Optimal binary decision tree

Mohammad Mahdi Heydari

Amirkabir University of Technology

June 1, 2019

Overview

Introduction

- An optimal binary decision tree is one which minimizes the expected number of tests required to identify the unknown object.
- Large amount of effort had been put into finding efficient algorithms for constructing optimal binary decision tree.
- On supposition that $P \neq NP$, such algorithm does not exist.
- It supplies motivation for finding efficient heuristics for constructing near-optimal decision trees.

Definition

Sets

Let $X = \{x_1, ..., x_n\}$ be a finite set of objects and let $\tau = \{T_1, ..., T_t\}$ be a finite set of tests. for each test T_i, $1 \leq i \leq t$ and object x_j, $1 \leq j \leq n$, we either have $T_i(x_j) = \mathit{true}$ or $T_i(x_j) = \mathit{false}$. T_i also denotes the set $\{x \in X | T_i(x) = \mathit{true}\}$.

Problem

The problem is to construct an identification procedure for the objects in X such that the expected number of tests required to completely identify an element of X is minimal. At each non-terminal node, a test is specified and terminal nodes specify objects in X.

Definition

Procedure

To apply the identification procedure one first applies the test specified at the root to the unknown object; if it is false one takes the left branch, otherwise the right. This procedure is repeated at the root of each successive subtree until one reaches a terminal node which names the unknown object.

Cost

Let $p(x_i)$ be the length of the path from the root to the terminal node naming x_i, that is, the number of tests required to identify x_i. Then the cost of this tree will be:

$$\sum_{x_i \in X} p(x_i)$$

Any Question?