Classification is one of the most important issues in machine learning. In the $n$-dimensional binary space $\{0, 1\}^n$, we are given a set of examples $x \in \{0, 1\}^n$, where each example $x$ is labeled as $y(x)$ by a Boolean function $y: \{0, 1\}^n \rightarrow \{0, 1\}$, called an oracle, and the classification problem asks to find a classifier $c$, a Boolean function $c: \{0, 1\}^n \rightarrow \{0, 1\}$ that is (approximately) identical to $y$. We aim at representing $y$ as a compact concept in the spirit of Occam’s Razor. As a tool to describe classifiers, we assume a representation model $R$, on which classifiers can be implemented and the complexity of a representation is defined as its length of description. From our assumption on oracles, we wish to construct a classifier $c$ with small complexity. In this paper, we discuss two representation models, iteratively composed features and decision trees, and prove that, for any classifier $c$, the former has a representation which is at least as compact as that by the latter.

Keywords: decision trees, iterative composition of features, learning algorithms, machine learning, representation complexity