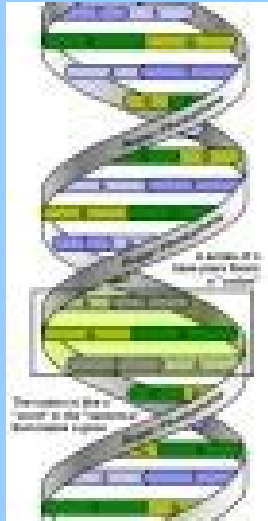




ΜΕΡΟΣ Α

Κεφάλαιο 7



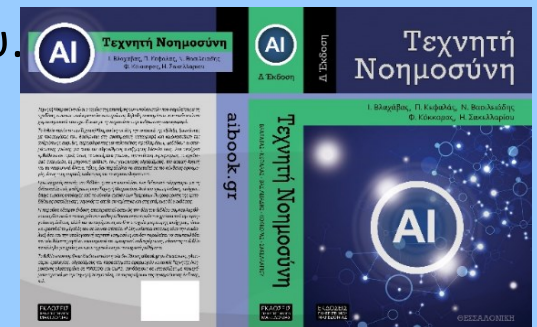
Γενετικοί Αλγόριθμοι

Ι. Βλαχάβας, Π. Κεφαλάς, Ν. Βασιλειάδης, Φ. Κόκκορας, Η. Σακελλαρίου.

Τεχνητή Νοημοσύνη - Δ' Έκδοση, ISBN: 978-618-5196-44-8

Έκδοση/Διάθεση: Εκδόσεις Πανεπιστημίου Μακεδονίας, 2020

Κωδικός Βιβλίου στον Εύδοξο: 94700120





Εισαγωγή

- ❖ Σε αρκετές περιπτώσεις το μέγεθος ενός προβλήματος καθιστά απαγορευτική τη χρήση κλασικών μεθόδων αναζήτησης για την επίλυσή του.
 - ❑ Αυτό εμφανίζεται σε μια μεγάλη κατηγορία προβλημάτων και ειδικότερα σε προβλήματα βελτιστοποίησης (optimization) που απαντώνται συχνά σε βιομηχανικές εφαρμογές, όπως
 - η σχεδίαση VLSI κυκλωμάτων
 - κοπή υλικών (ύφασμα, δέρμα, μέταλλο, ξύλο, γυαλί)
 - τοποθέτηση κεραιών κινητής τηλεφωνίας,
 - κλπ
- ❖ Στις περιπτώσεις αυτές βρίσκουν εφαρμογή πιθανοκρατικοί αλγόριθμοι οι οποίοι αν και δεν εγγυώνται ότι θα βρουν τη βέλτιστη λύση, είναι ικανοί να επιστρέψουν μια αρκετά καλή λύση σε εύλογο χρονικό διάστημα.
- ❖ Μια κατηγορία τέτοιων αλγορίθμων επίλυσης προβλημάτων είναι οι **γενετικοί αλγόριθμοι** (*genetic algorithms*), των οποίων ο βασικός μηχανισμός είναι εμπνευσμένος από τη Δαρβινική θεωρία της εξέλιξης των ειδών (evolution, Charles Darwin - 1875).



Θεωρία της εξέλιξης (evolution)



Κανόνας της φυσικής επιλογής

- ☐ Οι οργανισμοί που δε μπορούν να επιβιώσουν στο περιβάλλον τους πεθαίνουν, ενώ οι υπόλοιποι πολλαπλασιάζονται μέσω της **αναπαραγωγής**.
- ☐ Οι απόγονοι παρουσιάζουν μικρές διαφοροποιήσεις από τους προγόνους τους, ενώ συνήθως **υπερισχύουν αυτοί που συγκεντρώνουν τα καλύτερα χαρακτηριστικά**.
- ☐ Σποραδικά συμβαίνουν τυχαίες **μεταλλάξεις**, από τις οποίες οι περισσότερες οδηγούν τα μεταλλαγμένα άτομα στο θάνατο, αν και είναι πιθανό, πολύ σπάνια όμως, να οδηγήσουν στη δημιουργία νέων "καλύτερων" οργανισμών.
- ☐ Αν το περιβάλλον μεταβάλλεται με αργούς ρυθμούς, τα διάφορα είδη μπορούν να εξελίσσονται σταδιακά ώστε να **προσαρμόζονται** σε αυτό.



