Concept Learning

**Concept Learning**: Inferring a boolean-valued function from training examples of its input and output.

- General-to-specific ordering of hypotheses
- Version spaces and candidate elimination algorithm
- Need for inductive bias
# A Concept Learning Task

- **EnjoySport concept (Table 2.1)**

<table>
<thead>
<tr>
<th>Example</th>
<th>Sky</th>
<th>Temp</th>
<th>Humid</th>
<th>Wind</th>
<th>Water</th>
<th>Forecst</th>
<th>EnjoySpt</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sunny</td>
<td>Warm</td>
<td>Normal</td>
<td>Strong</td>
<td>Warm</td>
<td>Same</td>
<td>Yes</td>
</tr>
<tr>
<td>2</td>
<td>Sunny</td>
<td>Warm</td>
<td>High</td>
<td>Strong</td>
<td>Warm</td>
<td>Same</td>
<td>Yes</td>
</tr>
<tr>
<td>3</td>
<td>Rainy</td>
<td>Cold</td>
<td>High</td>
<td>Strong</td>
<td>Warm</td>
<td>Change</td>
<td>No</td>
</tr>
<tr>
<td>4</td>
<td>Sunny</td>
<td>Warm</td>
<td>High</td>
<td>Strong</td>
<td>Cool</td>
<td>Change</td>
<td>Yes</td>
</tr>
</tbody>
</table>

What is the general concept?
Learning Task

- **Given:**
  - Instances $X$: Possible days, each described by the attributes $Sky$, $AirTemp$, $Humidity$, $Wind$, $Water$, $Forecast$
  - Target function $c$: $EnjoySport : X \rightarrow \{0, 1\}$
  - Hypotheses $H$: Conjunctions of literals. E.g.
    $$\langle ?, Cold, High, ?, ?, ? \rangle.$$
  - Training examples $D$: Positive and negative examples of the target function
    $$\langle x_1, c(x_1) \rangle, \ldots, \langle x_m, c(x_m) \rangle$$

- **Determine:** A hypothesis $h$ in $H$ such that $h(x) = c(x)$ for all $x$ in $D$.

  - Most general hypothesis $\langle ?, ?, ?, ?, ?, ? \rangle$
  - Most specific hypothesis $\langle \emptyset, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset \rangle$
Terminology

- Instances or instance space $X$
  - Set of all possible input items
  - E.g., $x = \langle\text{sunny, warm, normal, strong, warm, same}\rangle$
    * $(3 \cdot 2 \cdot 2 \cdot 2 \cdot 2) = 96$ instances

- Target concept $c : X \rightarrow \{0, 1\}$
  - Concept or function to be learned
  - E.g., $c(x) = 1$ if EnjoySport=yes, $c(x) = 0$ if EnjoySport=no

- Training examples $D = \{\langle x, c(x)\rangle\}$
  - Instance $x$ from $X$ accompanied by target concept value $c(x)$
  - Positive examples, $c(x) = 1$, members of target concept
  - Negative examples, $c(x) = 0$, non-members of target concept
Terminology

- Hypothesis space $H$, set of all possible hypotheses
  - Depends on choice of representation
  - E.g., conjunctive concepts for EnjoySport
    * $(5 \cdot 4 \cdot 4 \cdot 4 \cdot 4) = 5120$ syntactically distinct hypotheses
    * $(4 \cdot 3 \cdot 3 \cdot 3 \cdot 3) + 1 = 973$ semantically distinct hypotheses
    * Any hypothesis with $\emptyset$ classifies all examples negative
  - Want $h$ in $H$ such that $h(x) = c(x)$ for all $x$ in $X$

- Inductive learning hypothesis
  - Any hypothesis approximating the target concept well over a sufficiently large set of training examples will also approximate the target concept well for unobserved examples
Concept Learning as Search

- Learning viewed as a search through hypothesis space $H$ for hypothesis consistent with training examples
- General-to-specific ordering of hypotheses
  - Allows more directed search of $H$
  - Definition [fig2.1]
    * Hypothesis $h_1$ is more general than or equal to hypothesis $h_2$ iff $\forall x \in X, h_1(x) = 1 \leftrightarrow h_2(x) = 1$
    * Written $h_1 \geq_g h_2$
    * $h_1$ (strictly) more general than $h_2$ ($h_1 >_g h_2$) when $h_1 \geq_g h_2$ and $h_2 \ngeq_g h_1$
      - Also implies $h_2 \leq_g h_1$, $h_2$ more specific than $h_1$
  - Defines partial order over $H$
Finding Maximally-Specific Hypothesis: Find-S

