ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ	
Θεματική Ενότητα	**ΠΛΗ 31: Τεχνητή Νοημοσύνη - Εφαρμογές**
Ακαδημαϊκό Έτος	**2021 – 2022**
Γραπτή Εργασία	**#3**
Ημερομηνία Παράδοσης	**Παρασκευή, 1.4.2022, ώρα 23:59:59**

Παρατηρήσεις

Η καταληκτική ημερομηνία υποβολής της εργασίας είναι η *Τετάρτη, 6 Απριλίου 2022, ώρα 23:59:59*, μετά από την οποία οποιαδήποτε υποβολή εργασίας θεωρείται ως εκπρόθεσμη.

Η παραπάνω ημερομηνία συνυπολογίζει την καθολική προκαταβολική έγκριση τυχόν αιτήματος παράτασης, κατά τον κανονισμό του ΕΑΠ (https://www.eap.gr/wp-content/uploads/2020/11/eap_kanonismos_spoydwn.pdf), με την υπόθεση πως η αρχικά οριζόμενη ημερομηνία παράδοσης θα ήταν η ακριβώς προηγούμενη Παρασκευή.

Την *Παρασκευή, 8 Απριλίου 2022, ώρα 11:59:59* (το αργότερο), θα δημοσιευθεί ενδεικτική επίλυση της εργασίας στο διαδίκτυο.

Περιμένουμε όλες οι εργασίες να υποβληθούν μέσω του συστήματος study. Πρέπει να υποβάλετε ΕΝΑ μόνο αρχείο (συμπιεσμένο), στο οποίο θα περιλαμβάνονται το πρωτότυπο κείμενο, σε οποιοδήποτε επεξεργαστή κειμένου έχετε χρησιμοποιήσει, και το παραγόμενο PDF, καθώς και όποια άλλα συνοδευτικά αρχεία. Αν υποβάλετε τμήματα κώδικα, θα πρέπει να βρίσκονται σε ξεχωριστά αρχεία, μέσα στο συμπιεσμένο αρχείο, και θα πρέπει ν' αναφέρονται στο κείμενο της εργασίας.

Αν υπάρχουν διευκρινιστικά ερωτήματα που θεωρείτε ότι είναι αναγκαία για την επίλυση της εργασίας και τα οποία δεν έχουν απαντηθεί, μπορείτε στην επίλυσή σας να τα αναφέρετε μαζί με τις παραδοχές που θα κάνετε και τις οποίες θα πρέπει να ακολουθήσετε στην επίλυσή σας.

ΘΕΜΑ		ΜΟΝΑΔΕΣ
	ΠΕ3.Α	10
	ΠΕ3.Β	10
1.	**Βασικά Ζητήματα Γενετικών Αλγορίθμων**	10 (2 + 4 + 4)
2.	**Συνάρτηση Αξιολόγησης σε Γενετικούς/Εξελικτικούς Αλγορίθμους**	15 (3 + 4+4+4)
3.	**Θέματα Θεμελίωσης Γενετικών Αλγορίθμων**	25 (1+2+2 + 5+5+5+5)
4.	**Εύρεση Στρατηγικής με Χρήση Γενετικού Προγραμματισμού**	30 (10 + 10 + 10)
	ΣΥΝΟΛΟ	**100**

Θέμα 1: Βασικά Ζητήματα Γενετικών Αλγορίθμων [*10 μονάδες*]

Α. [*2 μονάδες*] Ποια είναι τα πλεονεκτήματα και τα μειονεκτήματα χρήσης της Επιλογής Ανάλογα με την Απόδοση (Fitness Proportional Selection - FPS) σε σχέση σε σχέση με την Επιλογή Βασισμένη σε Βαθμολόγηση (Rank Based Selection). Πως συγκρίνεται η Επιλογή Τουρνουά (Tournament Selection) σε σχέση με τις άλλες δύο; Στις απαντήσεις σας να δώσετε έμφαση στο πως αυτές οι επιλογές επιδρούν στο ΓΑ για την εξερεύνηση του χώρου αναζήτησης.

(Υπόδειξη: μπορείτε να συμβουλευτείτε και εξωτερικές πηγές για να αναζητήσετε πληροφορίες για τις παραπάνω τεχνικές, όπως η https://www.obitko.com/tutorials/genetic-algorithms/selection.php.)

Απάντηση:
πηγή που χρησιμοποίησα:
https://en.wikipedia.org/wiki/Selection_(genetic_algorithm)#Roulette_Wheel_Selection
επιπλέον πηγή ανάγνωσης που χρησιμοποίησα:
https://www.tutorialspoint.com/genetic_algorithms/genetic_algorithms_parent_selection.htm

Η επιλογή της χρήσης εξαναγκασμένης ρουλέτας ως πιο σύνηθες αλγόριθμος επιλογής διότι πλεονέκτημα είναι ότι η επιλογή γίνεται με βάση απόδοσης του κάθε ατόμου. Το μειονέκτημα είναι ότι αν ένα άτομο έχει μεγάλη απόδοση και πάρει μεγαλύτερη μερίδα στην ρουλέτα θα επιλέγεται τις περισσότερες φορές το συγκεκριμένο άτομο παρά τα υπόλοιπα άτομα. Συγκριτικά με τον αλγόριθμο επιλογής Rankes Based Selection λειτουργεί και με αρνητική απόδοση διότι επιλέγεται με βάση την τάξη του (Rank) χρησιμοποιείται όταν έχουν οι πληθυσμοί σχεδόν ισόποση απόδοση ώστε να έχουν ισόποση κατανομή στην πίτα της επιλογής.

