

Quiz 2 - sample, with solutions

Note Title

9/12/2007

1. Compute the entropy of the following probability distributions for the random variable X :

(a)

x	apple	banana	lemon	grape	plum
$p(X=x)$	$\frac{1}{16}$	$\frac{1}{2}$	$\frac{1}{16}$	$\frac{1}{4}$	$\frac{1}{4}$

(b)

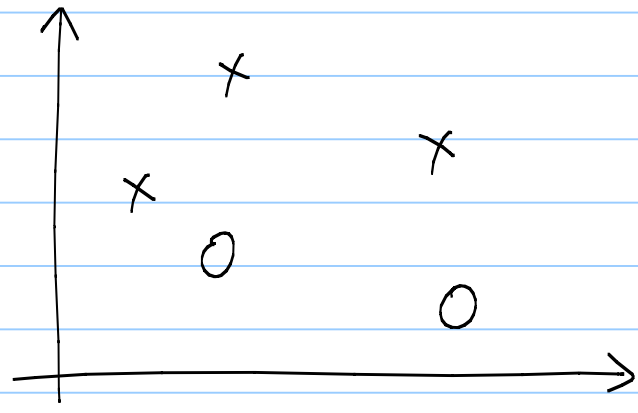
x	1	2	3	-----	256
$p(X=x)$	$\frac{1}{256}$	$\frac{1}{256}$	$\frac{1}{256}$		$\frac{1}{256}$

2. Using the greedy algorithm described in class, with entropy as the split criterion, construct a decision tree whose nodes are as pure as possible for the following data set:

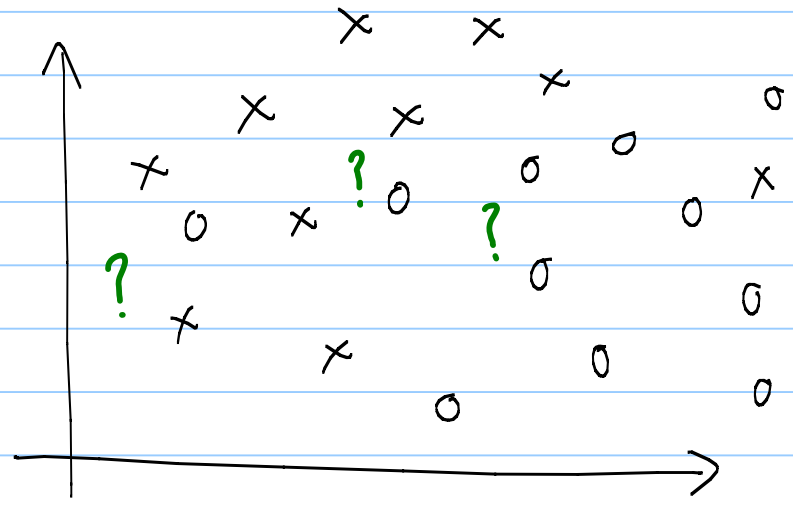
attributes : color { Blue, Green, Yellow }
 sound { Quiet, Loud }
 texture { Rough, Smooth }
(class vble) → material { Wood, Metal, Fibreglass }

Color	Sound	Texture	Material
B	Q	R	W
B	L	R	W
B	Q	S	M
B	L	S	M
G	Q	S	W
G	L	S	W
Y	Q	R	F
Y	L	R	F

3. a) Draw the decision boundaries for 1-nearest-neighbor on the dataset:



b) Indicate the class of each '?' according to the 3-nearest-neighbor algorithm:



Solutions

$$1) a) \text{ entropy} = \sum_i p_i \log_2 \frac{1}{p_i}$$

$$= \frac{1}{16} \times \log_2 16 + \frac{1}{2} \times \log_2 2 + \dots$$

$$= \frac{4}{16} + \frac{1}{2} + \frac{4}{16} + \frac{2}{4} + \frac{2}{4}$$

$$= 2 \text{ bits}$$

$$b) \text{ entropy} = \sum_i p_i \log_2 \frac{1}{p_i}$$

$$= \sum_{i=1}^{256} \frac{1}{256} \times \log_2 256$$

$$= 256 \times \frac{1}{256} \times 8$$

$$= 8 \text{ bits}$$

2) Try splitting on each attribute and see which results in lowest expected entropy.

