ML HW2

SHREYASH SHRIVASTAVA
1001397477

Decision Tree (PROBLEM 2)

1. Compute the entropy for the data set y

50 Edible 50 Poisonous

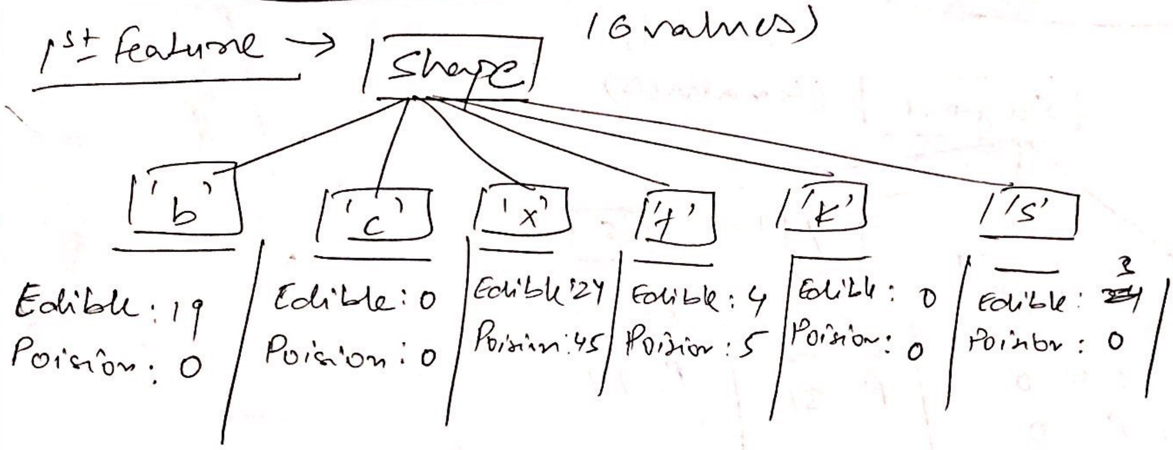
$E(S) = -P(\text{Edible}) \log_2 P(\text{Edible}) - P(\text{Poision}) \log_2 P(\text{Poision})$

$= -(1/2)\log_2(1/2) - (1/2)\log_2(1/2)$

$E(S) = 1$ (Total Entropy)

2. which node / attribute as root node?

1st feature → Shape (6 values)



Calculate entropy and information gain for all of the different nodes

$E(\text{shape}=b) = -\frac{19}{19}\log_2\frac{19}{19} - 0$

$= 0$

$E(\text{shape}=k)$ (irrelevant)

$E(\text{shape}=s) = 0$

$E(\text{shape}=c) =$ (irrelevant feature)

$E(\text{shape}=x) = -\frac{24}{69}\log\left(\frac{24}{69}\right) - \frac{45}{69}\log_2\left(\frac{45}{69}\right)$

$= 0.932$

$E(\text{shape}=f) = -\frac{4}{9}\log\left(\frac{4}{9}\right) - \frac{5}{9}\log\left(\frac{5}{9}\right)$

$= 0.9910$

Scanned with CamScanner

Information from Cap-shape

$$I(\text{Cap-shape}) = \frac{19}{100} \times 0 + \frac{69}{100} \times 0.982 + \frac{9}{100} \times .991$$

$$= 0.7322$$

Information gained from cap-shape

$$\text{Gain}(\text{Cap-shape}) = E(S) - I(\text{Cap-shape})$$

$$= 1 - 0.7322$$

$$\boxed{= 0.2678}$$

2nd feature → Surface (4 values)

f: Edible: 10, Poison: 0

g: E: 0, P: 0

y: E: 22, P: 31

s: E: 18, P: 19

$$E(\text{Surface} = f) = 0$$

$E(\text{Surface} = g)$ (irrelevant feature)

$$E(\text{Surface} = y) = -\frac{22}{53} \log\left(\frac{22}{53}\right) - \log\left(\frac{31}{52}\right) \log\left(\frac{31}{53}\right)$$

$$= 0.979$$

$$E(\text{Surface} = s) = -\frac{18}{37} \log_2\left(\frac{18}{37}\right) - \frac{19}{37} \log_2\left(\frac{19}{37}\right)$$

$$= 0.999$$

Information gain from Cap-surface

$$I(\text{Cap-Surface}) = \frac{53}{100} \times 0.979 + \frac{37}{100} \times 0.999 = 0.888$$

$$\boxed{\text{Gain} = 1 - 0.888 = 0.112}$$

Scanned with CamScanner

3rd feature → Color (10 values)

n	b	c	g	r	p	u	e	w	y	~~t~~	~~f~~
p: 6	p: 0	p: 0	p: 5	p: 0	p: 0	p: 0	p: 0	p: 12	p: 22	p:	p:
e: 27	e: 0	e: 0	e: 0	e: 0	e: 0	e: 0	e: 0	e: 23	e: 0	e:	e:
	(ir)	(ir)									