Στην χρήση του αλγορίθμου επιλογής τουρνουά μπορεί υπάρχουν εκδοχές όπως ο κ-τρόπος: επιλογή τυχαίων και ομοιόμορφων k άτομα του πληθυσμού και επιλογή βάση απόδοσης καλύτερου ατόμου. Αν συγκρίνουμε με βάση της χαμηλής απόδοσης ο αλγόριθμος επιλογής τουρνουά γενικά κάνει επιλογή ακόμα και άτομα με πάρα πολύ χαμηλή απόδοση.

Η επίδραση στον ΓΑ για την εξερεύνηση του χώρου αναζήτησης βάση του αλγόριθμου επιλογής είναι ότι: όσο ποιο ιδανική είναι η χρήση του αλγορίθμου ανάλογα με το πρόβλημα, τόσο και θα γίνει πιο σωστή και πιο ποικιλία στην ανακατανομή του πληθυσμού.

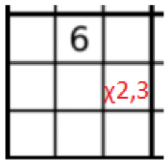
B. [*8 μονάδες*] Το Sudoku puzzle (ένα παράδειγμα φαίνεται παρακάτω) αποτελείται από ένα πλέγμα 9 x 9 χωρίζεται σε τμήματα 3 x 3 (blocks) με σύνολο 81 κελιά. Κάθε puzzle, που έχει μια μοναδική λύση, έχει μερικά κελιά προ-συμπληρωμένα. Στόχος του puzzle είναι να συμπληρωθούν τα υπόλοιπα κελιά με αριθμούς από 1 μέχρι 9, έτσι ώστε να ικανοποιούνται οι παρακάτω περιορισμοί:

- Κάθε οριζόντια γραμμή πρέπει να περιέχει αριθμούς από 1 μέχρι 9, χωρίς επαναλήψεις.
- Κάθε κάθετη στήλη πρέπει να περιέχει αριθμούς από 1 μέχρι 9, χωρίς επαναλήψεις.
- Κάθε 3 x 3 τμήμα πρέπει να περιέχει αριθμούς από 1 μέχρι 9, χωρίς επαναλήψεις.

5	3			7				
6			1	9	5			
	9	8					6	
8				6				3
4			8		3			1
7				2				6
	6					2	8	
			4	1	9			5
				8			7	9

B.1. [*4 μονάδες*] Αν θέλουμε να λύσουμε το puzzle με χρήση ΓΑ, πως θα αναπαραστήσετε ένα χρωμόσωμα του προβλήματος; Θα προτιμούσατε την αναπαράσταση με ένα 2-διάστατο πίνακα;

Απάντηση:
Για την γραμμική αναπαράσταση χρωμοσώματος του προβλήματος Sudoku puzzle, όπου για κάθε κενό κελί θα το αντιστοιχήσω με το σύμβολο x το οποίο θα έχει ως δείκτη i, j όπου είναι οι συντεταγμένες του πίνακα Sudoku (i = γραμμή και j = στήλη του πίνακα). Παράδειγμα αν υποθέσουμε έναν αριθμό σε ένα κελί όπου βρίσκεται στη δεύτερη γραμμή και στην τρίτη στήλη τότε θα είναι $x_{2,3}$



Για αναπαράσταση χρωμοσώματος προβλήματος Sudoku puzzle, με δυσδιάστατο πίνακα θα προτιμούσα μια πιο συνδυαστική προσέγγιση της αναπαράστασης όπως αντίστοιχα του προβλήματος knapsack 0/1 όπου η κάθε διακριτή τιμή x του πίνακα Sudoku θα έχει επιπλέον ως βάρος 1 bit για δύο κωδικοποιημένες καταστάσεις $\log_2(2)=1$. Την κατάσταση συντεταγμένης της στήλης και την κατάσταση της της γραμμής.
Επίσης Θα χωρίσω τον τετραγωνικό πίνακα σε υποσύνολα τετραγωνικών πινάκων όπου θα προστεθεί ως επιπλέον βάρος τα bit αρίθμησης του πίνακα ανάλογα με τον όγκο του πίνακα. Παράδειγμα στο συγκεκριμένο πρόβλημα ο πίνακας χωρίζεται σε 9 υπό πίνακες όπου $\log_2(9)=\lceil 3.16 \rceil = 4bits$
Άρα 0001 για την πρώτη θέση του υπό πίνακα, 0002 για τη δεύτερη και ούτω καθ' εξής μέχρι την 9 η θέση του υπό πίνακα 1001

Θα προτιμήσω την γραμμική κωδικοποίηση παρά τον 2-διάστατο πίνακα διότι θεωρώ ότι ο 2-διάστατος πίνακας χρειάζεται περισσότερους πόρους συστήματος για τον υπολογισμό του ολικού βέλτιστου. Όπως και είναι πιο πιθανό λόγω σύνθετης συνάρτησης αξιολόγησης να βρεθεί και να κλειστεί σε ένα τοπικό βέλτιστο.