(i) <u>color</u>		W	M	F	tot	entropy	weight
	B	2	2	0	4	1	$\frac{4}{8}$
	G	2	0	0	2	0	$\frac{2}{8}$
	Y	0	0	2	2	0	$\frac{2}{8}$

$$\therefore \text{expected entropy} = \frac{4}{8} \times 1 + \frac{2}{8} \times 0 + \frac{2}{8} \times 0 = \frac{1}{2}$$

(ii) <u>texture</u>		W	M	F	tot	entropy	weight
	R	2	0	2	4	1	$\frac{4}{8}$
	S	2	2	0	4	1	$\frac{4}{8}$

$$\therefore \text{expected entropy} = \frac{4}{8} \times 1 + \frac{4}{8} \times 1 = 1$$

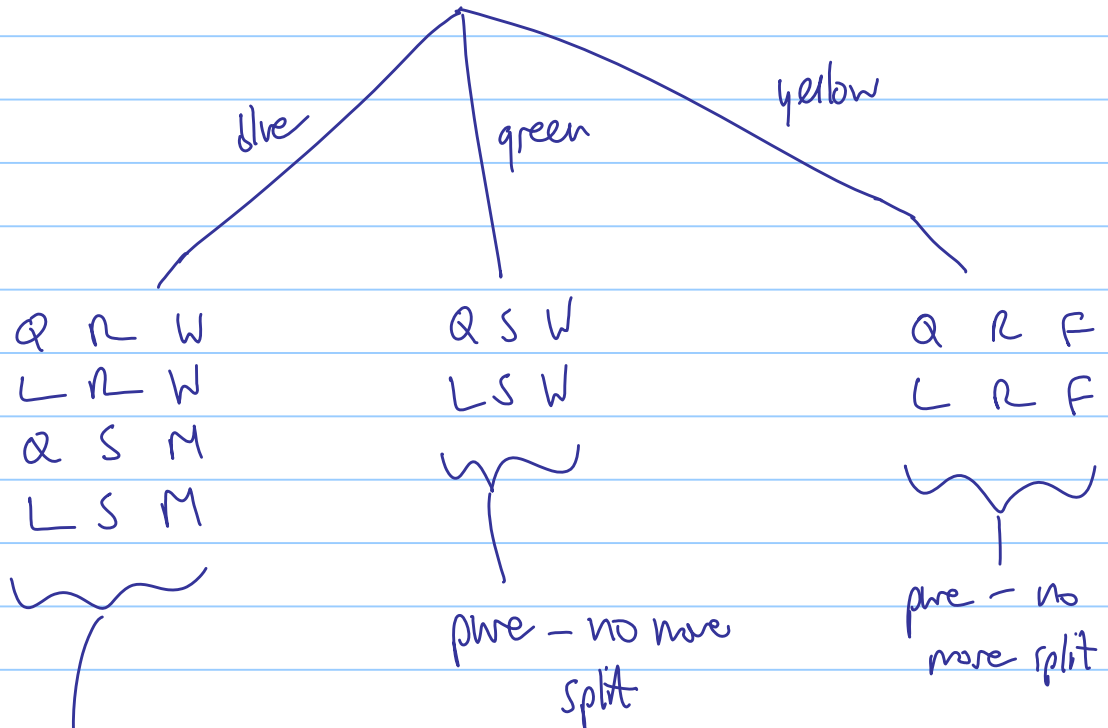
(iii) <u>sound</u>		W	M	F	tot	entropy	weight
	Q	2	1	1	4	$\frac{3}{2}$	$\frac{4}{8}$
	L	2	1	1	4	$\frac{3}{2}$	$\frac{4}{8}$

$$\therefore \text{expected entropy} = \frac{4}{8} \times \frac{3}{2} + \frac{4}{8} \times \frac{3}{2} = \frac{3}{2}$$

