Sample Exam Questions for Nearest Neighbors & Decision Trees

Note Title

[Solutions are at the end of this document]

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

(a)

x	apple	banana	lemon	grape	plum
$p(X=x)$	$1/16$	$1/2$	$1/16$	$1/8$	$1/4$

(b)

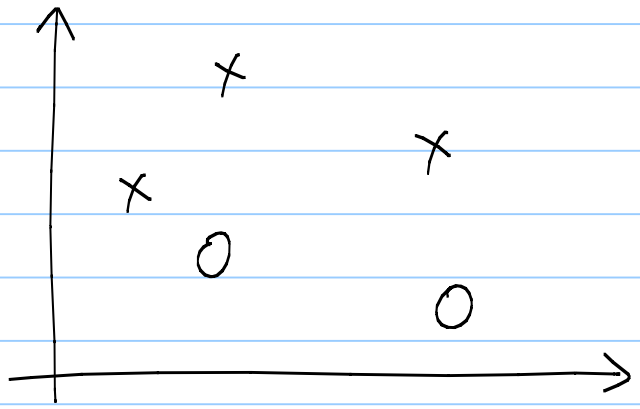
x	1	2	3	-----	256
$p(X=x)$	$1/256$	$1/256$	$1/256$		$1/256$

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:

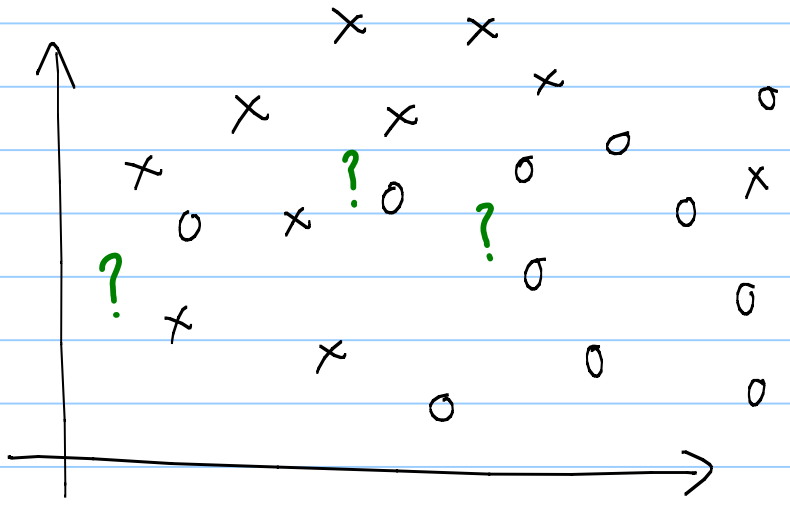
attributes : color { Blue, Green, Yellow }
sound { Quiet, Loud }
texture { Rough, Smooth }
(class vble) → material { Wood, Metal, Fibreglass }

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

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

$= \frac{4}{16} + \frac{1}{2} + \frac{4}{16} + \frac{3}{8} + \frac{2}{4} = \frac{15}{8}$

$= 1.875$ bits

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

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

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

$= 8$ bits

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

(i) color

	W	M	F	tot	entropy	weight
B	2	2	0	4	1	4/8
G	2	0	0	2	0	2/8
Y	0	0	2	2	0	2/8

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

(ii) texture

	W	M	F	tot	entropy	weight
R	2	0	2	4	1	4/8
S	2	2	0	4	1	4/8

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

(iii) sound

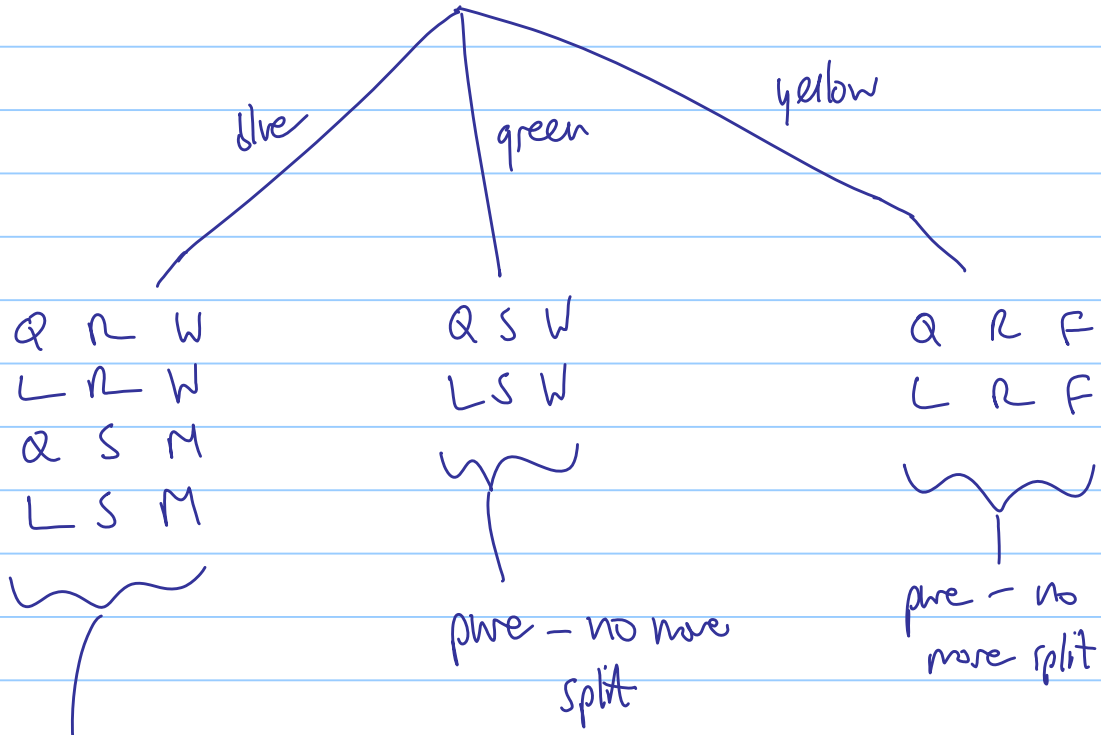
	W	M	F	tot	entropy	weight
Q	2	1	1	4	3/2	4/8
L	2	1	1	4	3/2	4/8