$$E(\text{color}=n) = -\frac{6}{33}\log_2\left(\frac{6}{33}\right) - \frac{27}{33}\log_2\left(\frac{27}{33}\right) = 0.684$$

$$E(\text{color}=g) = 0$$

$$E(\text{color}=w) = -\frac{12}{35}\log_2\left(\frac{12}{35}\right) - \frac{23}{35}\log_2\left(\frac{23}{35}\right) = 0.927$$

$$E(y) = 0$$

Information gain from Cap-Color

$$I(\text{Cap-Color}) = \frac{33}{100} \times 0.684 + \frac{35}{100} \times 0.927 = 0.649$$

$$\boxed{\text{Gain} = 1 - 0.649 = 0.350}$$

4th feature → Bruises

t	f
E: 41	E: 9
P: 50	P: 0

$$E(\text{Bruises}=t) = -\frac{41}{91}\log_2\left(\frac{41}{91}\right) - \frac{50}{91}\log_2\left(\frac{50}{91}\right) = 0.99$$

$$E(\text{bruises}=f) = 0$$

$$I(\text{bruises}) = \frac{91}{100} \times 0.99 = 0.909$$

$$\boxed{\text{Gain} = 1 - 0.909 = 0.0991}$$

5th feature = $\boxed{\text{odor}}$

odor	a	l	c	y	m	n	p	s
E:	19	22	0	0	0	9	0	0
P:	0	0	0	0	0	0	50	0
			(i'r)	(i'r)	(i'o)			(i'o)

