

# Herkansingstoets Parel 100 — Intelligente Interactie

5 November 2014

Dit tentamen bestaat uit 5 vragen. Het cijfer voor deze toets is het aantal punten gedeeld door 100.

- ☞ **1 (20 punten)** Voor het vroegtijdig detecteren van de ziekte  $Z$  wordt een test  $T$  gedaan bij een populatie van vogels. De uitkomst van de test kan positief zijn:  $T = pos$ , een indicatie dat de vogel de ziekte  $Z$  heeft, of negatief:  $T = neg$ , er is geen indicatie dat de vogel de ziekte  $Z$  heeft. Het volgende is bekend over het voorkomen van de ziekte onder vogels en de betrouwbaarheid van de test  $T$ .

- $P(Z = true) = 0.01$ , de kans dat een vogel de ziekte  $Z$  heeft.
- $P(T = pos|Z = true) = 0.9$ , de kans dat een vogel die de ziekte  $Z$  heeft positief op de test reageert.
- $P(T = pos|Z = false) = 0.001$ , de kans dat een vogel die ziekte  $Z$  *niet* heeft toch positief op de test reageert.

Bereken de kans  $P(Z = true|T = pos)$ , de kans dat een vogel de ziekte  $Z$  heeft als de uitkomst van de test positief is. (Hint: gebruik Bayes' rule. Gebruik: de wet van de totale kans om  $P(T = pos)$  te berekenen.)

- ☞ **2 (20 punten)** In urn  $H_1$  zitten **10** knikkers: 2 rode, 4 witte en 4 blauwe. In urn  $H_2$  zitten ook **10** knikkers, maar nu: 4 rode, 2 witte en 4 blauwe. Iemand gooit met een eerlijke dobbelsteen en kiest dan urn  $H_1$  als de uitkomst deelbaar is door 3 anders urn  $H_2$ . Vervolgens trekt hij 5 keer (*met teruglegging na iedere trekking*) een knikker uit de geselecteerde urn. De uitkomst  $D$  van de trekking is: 3 rode, 1 witte en 1 blauwe. Welke is de meest waarschijnlijke urn, ofwel: welke kans is groter  $P(H_1|D = \langle 3, 1, 1 \rangle)$  of  $P(H_2|D = \langle 3, 1, 1 \rangle)$ ? Motiveer je antwoord en geef de berekeningen die je hebt gedaan ter ondersteuning van je antwoord. (Let op: er wordt alleen gevraagd welke *de meest waarschijnlijke* urn is waaruit getrokken is gegeven de uitkomst van de trekking en de keuze procedure.)
- ☞ **3 (20 punten)** Beschouw de volgende dataset met attributen (features)  $A$  en  $B$ , beide met waardes  $T$  en  $F$ . Het klasse-label wordt gegeven door  $+$  en  $-$ .

Ex.	A	B	Klasse-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	-

- (a) Wat is de information gain van feature  $A$  en van feature  $B$ ?  
De tabel voor de waardes van  $-p \log(p)$  staat op de andere zijde.
- (b) Welk feature komt in de top van de beslisboom?
- (c) Bereken en schets de gehele beslisboom.

- 4 (20 punten) Beschouw de volgende confusion matrix behorende bij een classifier:

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

- (a) Wat is de accuracy van deze classifier?  
 (b) Wat is recall voor klasse  $C_1$ ?  
 (c) Wat is de precision voor klasse  $C_2$ ?
- 5 (20 punten) Gegeven de classificatielij  $3 - 2x_1 + 2x_2 = 0$  en het featurepunt  $x = (-2, 2)$ .
- (a) Hoe wordt aan de hand van deze classificatielij het punt  $x$  geclassificeerd (1 of 0)?  
 (b) Stel dat het punt  $x$  fout geclassificeerd wordt. Hoe wordt de gewichtenvector  $w = (3, -2, 2)$  aangepast vanwege deze misclassificatie? Neem een learning rate van 0.3.  
 (c) Hoe wordt  $x$  nu geclassificeerd, dat wil zeggen na aanpassing van de gewichtenvector. Is de aanpassing een stap in de goede richting? **Motiveer je antwoord!**

Table for  $-p \log_2(p)$

$p$	$-p \log_2(p)$	$p$	$-p \log_2(p)$
0	0	1/6	0.43
1	0	2/6	0.53
1/2	0.50	3/6	0.50
1/3	0.53	4/6	0.39
2/3	0.39	5/6	0.22
1/4	0.50	1/7	0.40
2/4	0.50	2/7	0.51
3/4	0.31	3/7	0.52
1/5	0.46	4/7	0.46
2/5	0.53	5/7	0.35
3/5	0.44	6/7	0.19
4/5	0.26		