Παραγωγή Λύσεων

- ❖ Στους γενετικούς αλγόριθμους, αντί να ψάχνουμε για λύσεις, χρησιμοποιούμε υπάρχουσες που συνήθως δεν είναι ποιοτικά αποδεκτές και συνδυάζοντάς τις με κάποιο τρόπο κατασκευάζουμε νέες (απογόνους), ελπίζοντας ότι αυτές θα είναι καλύτερες από τους "γονείς" τους.
- ❖ Δηλαδή οι γενετικοί αλγόριθμοι εκτελούν μία αναζήτηση στο χώρο των υποψηφίων λύσεων, με στόχο την εύρεση αποδεκτών, σύμφωνα με κάποιο κριτήριο, λύσεων.
 - ❑ Έχουν εφαρμοστεί επιτυχώς σε προβλήματα βελτιστοποίησης, όπως
 - δρομολόγηση καλωδίων (wire routing),
 - χρονοπρογραμματισμό (scheduling),
 - παίγνια (game playing),
 - προβλήματα εφοδιαστικής (logistics),
 - προβλήματα τύπου πλανόδιου πωλητή,
 - προβλήματα βέλτιστου ελέγχου,
 - βελτιστοποίηση ερωτήσεων σε βάσεις δεδομένων,
 - κτλ.
 - ❑ Να τονιστεί εδώ ότι ο όρος "λύση" χρησιμοποιείται με αρκετά ευρεία ερμηνεία.
 - Για παράδειγμα, στο πρόβλημα του πλανόδιου πωλητή, οποιαδήποτε διαδρομή που περνά από όλες τις πόλεις θεωρείται λύση αλλά όχι η καλύτερη και φυσικά είναι εύκολο να βρει κανείς τέτοιες λύσεις.
 - Αυτό που δεν είναι εύκολο είναι να βρει κανείς τη βέλτιστη διαδρομή (με κάποιο κριτήριο ποιότητας).



Ιστορικά στοιχεία

- ❑ Το 1960 ο L.J.Fogel (1928-2007), αποπειράθηκε να βελτιστοποιήσει προγράμματα κρατώντας σταθερή τη δομή τους αλλά εξελίσσοντας τις τιμές των παραμέτρων τους (*εξελικτικός προγραμματισμός - evolutionary programming*),
- ❑ Το 1975 ο Holland (1929-2015) έδωσε νέα ώθηση στο χώρο, χρησιμοποιώντας σειρές bits για να αναπαραστήσει λειτουργίες, με τρόπο τέτοιο ώστε κάθε συνδυασμός bits να είναι μια έγκυρη λειτουργία.



Γενική Μορφή Γενετικού Αλγόριθμου



Βήματα:

- ☐ Δημιούργησε τυχαία έναν αρχικό πληθυσμό Π , με N υποψήφιος (μη αποδεκτές, δηλαδή είτε μη έγκυρες ή μη βέλτιστες, κλπ) λύσεις.
- ☐ Βαθμολόγησε (δηλ. πόσο κοντά σε μια αποδεκτή λύση είναι) κάθε υποψήφια λύση χρησιμοποιώντας μια συνάρτηση καταλληλότητας (*fitness function*).
- ☐ Σχημάτισε $N/2$ ζευγάρια όχι απαραίτητα μοναδικών γονέων
 - Δηλαδή μια λύση μπορεί να είναι γονέας περισσότερες φορές, δίνοντας μεγαλύτερη προτεραιότητα στις καλύτερες λύσεις.
- ☐ Κάθε ζευγάρι ζευγαρώνει (*mates*), δίνοντας δύο νέες λύσεις, τους απογόνους (*offsprings*).
- ☐ Πιθανοκρατικά, σε κάποια από τα άτομα/απογόνους, εφαρμόζεται ένας τελεστής μετάλλαξης (*mutation*).
- ☐ Ο νέος πληθυσμός Π' αποτελείται από το σύνολο των απογόνων και συνήθως αποτελεί βελτίωση του προηγούμενου πληθυσμού.
 - Η ανανέωση του πληθυσμού μπορεί να μην είναι πλήρης αλλά ο νέος πληθυσμός Π' να αποτελείται και από απογόνους του αρχικού πληθυσμού Π .
- ☐ Η διαδικασία επαναλαμβάνεται για το νέο πληθυσμό Π'



Γενική Μορφή Γενετικού Αλγόριθμου (συνεχ.)

