CS420+500: Advanced Algorithm Design and Analysis

Lectures: Jan 16 + Jan 18, 2017

Prof. Will Evans *Scribe: David Penco*

In this lecture we discussed:

- Linear decision trees
- Connected sets
- Complexity of Element Uniqueness

Handouts (posted on webpage):

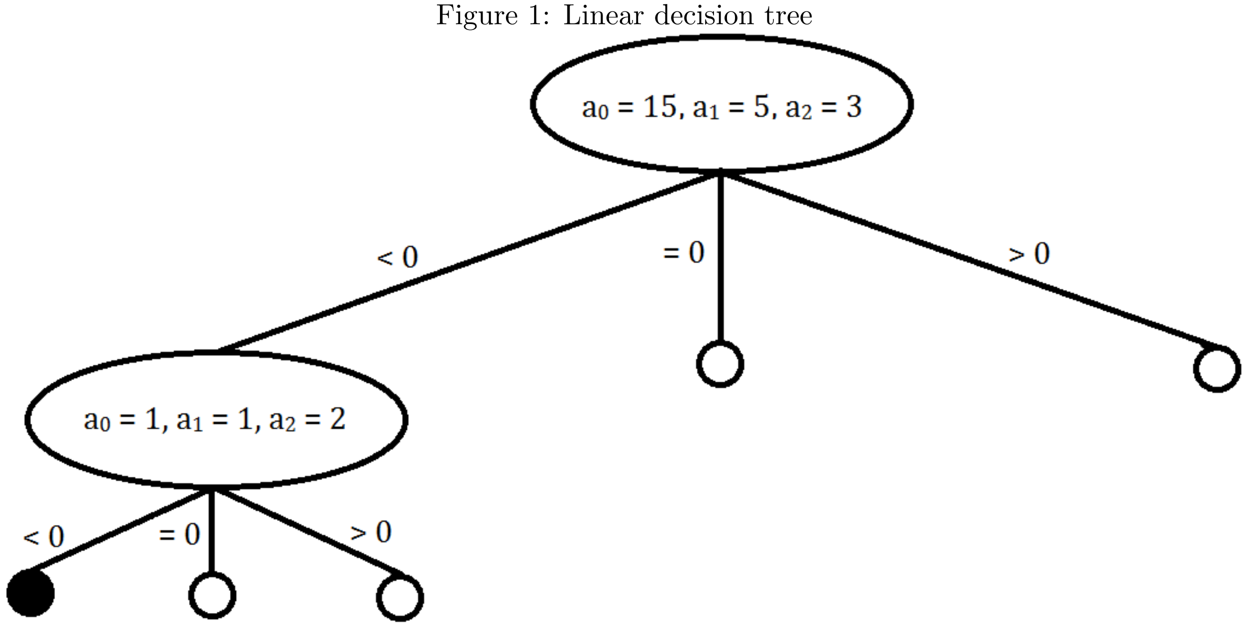
- In-class exercises

Reading: None

1 LINEAR DECISION TREES

A **linear decision tree** is a decision tree whose nodes are denoted by an $(n+1)-$tuple $(a_0, a_1, a_2, ..., a_n)$, where the a_i's represent the coefficients of the linear combination $a_0 + a_1x_1 + a_2x_2 + ... + a_nx_n$. See Figure 1 for an example of a linear decision tree where $n = 2$.

