

Dit tentamen duurt 3 uur. Er zijn 20 vragen, waarvan 4 open vragen en 16 meerkeuze.

Het is verboden literatuur, aantekeningen, een programmeerbare rekenmachine, of een telefoon te gebruiken. Het gebruik van een standalone niet-programmeerbare rekenmachine is toegestaan.

Het invullen van de tabellen op het antwoordenblad is verplicht.

Veel succes!

### Open vragen

Elk antwoord op een open vraag dient te zijn voorzien van een uitwerking, dat wil zeggen, van een gemotiveerde berekening. Van opgaven 1 en 4 dient het uiteindelijke antwoord te worden ingevuld op het antwoordenblad.

1. Gegeven is een 1-dimensionaal Kohonen netwerk met buurfunctie

$$g(x, y) = \begin{cases} 1 & \text{als } |x - y| = 0, \\ 0.5 & \text{als } |x - y| = 1, \\ 0 & \text{anders.} \end{cases}$$

Het netwerk ligt in  $R^2$  zoals weergegeven in Fig. 1 in de bijlage.

Bepaal de ligging van dit netwerk na updates met de volgende twee voorbeelden, waarbij de leerfactor,  $\alpha$ , gelijk aan 0.5 is. Schrijf de tussenresultaten van je berekening op in de tabel op je antwoordenblad.

Voorbeeld 1: (0, 3). Voorbeeld 2: (3, 3). De voorbeelden worden aangeboden in deze volgorde. **Antw.**

	A	B	C	D	E	F
0.	(1, 3)	(2, 4)	(3, 3)	(4, 2)	(3, 1)	(2, 2)
Na Voorbeeld 1.	(0.5, 3)	(1.5, 3.75)	(3, 3)	(4, 2)	(3, 1)	(1.5, 2.25)
Na Voorbeeld 2.	(0.5, 3)	(1.875, 3.5625)	(3, 3)	(3.75, 2.25)	(3, 1)	(1.5, 2.25)

2. De grootheid  $\partial E_d / \partial w_{ji}$  staat voor de invloed dat een gewicht  $w_{ji}$  op output-knoop  $j$  heeft op de totale fout  $E_d$  op data-voorbeeld  $d$ . Leid af dat

$$\frac{\partial E_d}{\partial w_{ji}} = -(t_j - o_j) o_j (1 - o_j) x_{ji}.$$

Herinnering: gebruik een aantal malen de kettingregel en het feit dat  $o'(x) = o(x)(1 - o(x))$ .

3. Gegeven is een reinforcement leerprobleem waarvan de transitiegraaf is gegeven in Fig. 2. De robot hanteert een policy

$$\Pi = \{H \rightarrow p, L \rightarrow (q, r)\},$$

waarbij  $p$  de kans is dat de robot met een volle batterij gaat zoeken,  $q$  de kans is dat de robot met een bijna-lege batterij gaat zoeken, en  $r$  de kans is dat de robot met een bijna-lege batterij besluit te gaan laden.

(a) Geef de (recursieve) waardefuncties van  $H$  en  $L$ . Vergeet niet de verdisconteringsfactor  $\gamma$  hierin te betrekken. Je antwoord hoeft niet vereenvoudigd te worden. **Antw.** We schrijven  $H$  voor  $V(H)$  en  $L$  voor  $V(L)$ . Dan

$$\begin{aligned} H &= p(\alpha(\mathcal{R}^{\text{search}} + \gamma H) + (1 - \alpha)(\mathcal{R}^{\text{search}} + \gamma L)) + (1 - p)(\mathcal{R}^{\text{wait}} + \gamma H) \\ L &= q(\beta(\mathcal{R}^{\text{search}} + \gamma L) + (1 - \beta)(-3 + \gamma H)) + (1 - q - r)(\mathcal{R}^{\text{wait}} + \gamma L) \end{aligned}$$

(b) De robot stelt door experimenteren gemiddelde waarden vast voor  $\mathcal{R}^{\text{search}}$ ,  $\mathcal{R}^{\text{wait}}$ ,  $\alpha$  en  $\beta$ , te weten  $\mathcal{R}^{\text{search}} \sim 30$ ,  $\mathcal{R}^{\text{wait}} \sim 2$ ,  $\alpha \sim 0.7$  en  $\beta \sim 0.8$ .