- ❖ Οι πιο συνηθισμένες συνθήκες τερματισμού της είναι η εύρεση μιας τέλειας λύσης με βάση τη συνάρτηση καταλληλότητας ή η σύγκλιση όλων των λύσεων σε μια.
- ❖ Η αναζήτηση που εκτελείται δεν είναι πλήρης.
 - ❑ Αντίθετα, έχει στοιχεία προσομοιωμένης ανόπτησης (*simulated annealing*) καθώς σε κάθε κύκλο λειτουργίας δεν επιλέγονται μόνο οι καλύτερες λύσεις βάσει της συνάρτησης καταλληλότητας αλλά και κάποιες υποδεέστερες ποιοτικά
 - ❑ Έτσι, περιορίζεται μεν ο κίνδυνος να παγιδευτεί η διαδικασία σε τοπικό ακρότατο ως προς τη συνάρτηση καταλληλότητας, αλλά ταυτόχρονα καθίσταται και απίθανο να βρεθεί το απόλυτο ακρότατο.
- ❖ Συνήθως και καθώς εξελίσσονται οι γενιές, ο πληθυσμός συγκλίνει προς την περιοχή μιας καλής λύσης αλλά δυσκολεύεται να την "πετύχει" ακριβώς.
- ❖ Μια αποδεκτή λύση μπορεί να προκύψει ανά πάσα στιγμή, λόγω της "στοχαστικότητας" των επί μέρους διαδικασιών.
- ❖ Ο γενετικός αλγόριθμος ΔΕΝ εγγυάται ότι θα βρει την καλύτερη λύση αλλά συνήθως βρίσκει μια πολύ καλή



Γενική Μορφή Γενετικού Αλγόριθμου (συνεχ.)

1. Δημιούργησε έναν αρχικό πληθυσμό Π , με N υποψήφιας λύσεις.
2. Υπολόγισε την καταλληλότητα κάθε λύσης.
3. Όσο δεν ισχύει κάποια συνθήκη τερματισμού:
 - α. Επανάλαβε $N/2$ φορές τα ακόλουθα βήματα:
 - i. Επέλεξε δύο λύσεις από τον πληθυσμό Π .
 - ii. Συνδύασε τις δύο λύσεις για να βγάλεις δύο απογόνους.
 - iii. Υπολόγισε την καταλληλότητα των δύο απογόνων.
 - β. Δημιούργησε το νέο πληθυσμό Π' έχοντας υπόψη όλους τους νέους απογόνους που προέκυψαν από το βήμα 3α και θέσε $\Pi = \Pi'$



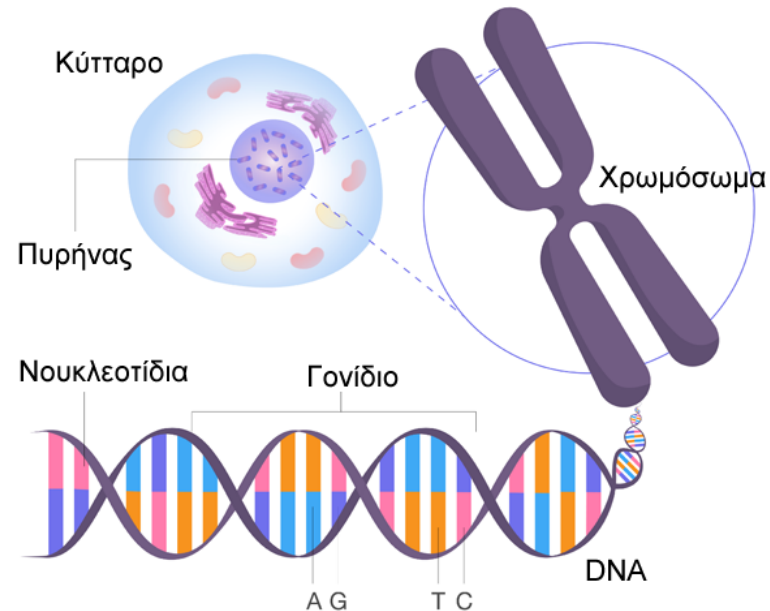
Λειτουργίες και Τελεστές Γενετικού Αλγόριθμου