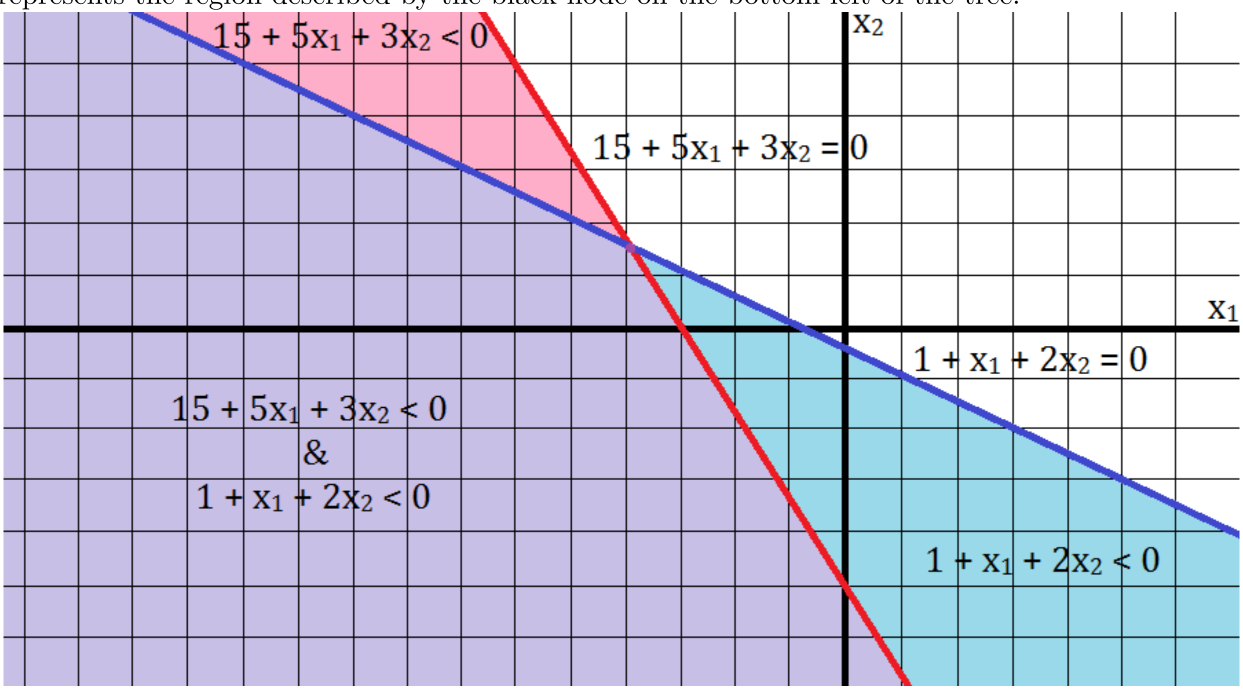
Figure 1: Linear decision tree



Some key facts about linear decision trees:

1. Each internal node tests the sign of a linear function.
2. Inputs that follow a child edge lie in some **halfspace** (for $>$ or $<$ edges) or a **hyperplane** (for $=$ edges). Both the halfspaces and the hyperplane are convex.
3. Inputs that reach a leaf lie in the intersection of a sequence of halfspaces or hyperplanes, which are all convex. The intersection of convex sets is convex, so the inputs that reach a node from the root form a convex set.

Figure 2: Geometric interpretation of the linear decision tree of Figure 1. The purple region represents the region described by the black node on the bottom left of the tree.



2 CONNECTED SETS

Let F be a function (like Element Uniqueness). Let $F_t = \{x | F(x) = t\}$. Let $\#F_t$ be the number of connected components in F_t. A set S is **connected** if for all points $a, b \in S$, there exists a path from a to b that lies completely in S. Note that this is similar to the definition of convexity, except for connectedness we do not have to take a straight line path.

For example, in Element Uniqueness, $\#F_{NO} = 1$, since you can always follow a path from any NO point to the origin. (We will prove this in Assignment 2, Exercise 6).

Lemma: Any linear decision tree that computes F has height at least $\lceil log_3(\sum_{\text{outputs } t} \#F_t)\rceil$.

Proof. Each leaf represents one connected component that has the same output. There are at least $\sum_{\text{outputs } t} \#F_t$ leaves, so the height of the tree is at least $\lceil log_3(\sum_{\text{outputs } t} \#F_t)\rceil$. ■

3 COMPLEXITY OF ELEMENT UNIQUENESS

Theorem: Any linear decision tree that computes Element Uniqueness has height $\Omega(nlog(n))$.

Proof Idea: Are $x = (1, 2, 3, 4, ..., n)$ and $y = (2, 1, 3, 4, ..., n)$ (both are examples of YES inputs) in the same connected component? No, they are not.

Proof. Let x be a vector of n unique numbers and let $y \neq x$ be any permutation of x. First off, we know there are $n!$ permutations of x, so there are $n! - 1$ possible choices of y for a given x. Now, since y is a permutation of x but x and y are not equal, there must be indices i, j such that $x_i < x_j$ and $y_i > y_j$. Any continuous path from x to y must contain a point z with $z_i = z_j$ (by the Intermediate Value Theorem).

To see this, let p be a continuous path from x to y. We can define p as a function like so: $p : [0, 1] \to \mathbb{R}^n$, $p(0) = x$, $p(1) = y$. In other words, over the time interval from $t = 0$ to $t = 1$, our path p goes from x to y. Keep in mind that p must be continuous. Now, define $q(t) = p(t)_j - p(t)_i$. Because $p(0) = x$ and $p(1) = y$, we know that $q(0) = x_j - x_i > 0$ and $q(1) = y_j - y_i < 0$. By the Intermediate Value Theorem, then, because our path p is continuous, there must exist a time $t \in (0, 1)$ such that $q(t) = 0$, i.e. there exists a time t such that $p(t)_j - p(t)_i = 0$, meaning $p(t)_j = p(t)_i$.

So there exists a point z on the path where $z_i = z_j$. But z is a NO point, since two of its coordinates are equal. So, any continuous path from x to y must pass through a NO point, and therefore x and y are not in the same connected component. Thus, none of the $n!$ permutations of x are in the same connected component as each other, which implies that the number of connected components of Element Uniqueness is at least $n!$. In mathematical notation, if we let $F =$ Element Uniqueness, we have $\#F_{YES} \geq n!$.

Therefore, by our earlier lemma, any linear decision tree that computes Element Uniqueness has height at least $\lceil log_3(\sum_{\text{outputs } t} \#F_t)\rceil = \lceil log_3(n!)\rceil$, which is $\Omega(nlog(n))$. ■