Bepaal  $H$  en  $L$  voor policy  $\Pi = \{H \rightarrow 0.5, L \rightarrow (0.6, 0.2)\}$  met verdiscontering  $\gamma = 0.9$ . **Antw.**

$$\begin{aligned} H &= 0.5(0.7(30 + 0.9H) + 0.3(30 + 0.9L)) + 0.5(2 + 0.9H) \\ L &= 0.6(0.8(30 + 0.9L) + 0.2(-3 + 0.9H)) + 0.2(2 + 0.9L) \end{aligned}$$

Dit is een stelsel van twee vergelijkingen met twee onbekenden. Oplossen geeft  $H = 106.5$  en  $L = 66.8$ .

4. We bekijken een reinforcement leerprobleem met discountfactor  $\gamma = 1.0$  waarvan de mogelijke toestanden en immediate rewards staan afgebeeld in Fig. 3.

Er worden vier runs uitgevoerd, resulterend in Epoch 1 =  $C, D, T_1$ , Epoch 2 =  $C, A, T_2$ , Epoch 3 = Epoch 1 en Epoch 4 =  $C, D, A, T_2$ .

- (a) Bepaal de waarden van alle knopen na Monte-Carlo sampling. **Antw.** Rewards-to-go wordt van achter naar voor uitgerekend. Deze waarden beïnvloeden elkaar niet. De (eind-) waarde van een knoop is het gemiddelde van alle rewards-to-go.

	$A$	$B$	$C$	$D$	$T_1$	$T_2$
0.	0	0	0	0	0	0
Epoch 1.			0	4	0	
Epoch 2.	-1		4			0
Epoch 3.			0	4	0	
Epoch 4.	-1		3	7		0
Eind.	-1		1.8	5	0	0

□

- (b) Bepaal de waarden van alle knopen na temporal difference learning met leerfactor  $\alpha = 0.5$ . **Antw.** Doe voor elke stap van  $i$  naar  $j$  in een epoch: Als  $j$  terminaal:

$$V(i) := V(i) + \alpha (R(i, j) - V(i))$$

Als  $j$  niet terminaal:

$$V(i) := V(i) + \alpha (R(i, j) + \gamma V(j) - V(i))$$

Voor een update gebruiken we dus de waarde van een volgende toestand. Idee: geef elke keer  $V(i)$  een duwtje in de gewenste richting. Updates beïnvloeden elkaar, of het nu in dezelfde run is of tussen runs in. De (eind-) waarde van een knoop is de waarde van de laatste update.

	$A$	$B$	$C$	$D$	$T_1$	$T_2$
0.	0	0	0	0	0	0
Epoch 1.			-2	2	0	
Epoch 2.	-0.5		1.5			0
Epoch 3.			-0.3	3	0	
Epoch 4.	-0.8		-0.6	5.3		0
Eind.	-0.8		-0.6	5.3		0

□

Bij beide berekeningen mag je er van uitgaan dat alle toestanden de beginwaarde nul hebben.

### Meerkeuze vragen

Bij elke meerkeuzevraag is steeds precies één antwoord het juiste. Wel kunnen andere antwoorden “bijna juist” of “deels juist” zijn. Mochten er meerdere goede antwoorden zijn, dan geldt het beste antwoord. Omdat er verschillende versies van de opgaven bestaan, correspondeert de volgorde van de opgaven niet altijd met de volgorde van de stof zoals die behandeld is in de colleges.