E(odor = a) = 0

E(odor = l) = 0

E(odor = n) = 0

E(odor = p) = 0

I(odor) = 0

$\boxed{\text{gain}() = 1 - 0 = \underline{1}}$

—— y ——

1. Shape

I(shape) = 0.732

Gain(shape) = 0.2678

3. Color

I(Color) = 0.649

Gain(color) = 0.350

selecting this

2. Surface

I(surface) = 0.888

Gain(surface) = 0.112

4. Bruises

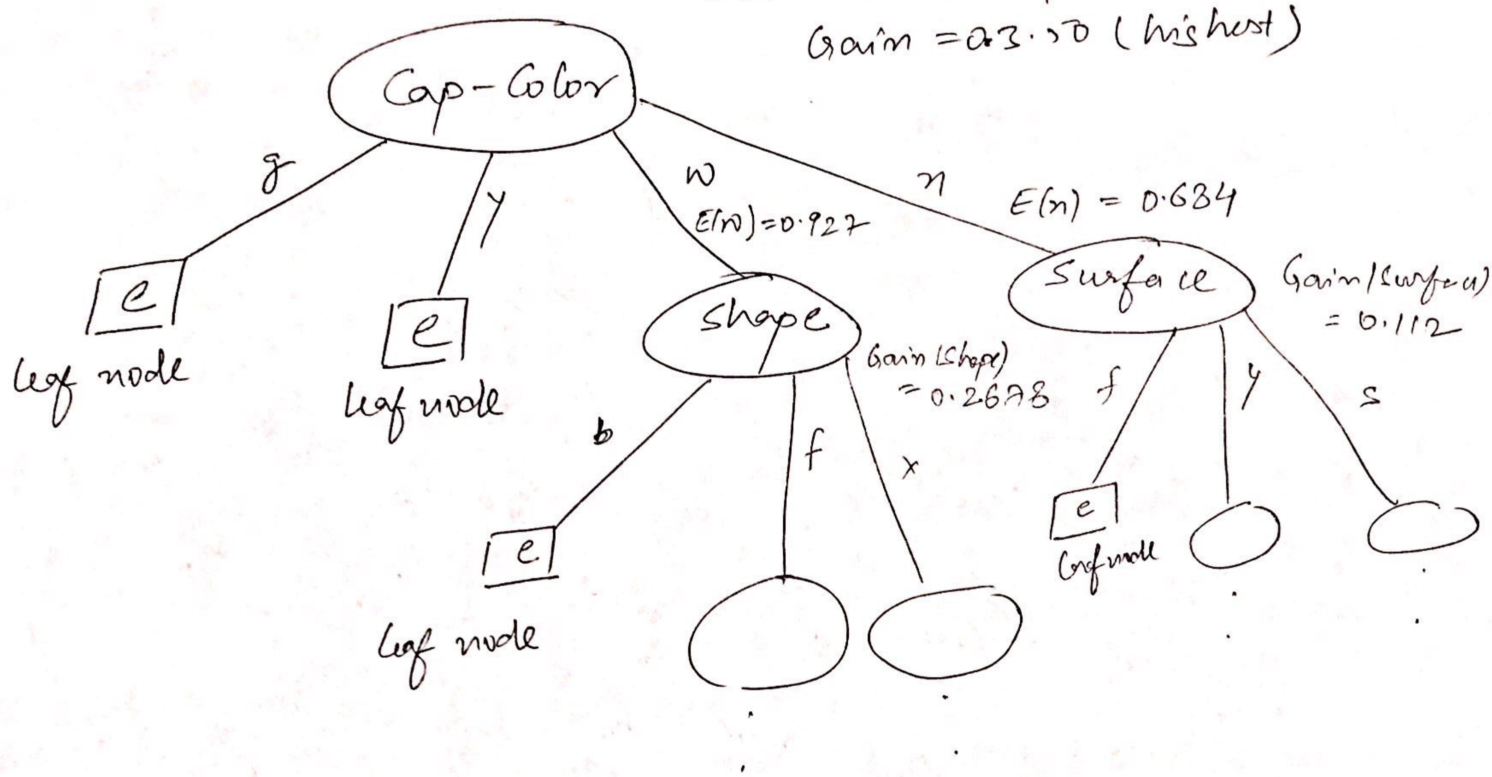
I(Bruises) = 0.909

Gain(Bruises) = 0.0991

The fifth feature 'odor' will not be considered since it has the highest gain and (0 Entropy), it will ~~onof~~ be the only feature required and all other features will become irrelevant.

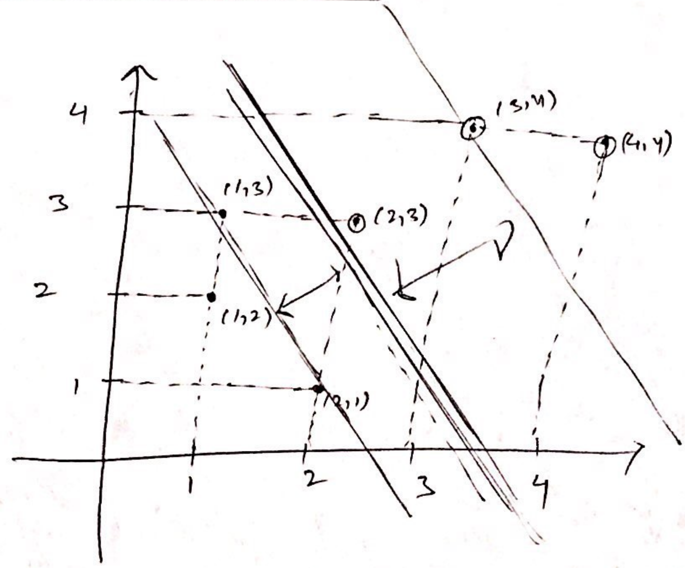
2-level decision tree

root node : Cap Color
Gain = 0.350 (highest)



Support Vector Machines

(a)



$\bullet \longrightarrow -1$

$\odot \longrightarrow +1$

$D = \{((1,2),-1),$
$((2,3),1),$
$((2,1),-1)$
$((3,4),1),$
$((1,3),-1),$
$((4,4),1)\}$

Aim: To find a descision line

$$x \cdot w + b = 0$$

$$x_i \cdot w + b \geq +1 \text{ for } i \in (1 \ldots N) \text{ where } y_i = +1$$
$$x_i \cdot w + b \leq -1 \text{ for } i \in (1 \ldots N) \text{ where } y_i = -1$$

- The margin of separation

$$d(w) = (x_+ - x_-) \cdot \frac{w}{\|w\|} = \frac{2}{\|w\|}$$

- Maximizing $d(w)$ is equivalent to minimizing $\|w\|^2$. So the constraint optimization problem in SVM is

$$\min : \frac{1}{2}\|w\|^2$$

$$\text{subject to } \quad y_i(x_i \cdot w + b) \geq 1 \quad \forall i = 1 \ldots N$$

Equivalently, minimizing a Lagrangian function

$$\min \; L(w, b, \{\alpha_i\}) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{N} \alpha_i [y_i(x_i \cdot w + b) - 1]$$

$$\text{subject to } \alpha_i \geq 0,\; y_i(x_i \cdot w + b) - 1 \geq 0,$$

$$\alpha_i [y_i(x_i \cdot w + b) - 1] = 0, \quad \forall i = 1 \ldots N$$

Now

$\downarrow$ (P.T.O)

$$\frac{\partial L(w, b, \{\lambda_i\})}{\partial b} = 0 \quad \text{and} \quad \frac{\partial L(w, b, \{\lambda_i\})}{\partial w} = 0$$

subject to constraint $\lambda_i \geq 0$

$$\rightarrow \sum_{i=1}^{N} \alpha_i y_i = 0 \quad \text{and} \quad w = \sum_{i=1}^{N} \alpha_i y_i x_i.$$

Substituting this back into the Lagrangian function

$$L(w, b, \{\lambda_i\}) = \frac{1}{2}(w \cdot w) - \sum_{i=1}^{N} \lambda_i y_i (x_i \cdot w) - \sum_{i=1}^{N} \lambda_i y_i b + \sum_{i=1}^{N} \lambda_i.$$

$$= \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j)$$

This results in

$$\max_{\lambda} \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j y_i y_j (x_i \cdot x_j)$$

subject to

$$\sum_{i=1}^{N} \alpha_i y_i = 0 \quad \text{and} \quad \lambda_i \geq 0, \; i = 1 \ldots N$$

(c) Using SVM solver

$w = [0.8, 0.4]$

The support vectors are (2, 1) and (3, 4)