\Rightarrow color split has lowest expected entropy ($\frac{1}{2}$ bit)

... so split on color first.

tree so far:



check each

attribute for lowest

expected entropy after split:

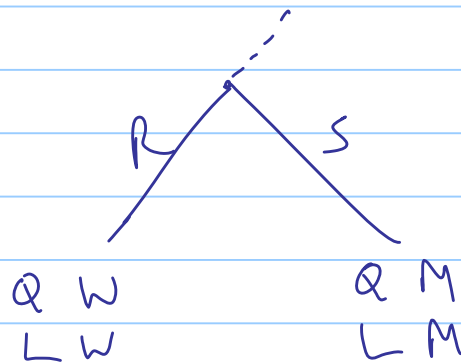
<u>Sound</u>	W	M	F	tot	entropy	weight
Q	1	1	0	2	1	$\frac{2}{4}$
L	1	1	0	2	1	$\frac{2}{4}$

$$\therefore \text{total expected Entropy} = \frac{2}{4} \times 1 + \frac{2}{4} \times 1 = 1$$

<u>texture</u>	W	M	F	tot	entropy	weight
R	2	0	0	2	0	$\frac{2}{4}$
S	0	2	0	2	0	$\frac{2}{4}$

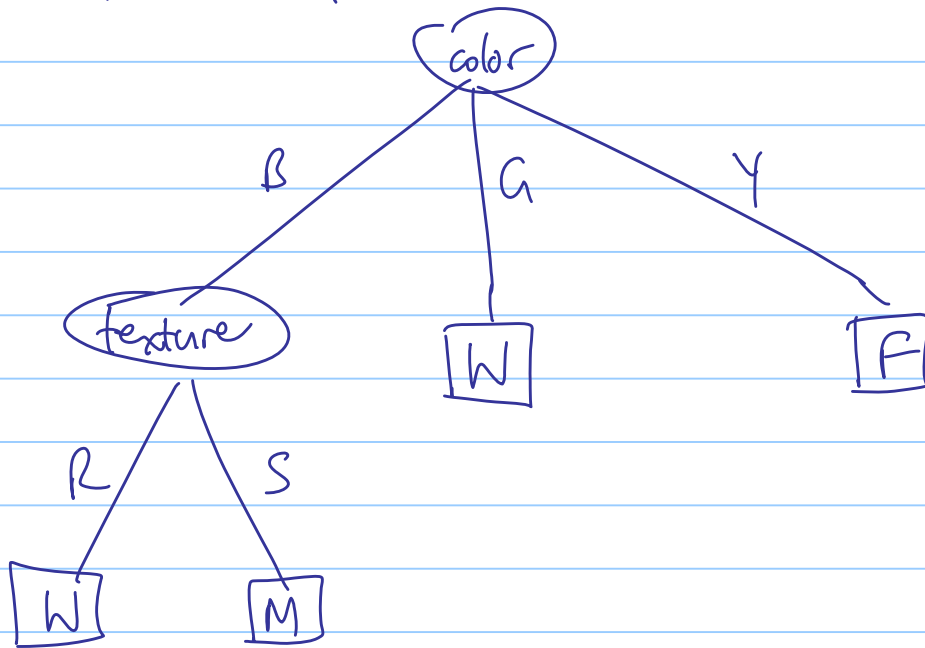
$$\therefore \text{expected entropy} = \frac{2}{4} \times 0 + \frac{2}{4} \times 0 = 0$$

So texture has least expected entropy. Splitting this node gives



- Both nodes pure, so we are done.

Final decision tree:



3 a)



b) from left to right: $x_1, x_2, 0$.