Β.2. [*4 μονάδες*] Ποια συνάρτηση αξιολόγησης θα χρησιμοποιήσετε;

Απάντηση:
Για τη γραμμική αναπαράσταση χρωμοσώματος του προβλήματος Sudoku puzzle θα χρησιμοποιήσω Fitness = 1/(1+ERROR). (πηγή: αρχείο ΟΣΣ <<3b EvolutionaryNeuralNets>>).
Όπου το error είναι μια συνάρτηση λάθους στην έξοδο του δικτύου και πρόκειται για ελαχιστοποίηση του λάθους.
Το λάθος στο συγκεκριμένο πρόβλημα είναι: βάση των κανόνων των περιορισμών του Sudoku puzzle όταν ένας αριθμός που δημιουργείται από τον αλγόριθμο υπάρχει ήδη τότε θεωρείται λάθος.

Παράδειγμα:
Όταν σε μία στήλη έχουμε ήδη {1,2,5} και ο αλγόριθμος δημιουργεί {5,7} τότε το 5 υπάρχει ήδη στην αρχική στήλη.

Θέμα 2: Συνάρτηση Αξιολόγησης στους Γενετικούς/Εξελικτικούς Αλγορίθμους [*15 μονάδες*]

Α. [*3 μονάδες*] Διαθέτουμε 10 κάρτες αριθμημένες από το 1 έως το 10. Θέλουμε να τις χωρίσουμε σε 2 στήλες, ώστε το άθροισμα των αριθμών στις κάρτες της 1ης στήλης να είναι όσο το δυνατόν πιό κοντά στον αριθμό **a** και το γινόμενο των αριθμών στις κάρτες της 2ης στήλης να είναι όσο το δυνατόν πιό κοντά στον αριθμό **b**.
Περιγράψτε το ΓΑ που θα προτείνατε για την επίλυση του προβλήματος, δίνοντας:

- τη μορφή της κωδικοποίησης των λύσεων (περιγραφικά και με παραδείγματα)
- την αντικειμενική συνάρτηση που θα χρησιμοποιούσατε
- τους τελεστές διασταύρωσης και μετάλλαξης που θα χρησιμοποιούσατε

Σημείωση: θεωρείστε ότι στο περιβάλλον που θα υλοποιηθεί ο προτεινόμενος ΓΑ η αναζήτηση γίνεται προς την ελάχιστη τιμή της αντικειμενικής συνάρτησης

Απάντηση:
Επειδή οι κάρτες είναι αριθμημένες με διακριτές τιμές στις οποίες δεν θα επαναλαμβάνονται, η μορφή της κωδικοποίησης θα είναι ένας δυσδιάστατος πίνακας όπου και θα πραγματοποιηθεί ακέραια κωδικοποίηση αριθμών στην οποία θα γίνεται παραγωγή μετάθεσης 1..N όπου N=10.

Ο πίνακας θα έχει δύο γραμμές όπου θα κωδικοποιηθεί με δύο bits το 0 bit για το αθροιστικό αποτέλεσμα a όπου αποτελεί την μια γραμμή του πίνακα και 1 για το αποτέλεσμα γινομένου b όπου αποτελεί τη δεύτερη γραμμή του πίνακα.

Παράδειγμα για a = 0 όπου θα περιέχει τις κάρτες {1,2,3,4,5} και b = 1 όπου θα περιέχει τις κάρτες{6,7,8,9,10}.
Ουσιαστικά υιοθετώ την σκέψη μου γύρω από το πρόβλημα knapsack όπου ανάγνωσα από το αρχείο μοντελοποίησης της ΟΣΣ όπου θα βάλω στους αριθμούς βάρος 0 ή 1, όπου 0 για το άθροισμα a και 1 για το γινόμενο b. Πρόκειται για ένα πρόβλημα ελαχιστοποίησης.

Κατά την εξέλιξη του γενετικού αλγορίθμου, το υπολογισμένο άθροισμα από τους επιλεγμένους αριθμούς απέχει από το ζητούμενο άθροισμα κατά κάποια ποσότητα, που πρέπει να μηδενιστεί.

Άρα όπως και αναφέρεται και από τα αρχεία της ΟΣΣ στο θέμα μοντελοποίηση με γενετικούς αλγορίθμους: ένα αντίστοιχο πρόβλημα ελαχιστοποίησης (subset sum).
a – (διακριτό πλήθος του πίνακα αθροίσματος) και (b- διακριτό πλήθος γινομένου)

Επειδή γίνεται ως προς την ελάχιστη τιμή της αντικειμενικής συνάρτησης τότε θα ισχύει f(x) = -f(x)+(c) όπου θα προστεθεί μια σταθερά ώστε να είναι η συνάρτηση μεγαλύτερη ή ίση με το μηδέν διότι θα χρησιμοποιηθεί η επιλογή με βάση την εξαναγκασμένη ρουλέτα.