1. Geef alle gemixte Nash-equilibria van het spel met payoff matrix

$$\begin{matrix} & \text{L} & \text{R} \\ \text{T} & (-1, -2) & (10, 6) \\ \text{B} & (-3, 7) & (10, -1) \end{matrix}$$

- (a) Hoek (1, 0), inwendig (1/2, 0), rand  $\{0\} \times [0, 1/2]$ .
- (b) Hoek (1, 0), inwendig (0, 1/2), rand  $\{0\} \times [0, 1/2]$ .
- (c) Hoek (0, 1), inwendig (1/2, 0), rand  $[0, 1/2] \times \{0\}$ .
- ✓ Hoek (0, 1), inwendig (0, 1/2), rand  $[0, 1/2] \times \{0\}$ .

2. In het geïtereerde prisoner’s dilemma laten we de volgende strategieën tegen elkaar spelen: ALL-C, ALL-D, RAND, en TFT. We krijgen de volgende matrix

	ALL-C	ALL-D	RAND	TFT
ALL-C	3	0	1.5	3
ALL-D	5	1	3	↓ 1
RAND	4	0.5	2.25	2.25
TFT	?	?	?	?

Vul de laatste rij in.

- (a) 2.25, ↑ 1, 3, 3
- (b) 2.25, 3, ↑ 1, 3
- ✓ 3, ↑ 1, 2.25, 3
- (d) 3, 2.25, ↑ 1, 3

3. De Pavlov strategie kan op tenminste twee manieren worden beschreven.

- i) Als TFT, maar pas vergelden na twee opeenvolgende defects van de tegenstander.
- ii) Werk nu samen als en slechts als zowel jij als de tegenstander in de vorige ronde dezelfde strategie hanteerden.
- iii) Wissel van strategie als de tegenstander in de vorige ronde verzaakte.
- iv) Wissel van strategie als die bij gelijkblijvend gedrag van de tegenstander meer zou opleveren.

- (a) i) en ii).
- ✓ ii) en iii).
- (c) ii) en iv).

(d) iii) en iv).

4. Waar lijkt, volgens Flake, de niet-spatieële evolutionaire variant van het IPD aan?

- (a) Het experiment is afhankelijk van te veel parameters, zoals het aantal strategieën  $K$ , de grootte van de startpopulatie per strategie  $P_i$ ,  $1 \leq i \leq K$ , het aantal ronden  $N$ , en het aantal episoden  $E$ .
- (b) De initiële proportie-keuze van strategieën beïnvloedt de uiteindelijke populatie-proporties. Dat zou niet mogen.
- ✓ Er wordt aangenomen dat elk individu van elke strategie met gelijke kans interacteert met elk individu van elke andere strategie. Deze aanname is onrealistisch.
- (d) De variantie van de gemiddelde opbrengst per episode is te groot, zodat de uitkomsten van een willekeurig uitgevoerd experiment insignificant zullen zijn.

5. Bekijk het volgende binaire classificatie-probleem. Mogelijke leer-instanties:

$$\mathcal{D} = \{(m, n) \in N \times N \mid 0 \leq m \leq 10 \text{ en } 0 \leq n \leq 10\}$$

Mogelijke classificaties: + en -. Hypotheseruimte:

$$\mathcal{H} = \{I \times J \mid I \text{ is een } t_{10}\text{-rij en } J \text{ is een } t_{10}\text{-rij}\}$$

Een  $t_K$ -rij is een niet-lege verzameling van opeenvolgende gehele getallen, waarvan het eerste getal niet-negatief is, en het laatste getal niet groter is dan  $K$ .

Gegeven zijn de leervoorbeelden

$$L = \left\{ \begin{array}{l} (1, 2) : + \quad (1, 5) : + \quad (1, 7) : - \\ (2, 1) : + \quad (2, 3) : + \quad (2, 4) : - \\ (3, 9) : - \quad (5, 6) : + \quad (6, 2) : - \\ (6, 7) : + \quad (7, 4) : + \quad (7, 6) : - \end{array} \right\}$$

Er worden twee hypothesen voorgesteld, te weten

$$\begin{aligned} H_1 &= \{0, \dots, 3\} \times \{0, \dots, 5\} \text{ en} \\ H_2 &= \{4, \dots, 8\} \times \{4, \dots, 7\}. \end{aligned}$$

Bereken van elk van deze hypothesen het bereik (range), de overdekking (coverage), de passing (match), en de nauwkeurigheid (accuracy).

