Artificial Intelligence
Tutorial 4 on Machine Learning
Including Answers

Introduction

The following multiple choice questions are examples of typical questions one can expect on the AI exam. The questions on the AI exam are also multiple choice, but for this tutorial one has to explain the answers given. Moreover at the end one can find some open questions.

After the tutorial the answers to the MC will be available on Canvas.

Questions on Decision Trees

1. For marketing purposes a retailer wants to distinguish between costumers younger than 35 (class Y) and customers older than 35 (class O). The following table summa-rizes the data set in the data base of the retailer in an abstract form. The relevant attributes, determined by domain knowledge, are for convenience denoted by A with values $a1$, $a2$ and $a3$, B with values $b1$ and $b2$, C with values $c1$ and $c2$ and D with values $d1$ and $d2$

A	B	C	D	Number of Instances	
				Y	O
a1	b1	c1	d1	12	4
a2	b1	c1	d2	4	6
a3	b1	c1	d1	6	0
a1	b2	c1	d2	0	12
a2	b2	c1	d1	4	2
a3	b2	c1	d2	0	4
a1	b1	c2	d1	0	8
a2	b1	c2	d2	8	0
a3	b1	c2	d1	4	0
a1	b2	c2	d2	0	4
a2	b2	c2	d1	8	0
a3	b2	c2	d2	4	0

Given a new customer x, what is the information content in the answer to the ques - tion: "What is the class (Y or O) of x"? Or in other words what is entropy of the

data set above with respect to the class labels Y and O. Choose the alternative which is closest to your answer.

You can find a table of $-p \log 2(p)$ values at the last page of this exam.

  (a) 1.00

  (b) 0.00

  (c) 0.99

  (d) 0.01

Answer: we have 50 examples of class Y and 40 examples of class O. Hence the entropy is $H(5/9,$

$4/9) = 5-/9 \log 2(5/9) 4/-9 \log 2(4/9) = 0.52 + 0.47 = 0.99$. See table at the last page.

2. The analyst wants to learn the above classification problem using decision trees. If he uses "information gain" as selection criteria what will be the information gain of attribute A? Choose the alternative which is closest to your answer.

You can find a table of $-p \log 2(p)$ values at the last page of this exam.

  (a) 0.83

  (b) 0.16

  (c) 0.82

  (d) 0.17

Answer: construct the following table:

A	Y	O	p_Y	p_O	entropy
a1	12	28	3/10	7/10	0.88
a2	24	8	3/4	1/4	0.81
a3	14	4	7/9	2/9	0.76

Hence the average entropy for splitting on attribute A is: $40/90 * 0.88 + 32/90 * 0.81 + 18/90 * 0.76 = 0.83$. So the information gain is $0.99 - 0.83 = 0.16$.

3. A data analyst has collected data (see table below) about customer loans. The goals is to predict, based on the customer profile, if a loan for a customer has a high risk or not.

payment history	debt	guarantee	income	risk
average	low	no	15-35 KEuro	low
average	low	no	0-15 KEuro	high
average	low	no	> 35 KEuro	low
average	low	sufficient	> 35 KEuro	low
bad	low	no	0-15 KEuro	high
bad	low	sufficient	> 35 KEuro	low
good	high	sufficient	> 35 KEuro	low
good	high	no	0-15 KEuro	high
good	high	no	15-35 KEuro	high
good	high	no	> 35 KEuro	low
bad	high	no	15-35 KEuro	high

What is the information gain of the attribute *payment history* ? Answer: following the same approach as in the previous exercise will give an information gain of $1.00 - [4/11 * 0.81 + 4/11 * 1.00 + 3/11 * 0.92] = 1.00 - 0.91 = 0.09$

4. Consider the following dataset with attributes (features) A en B, both with values T en F. The class label is given by $+$ en $-$.

Ex.	A	B	Class label
1	T	F	+
2	T	T	+
3	T	T	+
4	T	F	+
5	T	T	+
6	F	F	-
7	F	F	-
8	F	F	-
9	T	T	-
10	T	F	-

What is the entropy of this dataset with respect to the class label?

(a) 1

(b) 0

(c) 1/2

(d) None of the above.

Answer: 1

5. Once again consider the dataset of the previous question. What is the information gain of attribute A?

   (a) 0.86

   (b) 0.30

   (c) 0.60

   (d) 0.40

   Answer; $1 - [7/10 * 0.86 + 3/10 * 0] = 0.40$

6. Consider the following confusion matrix

		Predicted class		
		C_1	C_2	C_3
Actual	C_1	120	15	20
Class	C_2	16	150	10
	C_3	22	3	130

   What is the accuracy of this classifier?