- ❖ Ένας γενετικός αλγόριθμος για ένα συγκεκριμένο πρόβλημα περιλαμβάνει **οκτώ βήματα**:
 - ☐ Κωδικοποίηση (*encoding*) των υποψηφίων λύσεων,
 - ☐ Δημιουργία του αρχικού πληθυσμού (*initialization*),
 - ☐ Ορισμός της συνάρτησης καταλληλότητας (ή ευρωστίας) (*fitness function*) για την αξιολόγηση των λύσεων (*fitness evaluation*),
 - ☐ Ορισμός του μηχανισμού επιλογής γονέων,
 - ☐ Ορισμός του τελεστή αναπαραγωγής (ή ζευγαρώματος) (*mating operator*),
 - ☐ Ορισμός του τελεστή μετάλλαξης (*mutation operator*),
 - ☐ Ορισμός του πληθυσμού της επόμενης γενιάς, και
 - ☐ Ορισμός συνθηκών τερματισμού του αλγορίθμου.



Κωδικοποίηση Υποψήφιων Λύσεων

- ❖ Στους βιολογικούς οργανισμούς, ένα **χρωμόσωμα** είναι ένα μεγάλο μόριο (ακολουθία) DNA και περιέχει έναν αριθμό **γονιδίων**.



- ❖ Στο πραγματικό DNA το αλφάβητο έχει μήκος τέσσερα και αποτελείται από τα γράμματα A, G, T και C που αντιστοιχούν στα τέσσερα διαφορετικά νουκλεοτίδια (βάσεις) που το συνθέτουν (Adenine, Guanine, Thymine και Cytosine).
- ❖ Ένας οργανισμός μπορεί να έχει ένα ή περισσότερα χρωμοσώματα ενώ σε κάποιους οργανισμούς κάθε κύτταρο περιέχει δύο αντίγραφα για κάθε χρωμόσωμα.
 - ❑ Για παράδειγμα, ο άνθρωπος έχει 23 ζεύγη χρωμοσωμάτων.



Κωδικοποίηση Υποψήφιων Λύσεων (συνεχ.)

- ❖ Στην κλασική προσέγγιση των γενετικών αλγορίθμων, κάθε υποψήφια λύση αναπαρίσταται με μία **συμβολοσειρά** (*string*) ενός πεπερασμένου αλφάβητου.
 - ❑ Συνήθως χρησιμοποιείται το δυαδικό αλφάβητο, οπότε οι συμβολοσειρές ονομάζονται και δυαδικές συμβολοσειρές (bit-strings).
 - Υπάρχουν περιπτώσεις γενετικών αλγορίθμων που χρησιμοποιούν πιο σύνθετες μορφές αναπαράστασης.
 - ❑ Στα περισσότερα προβλήματα οι λύσεις περιγράφονται με μεταβλητές διαφόρων τύπων δεδομένων, επομένως η διαδικασία της κωδικοποίησης περιλαμβάνει τη μετατροπή των τιμών αυτών των μεταβλητών στις αντίστοιχες δυαδικές.
 - Για **παράδειγμα**, αν μια λύση περιγράφεται από 2 μεταβλητές, μία boolean με τιμή π.χ. 1 και μια short integer με τιμή π.χ. 73 ή 01001001, η κωδικοποίηση θα απαιτούσε 9 bits και το παραγόμενο bit-string θα ήταν το 101001001.
- ❖ Κατ' αναλογία με τη βιολογία, η συμβολοσειρά συνήθως αναφέρεται και σαν **χρωμόσωμα** (*chromosome*) ενώ τα επιμέρους τμήματά της που κωδικοποιούν κάποιο χαρακτηριστικό, δηλαδή κάποια μεταβλητή, ονομάζονται **γονίδια** (*gene*).
 - ❑ Στους γενετικούς αλγορίθμους έχουμε μόνο ένα χρωμόσωμα και ένα ή περισσότερα γονίδια.



Κωδικοποίηση Υποψήφιων Λύσεων (συνεχ.)

