Concept Learning

Inferring a boolean-valued function from training examples of its input and output is known as concept learning.

A Concept Learning Task

A set of items over which a concept is defined is called a set of instances, denoted by X.

The concept or function to be learned is denoted by c. In general, c can be any boolean-valued function defined over instances X such that c : X -> {0, 1}.

The task is said to be defined as

Given instances X, hypotheses H, target concept c and training examples D, find a hypothesis h in H such that h(x) = c(x) for all x in X.

Instances for which c(x) is 1 are called positive and c(x) = 0 are called negative.

The Inductive Learning Hypothesis is any hypothesis found to approximate the target function well over a sufficiently large set of training examples, and also approximates the target function well over other unobserved examples.

Concept Learning as a Search

Concept learning can be viewed as the task of searching through a large space of hypotheses implicitly defined by the hypothesis representation, to find the best fit hypothesis.

Many algorithms for concept learning organize the search through the hypothesis space by relying on a general-to-specific ordering of hypothesis, reducing the number of enumerated hypotheses.

For h1 and h2 being boolean-valued functions over X, h1 is more general than or equal to h2 if and only if for all examples x in X, if h2(x) = 1, then h1(x) is also 1.

This models the concept learning algorithm as

1. Initialise h to the most specific hypothesis in H
2. For each positive training instance x
    a. For each attribute constraint a in h
        i. If the constraint a is satisfied in h, do nothing
        ii. Else, replace a in h with the next more general constraint that is satisfied by x.
3. Output hypothesis h.

This is also called the find-S algorithm.