   (a) 150/486

   (b) 120/158+150/168+130/160

   (c) 120/135+150/176+130/155

   (d) 400/486
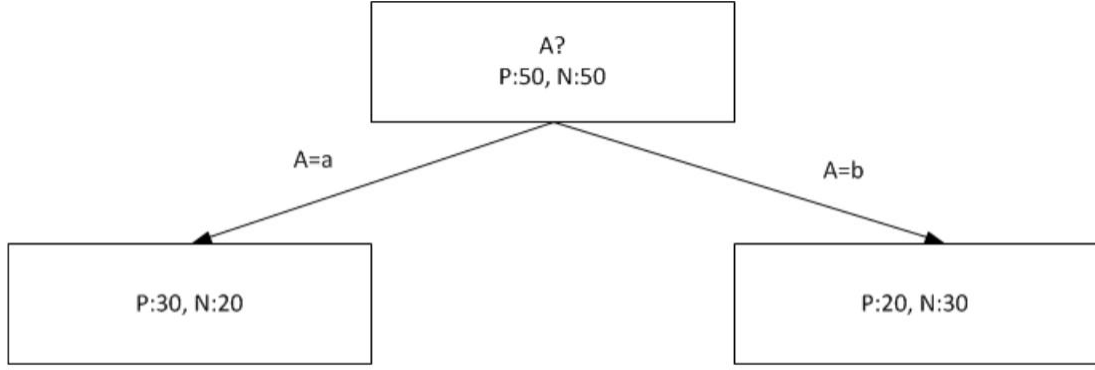
   Answer: (d)

7. Once again consider the confusion matrix of the previous question. What is the precision for class C_2?

   (a) 150/486

   (b) 150/168

   (c) 150/176

   (d) 150/400

   Answer: (b)

8. Consider the following part of the decision tree, with two leaf nodes and one parent node which splits on attribute A. The notation P:x N:y means that the node has x positive examples and y negative examples.

4

In order to apply χ^2 pruning one has to calculate the value of Δ:

$$\Delta = \sum_{i=1}^{i=d} [\frac{(p_i - \widehat{p_i})^2}{\widehat{p_i}} + \frac{(n_i - \widehat{n_i})^2}{\widehat{n_i}}]$$

What is the value of Δ in this case?

(a) 4

(b) 8

(c) 12

(d) None of the above

Answer: Observe that $p = 50 = n$, $p_1 = n_2 = 30$ and $p_2 = 20 = n_1$. The first step is to calculate $\widehat{p_1}$, $\widehat{n_1}$, $\widehat{p_2}$ and $\widehat{n_2}$: $\widehat{p_1} = \frac{p}{p+n} * (p_1 + n_1) = 0.5 * 50 = 25$. In similar way one gets $\widehat{n_1} = \widehat{p_2} = \widehat{n_2} = 25$. Hence

$$\Delta = \frac{(30 - 25)^2}{25} + \frac{(25 - 20)^2}{25} + \frac{(30 - 25)^2}{25} + \frac{(25 - 20)^2}{25} = 4$$

Table for $-p \log(p)$

p	$-p \log 2(p)$	p	$-p \log 2(p)$	p	$-p \log 2(p)$
0	0	1/8	0.38	1/10	0.33
1	0	2/8	0.50	2/10	0.46
1/2	0.50	3/8	0.53	3/10	0.52
1/3	0.53	4/8	0.50	4/10	0.53
2/3	0.39	5/8	0.42	5/10	0.50
1/4	0.50	6/8	0.31	6/10	0.44
2/4	0.50	7/8	0.17	7/10	0.36
3/4	0.31	1/9	0.35	8/10	0.26
1/5	0.46	2/9	0.48	9/10	0.14
2/5	0.53	3/9	0.53	1/11	0.31
3/5	0.44	4/9	0.52	2/11	0.45
4/5	0.26	5/9	0.47	3/11	0.51
1/6	0.43	6/9	0.39	4/11	0.53
2/6	0.53	7/9	0.28	5/11	0.52
3/6	0.50	8/9	0.15	6/11	0.48
4/6	0.39			7/11	0.42
5/6	0.22			8/11	0.33
1/7	0.40			9/11	0.24
2/7	0.51			10/11	0.13
3/7	0.52				
4/7	0.46				
5/7	0.35				
6/7	0.19				