

# Data Mining 2

Module 2 - 2020/2021

Name \_\_\_\_\_ Surname \_\_\_\_\_ ID: \_\_\_\_\_

Test id. AUTO

## Answers

Q1: 2, 5

Q2: no

Q3: Yes

Q4: *The trick is to transform the data from its original space  $x$  into a new space  $\Phi(x)$  so that a linear decision boundary can be used.*

Q5: 4 1

Q6: 3

Q7: 1

Q8: *Single,  $0.40.40.4*4 = 0.256$ .*

Q9: 3

Q10: 1

**Q11:**

<u>Plan</u>	Y	N	<u>Sex</u>	Y	N
travel	3/5	2/5	M	2/4	2/4
normal	3/4	1/4	F	2/5	3/5

Delta Gain =  $5/9 * 2/5 + 4/9 * 1/4 = 3/9$     Delta Gain =  $5/9 * 2/5 + 4/9 * 2/4 = 4/9$

plan	sex	minutes	churn	weight	new weight	norm weight
travel	F	90	N	1/9	0.11	0.07
travel	F	130	Y	1/9	0.11	0.07
travel	M	70	N	1/9	0.11	0.07
travel	M	80	N	1/9	0.11	0.07
normal	M	90	Y	1/9	0.41	0.255
normal	M	120	Y	1/9	0.11	0.07
normal	F	100	Y	1/9	0.11	0.07
normal	F	110	N	1/9	0.41	0.255
travel	F	100	N	1/9	0.11	0.07

Gain Function = Misclassification Error     $Z = 1.59$

70	80	90	90	100	100	110	120	130
N	N	N	Y	Y	N	N	Y	Y

<u>M&lt;90</u>	Y	N	<u>M&lt;100</u>	Y	N
<90	5/7	2/7	<100	1/4	3/4
>=90	0	2/2	>=100	3/5	2/5

Delta Gain =  $7/9 * 2/7 = 2/9$     Delta Gain =  $4/9 * 1/4 + 5/9 * 2/5 = 3/9$

<u>M&lt;120</u>	Y	N	<u>M&lt;110</u>	Y	N
<120	2/2	0	<110	2/6	4/6
>=120	2/7	5/7	>=110	1/3	2/3

Delta Gain =  $7/9 * 2/7 = 2/9$     Delta Gain =  $6/9 * 2/6 + 3/9 * 1/3 = 3/9$

Split by: Minutes

Error = 2/9

Alpha =  $1/2 * \ln((1-2/9)/(2/9)) = 1/2 \ln(7/2) = 0.626$

$w_i^{+1} = 2/9 * e^{0.626} = 0.41$   
misclassified

$w_i^{+1} = 2/9 * e^{-0.626} = 0.11$     classified  
correctly

List id questions: [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11]