Ο τελεστής διασταύρωσης που θα χρησιμοποιηθεί στον ΓΑ θα είναι ακέραιης κωδικοποίησης OX αντίστοιχο με αυτό που χρησιμοποιείται στο πρόβλημα του πλανόδιου πωλητή ώστε να αποφευχθεί το πρόβλημα των μεταθέσεων και να μην υπάρχουν διπλότυπα. Δηλαδή να μην υπάρχει επανάληψη στις κάρτες της ίδιας γενιάς.

Ο τελεστής μετάλλαξης που θα χρησιμοποιηθεί θα είναι επίσης Ακέραιας κωδικοποίησης όπου θα γίνεται ανταλλαγή θέσεων μεταξύ δύο αριθμών βάση πιθανότητας μετάλλαξης.

Β. [*12 μονάδες*] Δίνεται ο αρχικός πληθυσμός, στην 1η στήλη στον παρακάτω πίνακα και οι αντίστοιχες καταλληλότητες f(x), στη 2η στήλη. Υποθέστε ότι το ζητούμενο είναι η μεγιστοποίηση της καταλληλότητας. Παρατηρείστε πως, επειδή η καταλληλότητα περιέχει αρνητικές τιμές, δεν μπορούν να υπολογιστούν οι συσσωρευμένες πιθανότητες.

Αρχικός Πληθυσμός	Καταλληλότητα	Καταλληλότητα Β.1	Καταλληλότητα Β.2	Καταλληλότητα Β.3
100010111	4	14	15	104
100000001	-10	0	1	90
010101010	-6	4	5	96
010100110	1	11	12	101
001100111	-5	5	6	95
110110110	9	19	20	109
Συνολική Καταλληλότητα	-7	53	59	595

Β.1. [*4 μονάδες*] Πώς μπορεί να αντιμετωπιστεί αυτό το πρόβλημα; Να αιτιολογήσετε την απάντησή σας και να συμπληρώσετε την 3η στήλη του πίνακα.

<u>Απάντηση:</u>

Επειδή πρόκειται για ένα πρόβλημα μεγιστοποίησης καταλληλόλητας από τη θεωρία του βιβλίου στη σελίδα 62 γνωρίζω ότι η αντικειμενική συνάρτηση παίρνει μόνο θετικές τιμές. max g(x) = max{f(x)+C} όπου C είναι μια θετική σταθερά τιμή. Δεν δύναται η δυνατότητα επιλογής να είναι αρνητικές τιμές διότι δεν θα επιλεχτούν τα άτομα του συγκεκριμένου πληθυσμού κατά τον υπολογισμό της πιθανότητα επιλογής κάθε μέλους. $p_i = \frac{eval(v_i)}{F}$

Αν χρησιμοποιήσω ως θετική σταθερά το μέγιστο των αρνητικών τιμών της της καταλληλόλητας του πληθυσμού την πρώτης στήλης f(x) όπου είναι το -10 και το θέσω ως σταθερά θετικής τιμή δηλαδή 10 τότε f(x)+C ο πίνακας καταλληλόλητας Β.1 θα συμπληρωθεί ως εξής:

Καταλληλότητα	Καταλληλότητα Β.1
4	f(x)+C = 10 + 4 =14
-10	f(x)+C = 10 + (-10) =0
-6	f(x)+C = 10 + (-6) =4
1	f(x)+C = 10 + 1 =11
-5	f(x)+C = 10 + (-5) =5
9	f(x)+C = 10 + 9 =19

B.2. [*4 μονάδες*] Υπάρχει περίπτωση κάποιο άτομο να συσχετιστεί με μηδενική πιθανότητα επιλογής; Πως μπορεί να αντιμετωπιστεί το πρόβλημα αυτό; Αποτυπώστε τυχόν αλλαγές στην 4η στήλη του πίνακα.

Απάντηση:

Στην περίπτωση του υπό ερωτήματος Β.1 στη δεύτερη γραμμή το άτομο 100000001 αξιολόγηση καταλληλόλητας ίση με το 0 που σημαίνει ότι κατά τον υπολογισμό της πιθανότητας της επιλογής του μέλους θα είναι 0. Άρα δεν θα επιλεχτεί. Άρα δεν πρόκειται να συσχετιστεί κάποιο άτομο με μηδενική πιθανότητα επιλογής για την αντιμετώπιση του προβλήματος αυτού θα υπολογιστεί όπως και έγινε ο υπολογισμός σταθερής τιμής C του υπό ερωτήματος Β.1 δηλαδή θα επιλεχτεί το μέγιστο αρνητικής τιμής και θα προστεθεί +1. Άρα $F(x) = f(x) + (C+1)$

Καταλληλόλητα	Καταλληλόλητα B.3
4	f(x)+(C+1) = 10 + 4 = 14
-10	f(x)+(C+1) = 10 + (-10) = 0
-6	f(x)+(C+1) = 10 + (-6) = 4
1	f(x)+(C+1) = 10 + 1 = 11
-5	f(x)+(C+1) = 10 + (-5) = 5
9	f(x)+(C+1) = 10 + 9 = 19