- (a) Range: 15 en 12; coverage: 5 en 4; match: 2/3 en 7/12; accuracy: 4/5 en 3/4.
- (b) Range: 15 en 12; coverage: 5 en 4; match: 4/5 en 3/4; accuracy: 2/3 en 7/12.
- ✓ Range: 24 en 20; coverage: 5 en 4; match: 2/3 en 7/12; accuracy: 4/5 en 3/4.
- (d) Range: 24 en 20; coverage: 5 en 4; match: 4/5 en 3/4; accuracy: 2/3 en 7/12.

6. Bij HGS is er sprake van een zg. *beam size*.

- ✓ Een parameter die bepaalt hoeveel regels er maximaal simultaan in het geheugen kunnen worden gehouden om deze, als daar tijd voor is, verder te specialiseren.
- (b) De minimaal te behalen regel-accuratesse.
- (c) Het minimaal aantal te leren regels.
- (d) Een parameter die bepaalt hoeveel regels er voor elke conclusie er minimaal simultaan dienen te worden geëxploreerd.
7. Welke van de volgende update-methoden zijn gevoelig voor de volgorde waarin data  $\{d_k\}_{k=1}^{\infty}$  wordt aangeboden?
- i) Standaard gemiddelde update:  

$$x = x + k^{-1}(d_k - x)$$
- ii) Standaard geometrische update:  

$$x = x + \alpha(d_k - x), \text{ waarbij leerfactor } 0 \leq \alpha \leq 1.$$
- (a) i).
- ✓ ii).
- (c) i) en ii).
- (d) Geen.
8. We passen kwalitatief gesuperviseerd leren met vector-kwantisatie toe op drie cluster-representanten  $A = (1, 177, 0)$ ,  $B = (2, 177, 5)$  en  $C = (4, 177, 2)$ . Bepaal de bewegingen van deze cluster-representanten als het leervoorbeeld  $l = (2, 177, 3) \rightarrow B$  wordt aangeboden.
- (a) Er gebeurt niets.
- ✓ Repr- $B$  beweegt naar  $l$  toe.
- (c) Repr- $B$  beweegt naar  $l$  toe en Repr- $C$  beweegt van  $l$  af.
- (d) Repr- $B$  beweegt naar  $l$  toe en Repr- $A$  en Repr- $C$  bewegen van  $l$  af.
9. Marie beschikt over 1,000 voorbeelden om haar neurale netwerk te laten leren. Wat is de meest verstandige aanpak?
- (a) Laat maar zitten. Duizend voorbeelden is überhaupt te weinig om een netwerk significant te laten leren.
- (b) Herhaaldelijk laten leren over alle 1,000 voorbeelden. Stoppen als alle netwerkgewichten niet of nauwelijks meer veranderen.
- (c) Herhaaldelijk laten leren op de eerste 900 voorbeelden. De laatste 100 voorbeelden gebruiken als testvoorbeelden. Stoppen met leren als de performance over de testvoorbeelden gaat dalen.
- ✓ Voorbeelden opdelen in tien groepen. Dan tien keer laten leren over negen groepen met telkens één andere groep als contrôle. Telkens stoppen met leren als de performance over de testvoorbeelden gaat dalen. Run vervolgens een elfde keer met het gemiddeld aantal runs van alle vorige tien keren.
- Antw.** Zoals bediscussieerd in Chapt. 4, Artificial Neural Networks, Mitchell. □
10. We leren voorbeelden in  $R^2$  met een standaard één-dimensionaal Kohonen-netwerk ter lengte 100. Het netwerk is random geïnitieerd binnen het cirkeltje om  $(0, 0)$  met straal 0.01. De leervoorbeelden 1 t/m 1,000 bevinden zich uniform verdeeld over de eenheidscirkel  $C$  (oppervlakte nul). De leervoorbeelden 1,001 t/m 20,000 bevinden zich uniform verdeeld over het eenheidsvierkant  $I$  (oppervlakte één). De leerfactor is 0.4.
- We kunnen het volgende zeggen over het resulterende netwerk.
- (a) Het convergeert naar  $C$ .
- (b) Het convergeert naar  $I \setminus C$ .
- ✓ Het verdeelt zich uniform over  $I$ .
- (d) Het goede antwoord staat er niet bij.
11. Welke van de volgende leer-voorbeelden is **niet** lineair separabel?
- (a)  $\{(0, 0) \rightarrow -, (0, 1) \rightarrow -, (1, 0) \rightarrow -, (1, 1) \rightarrow +\}$
- (b)  $\{(0, 0) \rightarrow -, (0, 1) \rightarrow +, (1, 0) \rightarrow +, (1, 1) \rightarrow +\}$
- ✓  $\{(0, 0) \rightarrow -, (0, 1) \rightarrow +, (1, 0) \rightarrow +, (1, 1) \rightarrow -\}$
- (d)  $\{(0, 0) \rightarrow +, (0, 1) \rightarrow -, (1, 0) \rightarrow -, (1, 1) \rightarrow -\}$
12. Een neurale netwerk met veel connecties, en dus veel gewichten, zal in het algemeen minder snel in een lokaal minimum terechtkomen.
- ✓ Dat is waar, want meer connecties zorgen voor meer vrijheidsgraden om uit lokale minima te ontsnappen.
- (b) Dat is waar, want meer connecties zorgen voor een kleinere som-term in de fout-functie.
- (c) Dat is niet waar, want in netwerken met veel gewichten zijn er gewichten-verdelingen te bedenken waarbij er toch convergentie is naar een oneigenlijk minimum.
- (d) Dat is niet waar, want door een groot aantal gewichten kan een netwerk juist veel sneller naar een oneigenlijk minimum worden getrokken.
13. Waar word de hypotheseruimte in een neurale netwerk door vertegenwoordigt?
- ✓ Door de gewichten.
- (b) Door de netwerk-topologie, d.w.z., de knopen en de links.
- (c) Door de topologie en de gewichten.

