Test Pearl 110 — Intelligent Interaction

Pearls of Computer Science (201300070)
October 20 2017, 13:45–14:45
Module coordinator: Doina Bucur & Maurice van Keulen
Instructor: Mannes Poel

- You may use 1 A4 sheet with your own notes for this test, as well as a *simple* calculator
- Scientific or graphical calculators, laptops, mobile phones, books etc. are not allowed. ***Put those in your bag now (with the sound switched off)!***
- Total number of points: 100

20 points **Question 1**

A bag contains 10 fair dice: of which 2 are a 4 sided dice, one is a 6 sided dice, 2 are a 8-sided dice, 2 are a 12-sided dice and 3 are a 20-sided dice. All the dices are fair, so that for example the probability of throwing a 3 with a 20 sided dice is $1/20$. John randomly draws a dice from the bag and afterwards throws with this dice. The outcome is 5. This is the data or observation D.

H	$P(H)$	$P(D\|H)$	$P(D\|H).P(H)$	$P(H\|D)$
4				
6				
8				
12				
20				

(a) Copy the above table to your answer form and fill in the table with the correct entries (all empty entries should be filled in). In the table H stands for the possible dice (hypothesis), $P(H)$ is the prior probability, $P(D|H)$ the likelihood, and $P(H|D)$, the normalized posterior of H given the data D.
(b) What is the most likely value of H when we know D (i.e. that the outcome is 5)?

25 points **Question 2**

Consider the following piece of email, denoted by e:

```
we have display boxes with credit applications that we need to place in the
small owner-operated stores in your area . here is what you do : 1 . introduce
yourself to the store owner or manager . 2 . use our 90 % effective script
which tells them how this little display box will save their customers hundreds
of dollars , be a drawing card for their business
```

and a vocabulary V given by $V = \{business, display, lifetime, save, text\}$

(a) How will these piece of text e be coded if one uses a bag of word coding.

Now assume that one has the following database with spam and ham emails, only described by the words in V:

Id	words in V	class
1	business, lifetime, save, lifetime, save, text	spam
2	business, display, display, save, save, text	spam
3	business, text, text, save, save	ham
4	display, display, lifetime, text	ham
5	lifetime, save, text	ham
6	business, display, text	ham

(b) Compute the likelihood that the above email is generated by the spam class, i.e. compute $P(e|spam)$. Assume that no smoothing is applied and one uses a multinomial bag of words approach.
(c) Also compute $P(e|ham)$
(d) How will this email e be classified, *spam* or *ham*? **Clearly explain your answer**

25 points **Question 3**

Consider the following dataset with attributes (features) A and B, with A taking the values a_1, a_2 and a_3 and B has values T and F. The class label is given in the last column and is 1 or 0.

Id	A	B	*Class label*
1	a_1	T	1
2	a_2	T	1
3	a_3	T	1
4	a_1	F	1
5	a_2	T	1
6	a_3	F	1
7	a_1	F	0
8	a_2	F	0
9	a_3	F	0
10	a_1	F	0
11	a_2	T	1
12	a_3	T	1
13	a_2	T	0
14	a_2	T	1
15	a_2	T	1

(a) What is the information gain for attribute A and for attribute B?
A table with values for $-p\log_2(p)$ can be found at the end of this exam.
(b) Which attribute will be at the top of the decision tree and why?
(c) Compute and draw the complete decision tree.
(d) What is the error rate on the training set for this decision tree?

15 points **Question 4**

A certain classifier was tested on a test, resulting in the following confusion matrix:

		Predicted class		
		C_1	C_2	C_3
Actual Class	C_1	120	15	18
	C_2	16	150	10
	C_3	11	3	130

(a) What is the accuracy of this classifier?
(b) What is the definition of recall for class C_3 and what is the recall of this classifier for class C_3?
(c) What is the definition of precision for class C_1 and what is the precision of this classifier for class C_1?

15 points **Question 5**

Assume that we are training a linear classifier and that the current linear classifier is given by the line $2+1x_1-2x_2=0$. The next feature point in our training set is given by $x=(-1,1)$.

(a) How will the feature point x be classified, given the current weights $w=(2,1,-2)$ of the linear classifier, 0 or 1?

(b) Assume that the feature point x is misclassified How will the weights of the linear classifier be adapted. Assume a learning rate α of 0.2 and only one iteration of adapting the weights.

(c) How will x be classified after the above adaptation of the weight vectors w? Is this adaptation a step in the right direction? **Motivate your answer!**

Table for $-p\log_2(p)$

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				