B.3. [*4 μονάδες*] Δίνεται η συνάρτηση μετασχηματισμού της καταλληλότητας: g’(x)= f(x) + (max(|f(x)|)^2. Αποτυπώστε τις τιμές της στην 5η στήλη του πίνακα. Μπορεί να χρησιμοποιηθεί για το μετασχηματισμό της καταλληλότητας; Αν ναι, ποιον από τους 3 μετασχηματισμούς θα επιλέγατε και γιατί;

Απάντηση

Ως μέγιστο της απόλυτη τιμής του f(x) = -10 άρα: max(|f(x)|)^2 = max(|10)|)^2 = 100

Καταλληλότητα	Καταλληλότητα B.3
4	f(x)+ max(|f(x)|)^2 = 4 + 100 = 104
-10	f(x)+ max(|f(x)|)^2 = (-10) + 100 = 90
-6	f(x)+ max(|f(x)|)^2 = (-6) + 100 = 96
1	f(x)+ max(|f(x)|)^2 = 1 + 100 = 101
-5	f(x)+ max(|f(x)|)^2 = (-5) + 100 = 95
9	f(x)+ max(|f(x)|)^2 = 9 + 100 = 109

στο επόμενο βήμα του αλγορίθμου της επιλογής όπου και γίνεται η πιθανότητα επιλογής κάθε μέλους, βάση αυτού εφόσον θεωρώ ότι χρησιμοποιείται η εξαναγκασμένη ρουλέτα με χρήση τον υπολογισμό αθροιστικής πιθανότητας γίνεται ο επιμερισμός της κάθε μερίδας στη ρουλέτα.

Άρα για την Καταλληλότητα Β2 με συνολική καταλληλότητα F = 59 έχουμε:

Καταλληλότητα B.2	$p_i = \frac{eval(v_i)}{F}$	$q_i = \sum_{j=1}^{i} p_j$
15	$p_1 = \frac{15}{59} = 0.254$	q_1=0.254

1	$p2=\frac{1}{59}=0.016$	q2=0.27
5	$p_3=\frac{5}{59}=0.084$	q3=0.354
12	$p_3=\frac{12}{59}=0.203$	q4=0.557
6	$p_4=\frac{6}{59}=0.101$	q5=0.658
20	$p_6=\frac{20}{59}=0.338$	$q_6\simeq 1$

Άρα για την Καταλληλότητα Β3 με συνολική καταλληλότητα F = 595 έχουμε:

Καταλληλότητα Β.3	$p_i=\frac{eval(v_i)}{F}$	$q_i=\sum_{j=1}^{i} p_j$
104	$p_1=\frac{104}{595}=0.174$	q1=0.174
90	$p_2=\frac{90}{595}=0.151$	q2=0.325
96	$p_3=\frac{96}{595}=0.161$	q3=0.486
101	$p_4=\frac{101}{595}=0.169$	q4=0.655
95	$p_5=\frac{95}{595}=0.159$	q5=0.814
109	$p_6=\frac{109}{595}=0.183$	$q_6\simeq 1$

Παρατηρώντας τον υπολογισμό της πιθανότητας επιλογής παρατηρώ ότι όσο μεγάλη είναι η σταθερά τόσο και ποιο ισόποση είναι η κατανομή της εξαναγκασμένης ρουλέτας τόσο όσο που πλησιάζει εκδοχή της ισόποσης επιλογής τουρνουά.
Θα επέλεγα τον μετασχηματισμό Β2 διότι έτσι το μέγεθος της εξαναγκασμένης ρουλέτας με σχισμές η κάθε σχισμή είναι ανάλογη με την απόδοση του κάθε ατόμου.

Θέμα 3: Θέματα Θεμελίωσης Γενετικών Αλγορίθμων [*25 μονάδες*]

Α. [*5 μονάδες*] Στον παρακάτω πίνακα δίνεται ο παρακάτω αρχικός πληθυσμός και η εξέλιξή του κατά μία γενιά. Επίσης δίνονται οι αντίστοιχες καταλληλότητες.

Γενιά #0	Καταλ/λότητα	Γενιά #1	Καταλ/ότητα
1 1 0 1 1	2	1 1 0 1 1	2
0 1 0 1 1	1	0 0 1 1 1	3
1 1 0 0 1	3	1 1 0 1 1	2
1 0 1 1 1	4	1 0 1 0 1	5

Ζητείται να ιχνηλατήσετε την εξέλιξη των δύο σχημάτων 1###1 and ##01#, δείχνοντας για τις γενιές #0 και #1 τα εξής:
Α.1. [*1 μονάδα*] Τα άτομα στον πληθυσμό που ταιριάζουν σε κάθε σχήμα.

Απάντηση:
Τα άτομα του πληθυσμού για την γενιά #0 που ταιριάζουν στο σχήμα 1###1 είναι τρία: {11011,11001,10111}.
Τα άτομα του πληθυσμού για την γενιά #1 δεν έχει υποστεί αλλαγή και αυτά που ταιριάζουν στο σχήμα 1###1 είναι τρία: {11011,11011,10101}.

Τα άτομα του πληθυσμού για την γενιά #0 που ταιριάζουν στο σχήμα ##01# είναι δύο: {11011,01011}.
Τα άτομα του πληθυσμού για την γενιά #1 δεν έχει υποστεί αλλαγή και αυτά που ταιριάζουν στο σχήμα ##01# είναι δύο: {11011,11011}.

