Classification Tasks

Goal: Predict label of a given point

Formally:

Let $C = \{c_1, c_2, \ldots, c_k\}$ set of labels

$\mathcal{R}$ be domain for our data

Goal: Create a function (or model) $M: \mathcal{R} \to C$

Classification often called supervised learning

given training set

points from $\mathcal{R}$ and their label

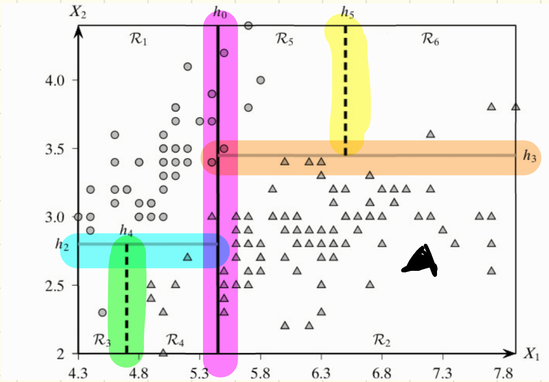
after we create M,

predict labels of pts not in training set

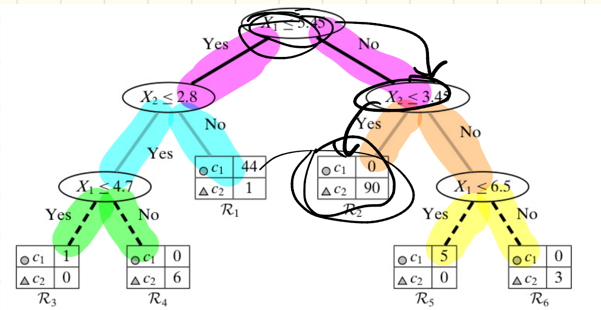
Decision Trees

Let $D = \{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n\}$ be n pts in $\mathcal{R} \subseteq \mathbb{R}^d$
$\vec{x}_i \in D$, $\hat{y}_i$ is the label of $\vec{x}_i$

Decision tree: uses axis-parallel hyperplanes
to split $\mathcal{R}$ into $\mathcal{R}_1$ & $\mathcal{R}_2$
$\Rightarrow$ a partion of D into D_1 & D_2
(repeat until regions are relatively homogeneous)



Key benifits:
- easy to understand
- simple to implement
- can handle $\boxed{\text{numeric}}$ and catagorical attributes
- decison rules are easy to interpret



Decision tree rep:
- internal nodes: rep decisions (splits defined by hyperplanes)
- leaf nodes: rep partitions of the data space
- label by a simple majority vote

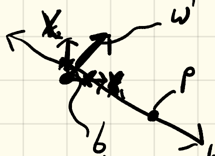
Axis-parallel hyperplanes

Let $\vec{w} \in \mathbb{R}^d$, $b \in \mathbb{R}$

hyperplane is all pts $\vec{x} \in \mathbb{R}^d$
$$h(\vec{x}) := \vec{w}^T \vec{x} + b = 0$$

normal to the hyperplane

$1 \cdot p_x + 2p_y - \frac{1}{2} = 0$



$1x_1 + 2x_2 - \frac{1}{2} = 0$

$w = (1, 2)$

$b = -\frac{1}{2}$

offset from the origin

$x_1 = 0 \quad x_2 = \frac{1}{4}$

$x_1 = \frac{1}{2} \quad x_2 = 0$

Decision tree: consider only axis-parallel hyperplanes

$$w = (0, 0, \ldots, 0, 1, 0, 0, \ldots, 0) = e_j$$

j^{th} entry is 1

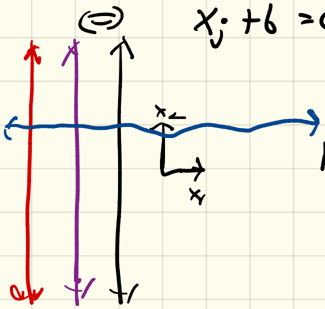
all others are 0

$$\Rightarrow \vec{x} = (x_1, x_2, \ldots, x_d)$$

$$h(\vec{x}) := \vec{e}_j^T \vec{x} + b = 0$$

$$\Leftrightarrow \quad x_j + b = 0$$

different b we get different pallell hyperplanes



hyperplane

$x_1 + 1 = 0$ $\quad$ $x_2 + 1 = 0$

$x_1 + 2 = 0$

$x_1 + 3 = 0$

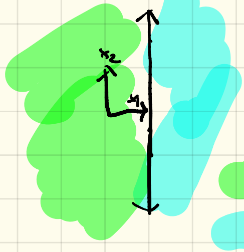
Split points

each hyperplan splits $\mathcal{R}$ into 2 halfspaces

Given $h(\vec{x})$

A: $\vec{y} \in \mathbb{R}^d \quad h(\vec{y}) \le 0$

$\vec{z} \in \mathbb{R}^d \quad h(\vec{z}) > 0$



$x_1 + -1 = 0$

$p \in \mathbb{R} \ni p \le 1$

$q \in \mathbb{R} \ni q > 1$

Since our hyperplanes are axis-aligned

sliplifies to

$\exists e_i \ni h(\vec{x}) = e_i^T \vec{x} + b$

$\iff x_i + b$

$\iff x_i \le -b$

In general

"generic" split pant is encoded as (x_j, v)

dim

$-b$

partions $\mathcal{R} = \{\mathcal{R}_Y, \mathcal{R}_N\}$

pts satifying inequality

pts not satisfy inequality