DECISION TREES

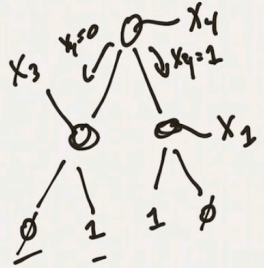
- HEURISTICS FOR LEARNING DEC TREES
- THEORETICAL PROPERTIES

DECISION TREE IS A BOOLEAN FUNCTION

(OUTPUTS +1 or -1
0 or 1)

INPUT:

$X \in \{0,1\}^n$ $\quad f(x) \to \{0,1\}$



SIZE OF DECISION TREE = # OF NODES IN THE TREE

DEPTH OF TREE = LENGTH OF LONGEST PATH FROM ROOT TO A LEAF.

GIVEN A SET OF LABELED EXAMPLES,
BUILD A TREE WITH LOW ERROR.

S = TRAINING SET $\quad (x^1, y^1)$ $\quad y^i \in \{0,1\}$

$\vdots$ $\quad x^i \in \{0,1\}^n$

(x^m, y^m)

- ERROR RATE
- TRAINING ERROR
- EMPIRICAL ERROR RATE

$=$ FIX T

$$\frac{\#\text{MISTAKES } T \text{ MAKES ON } S}{|S|}$$

NATURAL APPROACH FOR BUILDING DECISION TREES

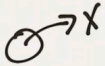
GIVEN A SET S.

S — ASSUME: TREE IS A LEAF

ALWAYS +1 OR ALWAYS 0

CHOOSE MAJORITY OF LABELS

- A MORE INTERESTING TREE:



— HOW TO DECIDE WHAT LITERAL TO PUT AT THE ROOT OF TREE?

DEFINE A POTENTIAL FUNCTION $\phi(a)$

$$\phi(a) = \min(a, 1-a)$$

PICK A LITERAL x_i

ASSUME 5 POS EXAMPLES

COMPUTE $\phi\left(\Pr_{(x,y)\sim S}(y=0)\right)$

ASSUME 10 NEG EXAMPLES

$$\phi\left(\Pr_{(x,y)\sim S}(y=0)\right) = \phi\left(\frac{1}{3}\right) = \min\left(\frac{1}{3}, \frac{2}{3}\right) = \frac{1}{3}.$$

$$\phi\left(\Pr_{(x,y)\sim S}(y=0)\right) = \phi\left(\frac{2}{3}\right) = \min\left(\frac{1}{3}, \frac{2}{3}\right) = \frac{1}{3}$$

ERROR RATE FOR TREE WITH JUST 1 LEAF

$$\phi\left(\Pr_{(x,y)\sim S}(y=0)\right)$$ ← ERROR RATE OF TRIVIAL DEC TREE

PICK LITERAL x_1

GAIN (x_1) =
OLD RATE −
NEW RATE USING x_1

WHAT IS NEW ERROR RATE?

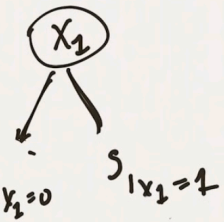
$$\Pr_{x,y\sim S}[x_1=0]\cdot\phi\left(\Pr_{x,y\sim S}(y=0\mid x_1=0)\right)+$$
$$\Pr_{x,y\sim S}[x_1=1]\cdot\phi\left(\Pr_{x,y\sim S}(y=0\mid x_1=1)\right)$$

} NEW ERROR RATE

$$\phi\left(\Pr_{(x,y)\sim S}(y=0)\right)$$ ← ERROR RATE OF TRIVIAL DEC TREE

PICK LITERAL x_1

GAIN (x_1) =
OLD RATE −
NEW RATE USING x_1



WHAT IS NEW ERROR RATE?