Α.2. [*2 μονάδες*] Τη μέση απόδοση κάθε σχήματος.

Απάντηση:
Η μέση απόδοση κάθε σχήματος που ταιριάζουν σε ένα σχήμα S κατά μια χρονική στιγμή t υπολογίζεται με:

$$eval(S,t) = (\sum_{j=1}^{p} eval(v_{i_j}))/\ p\ .$$

άρα:

Για την γενιά #0 και το σχήμα 1###1 $eval(S,t) = \dfrac{(2+3+4)}{3} = 3$

Για την γενιά #1 και το σχήμα 1###1 $eval(S,t) = \dfrac{(2+2+5)}{3} = 3$

Για την γενιά #0 και το σχήμα ##01# $eval(S,t) = \dfrac{(2+1)}{2} = 1.5$

Για την γενιά #1 και το σχήμα ##01# $eval(S,t) = \dfrac{(2+2)}{2} = 2$

Α.3. [*2 μονάδες*] Συμφωνεί αυτή η ιχνηλάτηση με το θεώρημα των σχημάτων ή όχι; Να αιτιολογήσετε σύντομα την απάντησή σας.

Απάντηση:
Από το θεώρημα των σχημάτων γνωρίζουμε ότι όταν το δ(s) είναι μικρό τότε έχουμε μικρή πιθανότητα καταστροφής.

Για το σχήμα 1###1 έχουμε δ(s) = 4-1 = 3 με όμοια μέση απόδοση

ξ(S,#0) = 3
eval(S,0) = (2+3+4) / 3 = 3
F(t) = (2+1+3+4) / 4 = 2.5
ξ(S,t+1) = ξ(S,t) * eval(S,t)/F(t) = (3*3) / 2.5 = 3.6

ξ(S,#1) = 3
eval(S,0) = (2+2+5) / 3 = 3
F(t) = (2+3+2+5) / 4 = 3
ξ(S,t+1) = ξ(S,t) * eval(S,t)/F(t) = (3*3) / 3 = 3

όχι δεν συμφωνεί η ιχνηλάτιση με το θεώρημα των σχημάτων παρότι το δ(s) είναι μεγάλο και υπάρχει μεγάλη πιθανότητα καταστροφής, η τάξη o(s) είναι μεγάλη αλλά και ότι το ξ(S,t+1) είναι πάνω από το μέσο όρο eval(S,0) όπου σημαίνει ότι αυξάνεται η πιθανότητα επιλογής. Στο συγκεκριμένο παράδειγμα και οι δύο γενεές έχουν το ίδιο αποτέλεσμα. Παρατηρώ όμως ότι το ξ(S,t+1) μειώθηκε και βάση του θεωρήματος καταλληλότητας μικρότερη της μέσης μειώνουν την συμμετοχή τους στον πληθυσμό με το πέρασμα τον επόμενων γενεών.

Για το συγκεκριμένο παράδειγμα θα πρέπει να λάβουμε υπόψιν την πιθανότητα διασταύρωσης όπως και την πιθανότητα μετάλλαξης βάση του θεωρήματος σχημάτων για καλύτερα αποτελέσματα.

$$\xi(S,t+1) \geq \xi(S,t) \cdot eval(S,t) / \overline{F(t)} \left[1 - p_c \cdot \frac{\delta(S)}{m-1} - o(S) \cdot p_m \right]$$

Για το σχήμα ##01# έχουμε δ(s) = 4-3 = 1

ξ(S,#0) = 2
eval(S,0) = (2+1) / 2 = 1.5
F(t) = (2+1+3+4) / 4 = 2.5
ξ(S,t+1) = ξ(S,t) * eval(S,t)/F(t) = (2*1.5) / 2.5 = 1.2

ξ(S,#1) = 2
eval(S,0) = (2+2) / 2 = 2
F(t) = (2+3+2+5) / 4 = 3
ξ(S,t+1) = ξ(S,t) * eval(S,t)/F(t) = (2*2) / 3 = 1.33

όχι δεν συμφωνεί η ιχνηλάτιση με το θεώρημα των σχημάτων παρότι το δ(s) υπάρχει μικρή πιθανότητα καταστροφής η τάξη είναι μικρή αλλά και ότι το ξ(S,t+1) είναι κάτω από το μέσο όρο eval(S,0) όπου αυξάνεται εκθετικά η συμμετοχή τους στον πληθυσμό με την πάροδο του χρόνου. Βλέπω ότι στην γενεά #1 η καταλληλότητα δεν είναι μεγαλύτερη της μέσης αλλά μικραίνει που σημαίνει ότι ενδέχεται να καταστραφεί.

