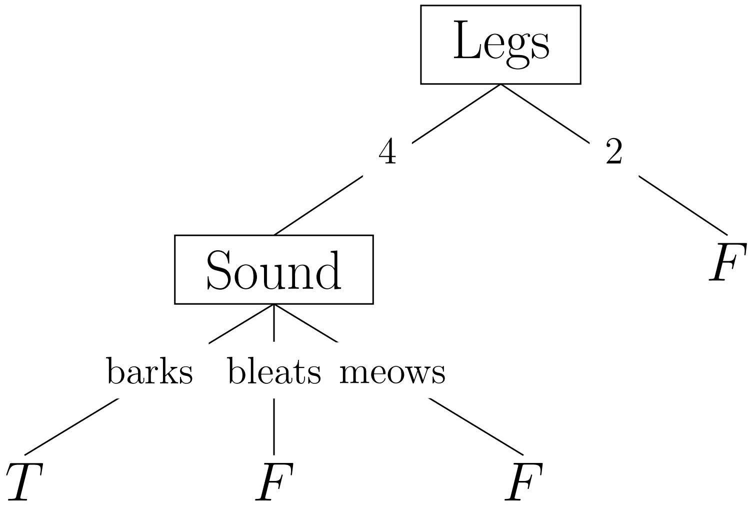


Classification: Use voting among concepts in $\mathrm{VS}_{H,D}$

- Unanimously $T \rightarrow T$, unanimously $F \rightarrow F$
- Otherwise $\rightarrow$ "Don't know"

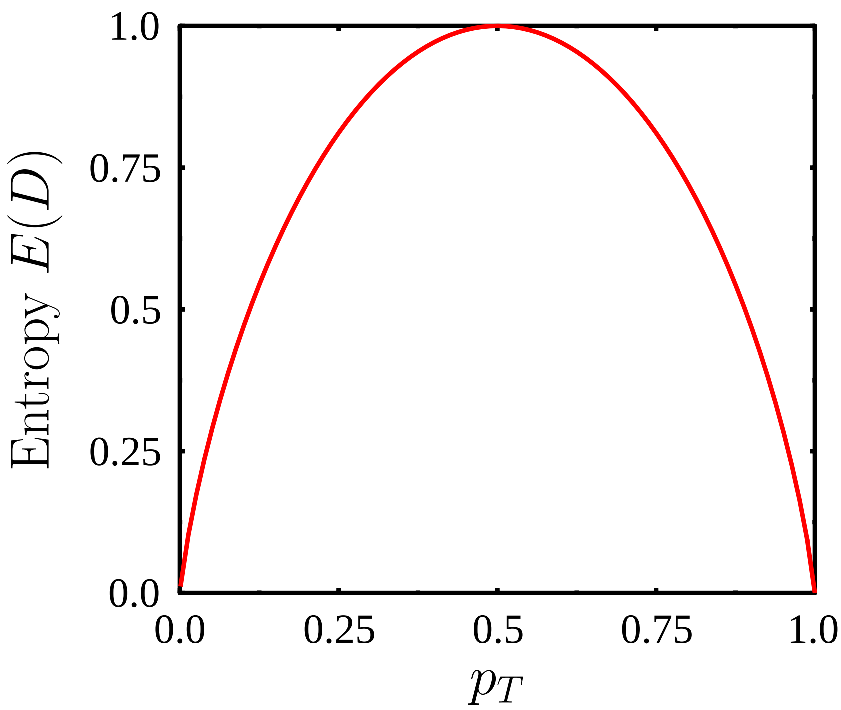
  $\{Barks, White\}$: $\boxed{\text{BA ?}} \rightarrow T$, $\boxed{\text{? B}} \rightarrow F$, $\boxed{\text{BA B}} \rightarrow F$

Decision Trees



Selecting Split Attributes

- Entropy $E(D) = \sum_{i=1}^{c} -p_i \cdot \log_2 p_i$
- $= -p_T \cdot \log_2 p_T - (1 - p_T) \cdot \log_2(1 - p_T)$



- Information gain $G(D, a) = E(D) - \sum_{v \in \text{values}(a)} \frac{|D_v|}{|D|} \cdot E(D_v)$
- Select at each node the attribute with the highest gain