- (d) Door de topologie, en de respectievelijke leervoorbeelden.

**Antw.** Alleen de gewichten, de rest ligt tijdens het leren vast.

14. In reinforcement leren is policy evaluation een techniek om via iteratie grootheden aan te passen. Daarbij gelden de volgende beweringen.

i)  $V$  verandert.

ii)  $Q$  verandert.

iii)  $\Pi$  verandert.

(a) i), ii) en iii).

i), ii) en niet iii).

(c) i), niet ii) en iii).

(d) i), niet ii) en niet iii).

15. Benoem het verschil tussen value iteration en temporal difference learning.

(a) Bij value iteration worden alle values van alle toestanden ge-update, bij temporal difference learning worden alle policies van alle toestanden ge-update.

(b) Bij value iteration worden alle values van alle toestanden in een epoch ge-update, bij temporal difference learning worden alle policies van alle toestanden in een epoch ge-update.

Bij value iteration worden alle states ge-update, bij temporal difference learning alleen de states in een epoch.

(d) Bij value iteration worden alle states in een epoch ge-update, bij temporal difference learning alle states.

16. Welke van de volgende zijn **geen** (versies van) lerende classificatiesystemen?

i) LCS

ii) ZCS

iii) XCS

iv) XCSR

v) XCS Redux

(a) iii).

(b) iv).

v).

- (d) Het goede antwoord staat er niet bij.

17. Waarom werd ZCS geïntroduceerd?

i) Om de werking van classificatie-systemen beter te begrijpen.

ii) Als uniform raamwerk voor classificatie-systemen.

iii) Om classificatie-systemen efficiënter te laten werken.

(a) i) en ii).

(b) ii) en iii).

i) en iii).

(d) Het goede antwoord staat er niet bij.

18. Waar staat de notie *bucket brigade* voor?

(a) Een korte heftige opeenvolging ("burst") van updates in een regel niche.