Για το συγκεκριμένο παράδειγμα πρέπει να λάβουμε υπόψιν την πιθανότητα διασταύρωσης όπως και την πιθανότητα μετάλλαξης βάση του θεωρήματος σχημάτων για καλύτερα αποτελέσματα.

$$\xi(S,t+1) \geq \xi(S,t) \cdot eval(S,t) / \overline{F(t)} \left[1 - p_c \cdot \frac{\delta(S)}{m-1} - o(S) \cdot p_m \right]$$

Β. [*20 μονάδες*] Υποθέστε ένα Γενετικό Αλγόριθμο (ΓΑ) με χρωμόσωμα μήκους n=6 bits και την ακόλουθη συνάρτηση καταλληλότητας:

$$F(C) = \sum_{i=0}^{n-2} g(C[i], C[i+1])$$

Όπου:

$$g(X,Y) = \begin{cases} 1 \; \alpha\nu \; X \neq Y \\ 0 \; \alpha\nu \; X = Y \end{cases}$$

Και C[i] είναι το i-στό bit του χρωμοσώματος C, χρησιμοποιώντας δεικτοδότηση από το 0.

Β.1. [*5 μονάδες*] Να καταγράψετε όλα τα άτομα που έχουν την μέγιστη καταλληλότητα στο πεδίο καταλληλότητας που ορίζεται από την F (.). Ποια είναι η μέγιστη τιμή της καταλληλότητας;

<u>Δεν απαντήθηκε</u>

Β.2. [*5 μονάδες*] Να δώσετε ένα απλό παράδειγμα διασταύρωσης μονού σημείου, για οποιαδήποτε δύο χρωμοσώματα 6-bit, C1 και C2, όπου |F(C1) - F(C2)| =1, και να δείξετε ότι για ένα από τα παιδιά, K, ισχύει: F(K) = F(C1)+F(C2) = $f_{max,}$ (η μέγιστη τιμή της καταλληλότητας).

<u>Δεν απαντήθηκε</u>

Β.3. [*5 μονάδες*] Εξηγείστε πως το παράδειγμά σας, στο ερώτημα 2, επεξηγεί την Υπόθεση Δομικών Στοιχείων (Building Block Hypothesis - BBH) για τους ΓΑ.

<u>Δεν απαντήθηκε</u>

Β.4. [*5 μονάδες*] Χρησιμοποιώντας διασταύρωση 3 σημείων (σε οποιαδήποτε σημεία της επιλογής σας, σε οποιαδήποτε δύο χρωμοσώματα της επιλογής σας) δώστε ένα παράδειγμα το οποίο αντίκειται στο βασικό πνεύμα του BBH και εξηγείστε γιατί συμβαίνει αυτό.

<u>Δεν απαντήθηκε</u>

Θέμα 4: Εύρεση Στρατηγικής με Γενετικό Προγραμματισμό [*30 μόρια*]

Για να εκπαιδεύσουμε το Nemo, θα σχεδιάσουμε ένα μηχανικό ψάρι ώστε να μπορεί να αναζητά το φως και το οξυγόνο. Το μηχανικό ψάρι θα έχει 5 αισθητήρες: εμπρός, αριστερά, δεξιά, επάνω και κάτω. Κάθε αισθητήρας μπορεί να ανιχνεύσει μεταβολές δύο ποσοτήτων: φως και οξυγόνο (O_2). Οι μεταβολές υπολογίζονται μεταξύ των λεπιών του ψαριού και της άκρης του αισθητήρα. Η συσκευή μέτρησης είναι τόσο ακριβής έτσι ώστε να εξασφαλίζεται ότι θα υπάρχει διαφορά μεταξύ των δύο μετρήσεων. Έτσι, κάθε αισθητήρας θα δώσει μια ένδειξη αύξησης ή μείωσης (από το σώμα του ψαριού στην άκρη του αισθητήρα) για κάθε μία από τις δύο ποσότητες.

Συνολικά, το μηχανικό ψάρι θα έχει 10 παραμέτρους εισόδου (5 αισθητήρες x 2 ανιχνευόμενες ποσότητες).

Οι ενέργειες που μπορεί να κάνει το ψάρι είναι στρίψιμο, περιστροφή και κολύμβηση, με μια μικρή διακύμανση των πιθανών γωνιών στροφής και περιστροφής:

- Στροφή: 45^o αριστερά, 45^o δεξιά, 0^o (καμία στροφή), τυχαία στροφή μεταξύ 45^o αριστερά και 45^o δεξιά.
- Περιστροφή: 45^o επάνω, 45^o κάτω, 0^o (καμία περιστροφή), τυχαία περιστροφή μεταξύ 45^o επάνω και 45^o κάτω (για παράδειγμα, αν το ψάρι κινείται κατακόρυφα προς τα πάνω, με το κεφάλι προς τα πάνω και την ουρά προς τα κάτω, τότε η περιστροφή «45^o κάτω» θα το κάνει να συνεχίσει να κινείται προς τα πάνω, αλλά πλάγια και υπό γωνία 45^o).
- Η ενέργεια της κολύμβησης είναι δυαδική: είτε το ψάρι κολυμπά (με σταθερή ταχύτητα) στην κατεύθυνση του σώματος του ψαριού ή δεν κολυμπά.