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

$\Rightarrow$ 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:

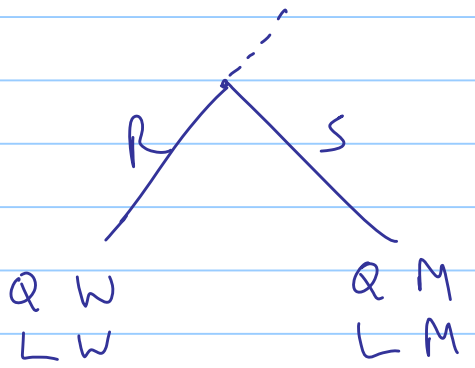
sound		W	M	F	tot	entropy	weight
	Q	1	1	0	2	1	2/4
	L	1	1	0	2	1	2/4

∴ total expected entropy = $\frac{2}{4} \times 1 + \frac{2}{4} \times 1 = 1$

texture		W	M	F	tot	entropy	weight
	R	2	0	0	2	0	2/4
	S	0	2	0	2	0	2/4

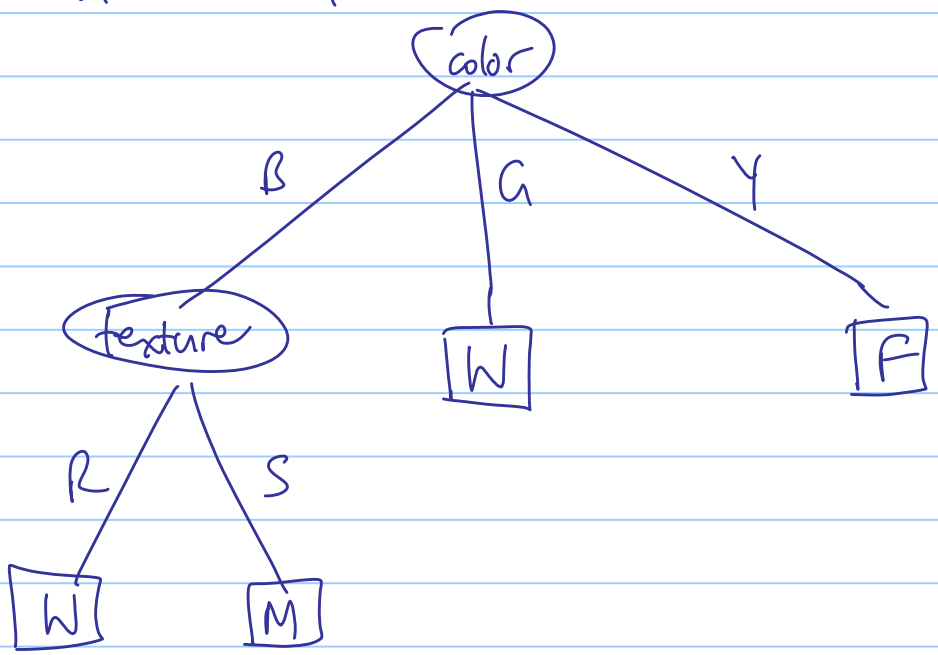
∴ expected entropy = $\frac{2}{4} \times 0 + \frac{2}{4} \times 0 = 0$

so texture has lowest expected entropy. Splitting this node gives

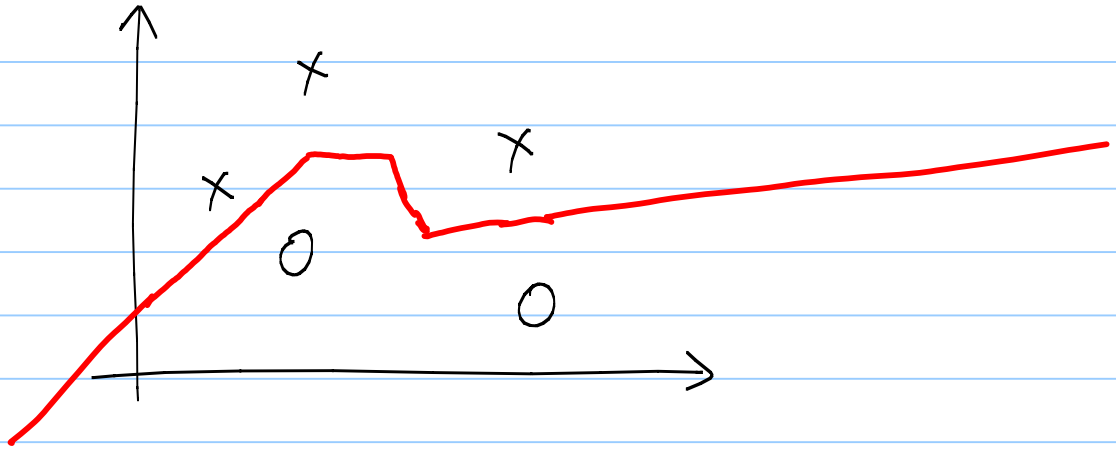


– both nodes pure, so we are done.

Final decision tree:



3 a)



b) from left to right: x, x, 0 .