Winnende classifiers sluizen een deel van de verkregen beloning (reward) door naar classifiers die eerder actief waren.

(c) Een buffer-mechanisme voor message lists in classificatie-systemen.

(d) Een internationaal consortium van onderzoekers en instituten dat zorgdraagt voor de promotie en instandhouding van lerende classificatie-systemen.

19. We beschouwen ZCS. In welke situaties treedt covering op?

(a) Als er geen regels matchen met de data.

(b) Als er te weinig regels matchen met de data.

(c) Als de gemiddelde regelsterkte van de match-set te klein wordt.

Als de gemiddelde regelsterkte van de match-set te klein wordt ten opzichte van de gemiddelde regelsterkte van de gehele regelset.

20. ZCS lijdt aan een aantal problemen.

i) Path habits.

ii) Panmictic crossover.

(a) ZCS heeft zeker problemen, maar daar behoren i) en ii) niet toe.

(b) i).

(c) ii).

i) en ii).

Einde van de meerkeuzevragen.

## Figuren

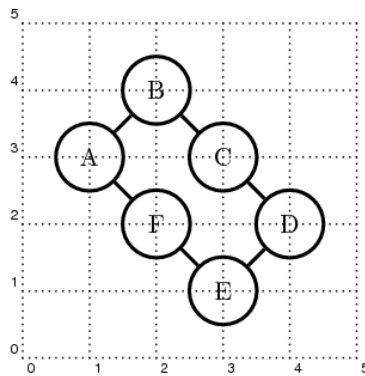


Fig. 1: Kohonen netwerk.

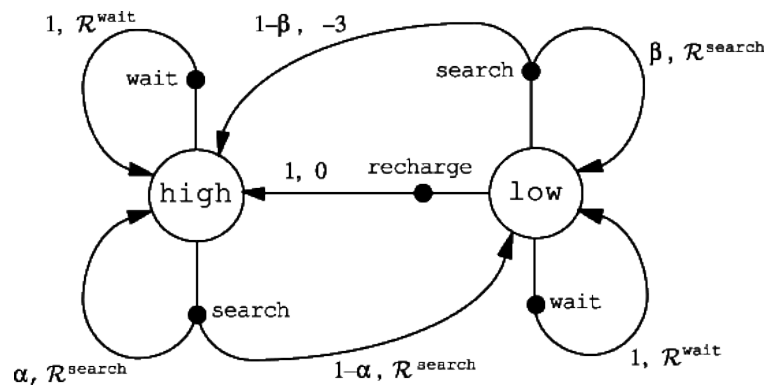


Fig. 2: Transitiegraaf voor de blikjesrobot.

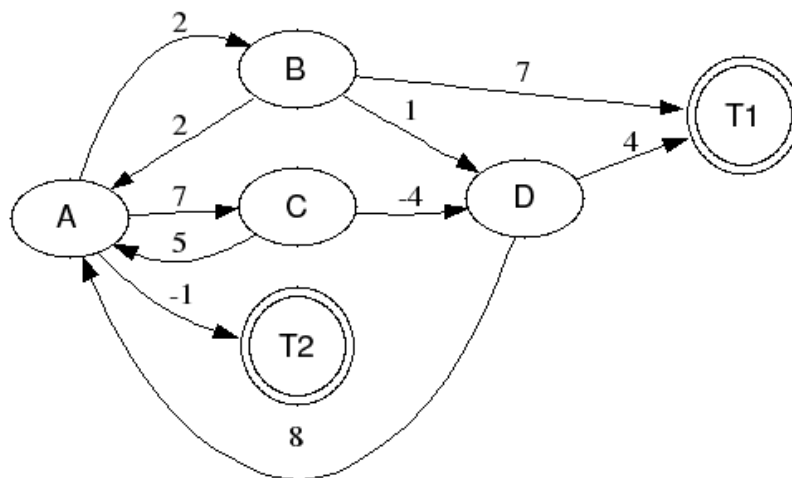


Fig. 3: Toestandengraaf.