Ζητείται να χρησιμοποιήσετε μια εξελικτική τεχνική για να εξελίξετε μια καλή στρατηγική ελέγχου του ψαριού. Οι στρατηγικές αποτελούνται από ένα σύνολο ενεργειών σε απόκριση διαφόρων προτύπων εισόδου των αισθητήρων. Πιο συγκεκριμένα, ζητείται:

Α. [*10 μονάδες*] Να δώσετε ένα σύντομο (αλλά λεπτομερές) προσχέδιο μιας μεθόδου για γραμμική προσέγγιση της στρατηγικής ελέγχου του ψαριού, έτσι ώστε μια πλήρης στρατηγική να μπορεί να κωδικοποιηθεί από $5*2^{10}$ bits σε ένα χρωμόσωμα δυαδικού γενετικού αλγορίθμου.

Υπόδειξη: Για την κωδικοποίηση είναι σκόπιμο να κατασκευάσετε ένα πίνακα, με τις πιθανές εισόδους και τις αντίστοιχες ενέργειες.

<u>Απάντηση:</u>

Το ψάρι βάση εκφώνησης πραγματοποιεί τις ακόλουθες ενέργειες:

Για την στροφή:
Αριστερά, Δεξιά, Τίποτα και μια τυχαία στροφή(αριστερά ή δεξιά)
Άρα έχουμε 4 επιλογές όπου θα χρειαστούμε $\log_2(4) = 2bit$

Για την περιστροφή:
πάνω, κάτω, Τίποτα και μια τυχαία περιστροφή(πάνω ή κάτω)
Άρα έχουμε 4 επιλογές όπου θα χρειαστούμε $\log_2(4) = 2bit$

Η ενέργεια έχει δύο επιλογές ή κολυμπάει ή δεν κολυμπάει.
Άρα: $\log_2(2) = 1bit$

Σύνολο Θα χρειαστούμε 5 bit κωδικοποίησης για την ενέργεια του ψαριού.

Το ψάρι έχει 5 αισθητήρες όπου ο κάθε ένας από τους αισθητήρες μπορεί να ανιχνεύσει δύο μεταβολές το φως και το οξυγόνο και είναι ανεξάρτητοι. Άρα ο αισθητήρας έχει βάση της συνδυαστικής του κανόνα του γινομένου m x n πιθανά αποτελέσματα 5x2=10, όπου έχουμε βάση του συνόλου αυτού θέλουμε δύο bit για την κάθε διάταξη. Άρα από τη συνδυαστική διάταξη με επανάληψη θα χρειαστούμε 2^{10} = 1024 διαφορετικά bit για την κωδικοποίηση.

Βάση συνδυαστικής του κανόνα του γινομένου έχουμε 5 bit για τις ενέργειες του ψαριού επί 1024 bit με τις δυνατές λειτουργίες του αισθητήρα. $5*2^{10}$.

Υπόδειξη:

αισθητήρες										ενέργειες				
Φως εμπρός	Φως αριστερά	Φως δεξιά	Φως επάνω	Φως κάτω	οξυγόνο εμπρός	οξυγόνο αριστερά	οξυγόνο δεξιά	οξυγόνο επάνω	οξυγόνο κάτω	στροφή		περιστροφή		κολύμβηση
0	0	0	0	1	0	0	0	0	1	0	0	0	0	1
0	1	0	1	0	1	0	1	0	1	0	1	0	1	0

Άρα παράδειγμα ο δεύτερος συνδυασμός: ο αισθητήρας αισθάνεται Φως αριστερά, φως επάνω, οξυγόνο εμπρός, οξυγόνο δεξιά και οξυγόνο κάτω -> τότε θα κάνει τις εξής ενέργειες: στροφή και περιστροφή. Όπου:

Στροφή				Περιστροφή				Κολύμβηση	
αριστερά	δεξιά	τίποτα	Τυχαία περιστροφή	αριστερά	δεξιά	τίποτα	Τυχαία περιστροφή	Κολυμπά	Δεν κολυμπά
00	01	10	11	00	01	10	11	0	1

Συγκεκριμένα: 01 , θα κάνει στροφή δεξιά και για 01 θα κάνει και μια περιστροφή δεξιά.

Β. [*10 μονάδες*] Για την εξέλιξη ενός πληθυσμού από NEMOBOT χρησιμοποιώντας Γενετικό Προγραμματισμό (ΓΠ), ζητείται να περιγράψετε το σύνολο συναρτήσεων και τερματικών. Πρέπει να συμπεριλάβετε και κάποιες βασικές λογικές συναρτήσεις στο σύνολο συναρτήσεων. Όλες οι λογικές συναρτήσεις θα παίρνουν δυαδικές τιμές σαν είσοδο και θα παράγουν μια δυαδική τιμή σαν έξοδο.

<u>Δεν απαντήθηκε</u>

Γ. [*10 μονάδες*] Να γράψετε ένα παράδειγμα που θα παρήγαγε ο ΓΠ αν χρησιμοποιούσε το σύνολο συναρτήσεων που προτείνατε και εξηγείστε σύντομα πως αυτό θα έτρεχε.

<u>Σημείωση:</u> Για λόγους απλούστευσης, αφού δεν καθορίζεται πως θα αξιολογηθεί η απόδοση του ψαριού, δεν χρειάζεται να ορίσετε τη συνάρτηση καταλληλότητας.

<u>Δεν απαντήθηκε</u>