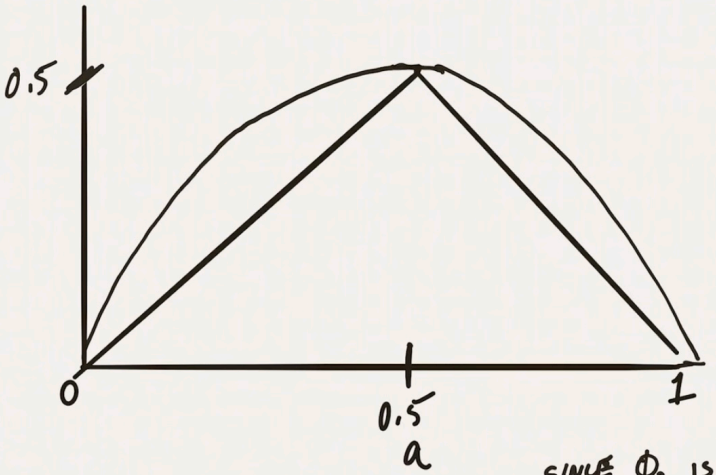
$$\Pr_{x,y\sim S}[x_1=0]\cdot\phi\left(\Pr_{x,y\sim S}(y=0\mid x_1=0)\right)+$$
$$\Pr_{x,y\sim S}[x_1=1]\cdot\phi\left(\Pr_{x,y\sim S}(y=0\mid x_1=1)\right)$$

} NEW ERROR RATE

STRUCTURE OF TREE IS DETERMINED BY CHOICE OF ϕ

$\phi(a) = \min(a, 1-a)$ CORRESPONDED TO TRAINING ERROR

$\phi(a) = 2 \cdot a \cdot 1-a$ "GINI FUNCTION" OR "GINI INDEX"



$\phi_1(a) = \min(a, 1-a)$

$\phi_2(a) = 2 \cdot a \cdot (1-a)$

SINCE ϕ_2 IS AN UPPER BOUND ON ϕ_1

$\Rightarrow$ SMALL VALUES OF ϕ_2 ⟨ TRAINING ERROR

$\Rightarrow$ SMALL VALUES OF ϕ_1

$S =$

X_1	X_2	POS	NEG
0	0	1	1
0	1	2	1
1	0	3	1
1	1	4	2

$\phi(a) = 2 \cdot a \cdot (1-a)$

$\phi(\Pr_S[\text{NEG}]) = \frac{4}{9}$

$\frac{5}{15} = \frac{1}{3}$

$2 \cdot \frac{1}{3} \cdot \frac{2}{3} = \frac{4}{9}$

PICK X_1 TO BE AT ROOT? $\frac{11}{25}$

X_2 TO BE AT ROOT?

LOOK AT X_1

$\Pr(X_1=0) \cdot \phi(\Pr(\text{NEG} \mid X_1=0)) + \Pr(X_1=1) \cdot \phi(\Pr(\text{NEG} \mid X_1=1))$

$\frac{1}{2} \cdot 2 \cdot \frac{2}{5} \cdot \frac{3}{5} + \frac{1}{2} \cdot 2 \cdot \frac{3}{5} \cdot \frac{2}{5} = \frac{2}{25} = \frac{11}{25}$

($\frac{4}{25}$)

$S =$

X_1	X_2	POS	NEG
0	0	1	1
0	1	2	1
1	0	3	1
1	1	4	2

$\phi(a) = 2 \cdot a \cdot (1-a)$

$\phi(\Pr_S[\text{NEG}]) = \frac{4}{9}$

$\frac{5}{15} = \frac{1}{3}$

$2 \cdot \frac{1}{3} \cdot \frac{2}{3} = \frac{4}{9}$

PICK X_1 TO BE AT ROOT? $\frac{11}{25}$ $\left(\frac{4}{9} - \frac{11}{25}\right) > 0$

X_2 TO BE AT ROOT? $\frac{4}{9}$ $\left(\frac{4}{9} - \frac{4}{9}\right) = 0$

LOOK AT $\frac{2}{5} \cdot 2 \cdot \frac{1}{3} \cdot \frac{2}{3} + \frac{3}{5} \cdot \frac{4}{9} = \frac{4}{9}$

($X_2 = 0$) ($X_2 = 1$)

ONE QUESTION IS:

WHEN SHOULD WE STOP?

→ MANY ANSWERS

ONE ANSWER: STOP WHEN GAIN IS EXTREMELY SMALL FOR ALL LITERALS

PRUNING: BUILD AN ENORMOUS TREE

PARAMETER INDICATING HOW MANY NODES YOU WANT.

RANDOM FORESTS

- BUILD MANY "SMALL" DECISION TREES
- TAKE A MAJORITY OF RESULTING TREES

ALGORITHM FOR BUILDING MANY TREES:

TAKE TRAINING SET S

1. RANDOMLY SUBSAMPLE FROM S to CREATE S'
2. RANDOMLY PICK SOME FEATURES FROM $\{X_1, \ldots, X_n\}$ OF SIZE K.

BUILD A DECISION TREE USING S' AND THE K RANDOM FEATURES.