- Find the most specific hypothesis covering all positive examples
- Hypothesis \( h \) covers positive example \( x \) if \( h(x) = 1 \)
- Find-S algorithm [tab2.3]
  - Example from data in [tab2.1]
  - Illustrate in [fig2.2]
- Will \( h \) ever cover a negative example?
  - No, if \( c \) in \( H \) and training examples consistent
- Problems with Find-S
  - Cannot tell if converged on target concept
  - Why prefer the most specific hypothesis?
  - Handling inconsistent training examples due to errors or noise
  - What if more than one maximally-specific consistent hypothesis?
Version Spaces and the Candidate-Elimination Algorithm

- CE finds all hypotheses consistent with training examples
  - Hypothesis $h$ consistent with training $D$ iff $h(x) = c(x)$ for all $(x, c(x)) \in D$
    * Differs from "satisfies" ($h(x) = 1$) in that $h(x)$ may be 0
  - Version Space $V_{S_{H,D}} = \{h \in H \mid \text{consistent}(h, D)\}$

- List-Then-Eliminate algorithm [tab2.4]
  - $V_S$ = list of every hypothesis in $H$
  - For each training example $(x, c(x))$
    * Remove from $V_S$ any $h$ where $h(x) \neq c(x)$
  - Return $V_S$
  - Impractical for all but most trivial $H$’s
Version Spaces and the Candidate-Elimination Algorithm

- Can represent $VS$ with only most general $G$ and most specific $S$ members
  - $VS$ after four EnjoySport training examples [fig2.3]
- $VS$ representation theorem (2.1)
  - All consistent hypotheses fall between $S$ and $G$ according to the ”more general than” partial ordering
- CE algorithm [tab2.5]
  - EnjoySport example [fig2.4-2.7]
  - Final $VS$ independent of training sequence
Remarks on VS’s and CE

- Will CE converge to correct hypothesis?
  - If no errors and target concept in $H$
  - Convergence: $S = G = \{h_{final}\}$
  - Otherwise, eventually $S = G = \{\}$

- Which training example requested next?
  - Learner may query oracle for example’s classification
  - Ideally, choose example eliminating half of $VS$
    * Need $\log_2 |VS|$ examples to converge
  - E.g., \langle sunny,warm,normal,light,warm,same\rangle or \langle sunny,warm,high,light,cool,change\rangle
Remarks on VS’s and CE

- Using partially learned concepts
  - If all of $S$ predict positive, then positive [tab2.6,A]
  - If all of $G$ predict negative, then negative [tab2.6,B]
  - If half and half, then don’t know [tab2.6,C]
  - If majority of $h$’s say pos (neg), then pos (neg) with some confidence [tab2.6,D]

- $G$ can grow exponentially in $|D|$, even for conjunctive $H$
Inductive Bias

• How does the choice for $H$ affect learning performance?

• Biased hypothesis space
  – EnjoySport $H$ cannot learn constraint [sky = sunny or cloudy]
  – How about $H = $ every possible hypothesis?

• Unbiased learner
  – $H = $ every teachable concept (power set of $X$)
    * E.g., EnjoySport $|H| = 2^{96} = 10^{28}$ (only 973 by previous $H$, biased!)
  – $H' = $ arbitrary conjunctions, disjunctions or negations of hypotheses from $H$
    * E.g., [sky=sunny or cloudy] →
      \( \langle \text{sunny},?,?,?,? \rangle \) or \( \langle \text{cloudy},?,?,?,? \rangle \)
  – Problem using CE with $H'$
    * $S = $ disjunction of positive examples
    * $G = $ negated disjunction of negative examples
    * Thus, no generalization
    * Each unseen instance covered by exactly half of $VS$
Futility of bias-free learning

- Fundamental property of inductive learning
  - Learners making no a priori assumptions about target concept have no rational basis for classifying unseen instances

- Inductive bias
  - Informal: any preference on the space of all possible hypotheses other than consistency with training examples
  - Formal: set of assumptions $B$ such that the classification of an unseen instance $x_i$ by a learner $L$ on training data $D$ can be inferred deductively [fig2.8]
  - E.g., for CE, $B = \{(c \in H)\}$
    * Where classification only by unanimous decision of $VS$

- Inductive bias permits comparison of learners
  - Rote learner: no bias
  - CE: $c \in H$
  - Find-S: $(c \in H)$ and $(c(x) = 0$ for all instances not covered)
Summary

- Concept learning as search
- General-to-specific ordering
- Version spaces and CE algorithm
- $S$ and $G$ boundaries characterize learner’s uncertainty
- Learner can generate useful queries
- Inductive leaps possible only if learner is biased