Title

9.66 Final Project

Leon Lin

abstract

Human minds have a striking ability to learn general concepts from just a small number of examples. Two or perhaps just one example of, say, a zebra or a wheel suffices to teach a child those respective categories.

In this project we model the problem much as Goodman, Tenenbaum, Feldman, & Griffiths (2008) do: the *world* consists of a set of *objects*, each of which may exhibit some subset of the *features* in the set of possible features. The learner is aware of exactly which features each object exhibits.

Following a suggestion of Goodman et al. (2008), we attempt to extend their work by employing "strong sampling": A subset C of the objects belong to a *concept*, which the learner seeks to learn. A few objects are selected randomly from C, and perhaps a few are also selected from the complement of C. These few objects and an indication of whether each belongs to C are given to the learner. The learner receives incorrect information about whether a particular object is in C with probability e^{-b}, for some *outlier parameter* b, as in Goodman et al. (2008).

Given a few positive examples of a concept, there are potentially many possible larger sets C that contain those examples. How is the learner to choose from among them? Feldman (2000) found that the most natural concepts are the simple ones, that is, those with short representations as boolean propositional formulas in terms of feature predicates f_i, where $f_i(x)$ is true if and only if object x exhibits feature i. Such a formula might look like

$$(f_1(x) \wedge f_3(x)) \vee f_2(x).$$

Again following Goodman et al. (2008), we will consider formulas in *disjunctive normal form*. These formulas are disjunctions of conjunctions of terms of the form $f_i(x)$ or $\neg f_i(x)$. A grammar for such formulas is given by

$$\begin{aligned}
S &\to x \in C \Leftrightarrow (\text{Disj})\\
\text{Disj} &\to (\text{Conj}) \vee \text{Disj}\\
\text{Disj} &\to \text{False}\\
\text{Conj} &\to P \wedge \text{Conj}\\
\text{Conj} &\to \text{True}\\
P &\to f_1(x)\\
P &\to \neg f_1(x)\\
P &\to f_2(x)\\
P &\to \neg f_2(x)\\
&\vdots
\end{aligned}$$

With the standard semantics, a formula generated by this grammar picks out a specific set of objects — a *hypothesis* for the concept.

Any partition of the objects into concepts and non-concepts is specified by some formula, so long as objects with the same features are categorized in the same way. Two formulas may define the same concept.

Results are presented in Table 1.

Table 1
Sample table.

AAA	BBB	CCC
1.0	2.0	3.0
1.0	2.0	3.0

References

Feldman, J. (2000). Minimization of boolean complexity in human concept learning. *Nature*, *407*(6804), 630–633.

Goodman, N. D., Tenenbaum, J. B., Feldman, J., & Griffiths, T. L. (2008). A rational analysis of rule-based concept learning. *Cognitive Science*, *32*(1), 108–154.