- ❖ Στη γενετική, το σύνολο των παραμέτρων που αναπαρίστανται από ένα συγκεκριμένο γονίδιο που μας ενδιαφέρει ή έναν αριθμό γονιδίων (ενδεχομένως και όλων) αναφέρεται σαν **γονότυπος** (*genotype*).
- ❖ Ένα γονίδιο μπορεί να κωδικοποιεί κάποιο χαρακτηριστικό του οργανισμού, όπως το χρώμα των ματιών, και προσδιορίζει ως ένα βαθμό την εμφάνιση του.
- ❖ Η συνολική φυσική εμφάνιση ή συγκεκριμένη εκδήλωση ενός χαρακτηριστικού (για παράδειγμα το χρώμα ματιών) ονομάζεται **φαινότυπος** (*phenotype*).
- ❖ Κάθε γονότυπος (δηλαδή ένα χρωμόσωμα στους γενετικούς αλγορίθμους) θα μπορούσε να αποτελεί μια πιθανή λύση στο πρόβλημα, η σημασία της οποίας (δηλαδή ο φαινότυπος) καθορίζεται από το χρήστη.
- ❖ Στο σχήμα αναπαρίσταται η δυαδική κωδικοποίηση μιας λύσης που περιγράφεται με 2 μεταβλητές, μια Boolean X (1 bit) και ενός ακεραίου Y (8 bit).
 - ❑ Επομένως, το χρωμόσωμα 101001001 θα περιέχει 2 γονίδια, ένα μήκους ενός bit στην αρχή και ένα μήκους 8 bit στη συνέχεια

Γονότυπος

X	Y
---	---

Φαινότυπος

1	0	1	0	0	1	0	0	1
---	---	---	---	---	---	---	---	---



Κωδικοποίηση Πραγματικών Αριθμών

- ❖ Όταν μία παράμετρος παίρνει συνεχείς (πραγματικές) τιμές σε κάποιο αυθαίρετο διάστημα τιμών μπορούμε είτε
 - ☐ Να τους χρησιμοποιήσουμε στη φυσική τους εκδοχή, ή
 - ☐ Να κάνουμε δυαδική κωδικοποίηση τους, η οποία
 - Προτιμάται καθώς είναι βολική προσέγγιση
 - Οι διάφοροι τελεστές που χρησιμοποιούνται λειτουργούν γρηγορότερα με δυαδικές κωδικοποιήσεις, άρα βελτιώνεται συνολικά η απόδοση του γενετικού αλγόριθμου
 - ☐ Μειονεκτήματα
 - Απαιτείται προσπάθεια για τη μετατροπή των παραμέτρων του προβλήματος σε δυαδική κωδικοποίηση
 - Είναι απαραίτητο να είναι γνωστό το πεδίο τιμών των παραμέτρων
 - Η ακρίβεια που παρέχει η δυαδική κωδικοποίηση πραγματικών αριθμών εξαρτάται από το πλήθος των bit που θα χρησιμοποιηθούν
 - Μεγαλύτερη ακρίβεια για δεδομένο εύρος τιμών απαιτεί περισσότερα bit, που σημαίνει επιβάρυνση σε μνήμη και χρόνο εκτέλεσης



Α) Δυαδική Κωδικοποίηση Πραγματικών Αριθμών

- ❖ Το πλήθος των bit που απαιτούνται για κωδικοποίηση πραγματικών αριθμών μεταξύ a και b με ακρίβεια k δεκαδικά ψηφία, υπολογίζεται από τη σχέση:

$$n = \left\lceil \log_2 \left((b - a) \cdot 2 \cdot 10^k \right) \right\rceil$$

- ❑ Για παράδειγμα, για να κωδικοποιήσουμε δεκαδικούς ακρίβειας τριών δεκαδικών ψηφίων ($k=3$) μεταξύ $a=0$ και $b=1000$, απαιτούνται 20 bit, ενώ για ακρίβεια πέντε δεκαδικών ψηφίων, 27 bit.
- ❑ Αν π.χ. το $x \in [19.1 .. 44.6]$ για ακρίβεια ενός δεκαδικού ψηφίου, ($k=1$) απαιτούνται 8 bit
- ❑ Ένας πραγματικός αριθμός με 6 δεκαδικά, π.χ. ο $x=1.666504$ αναπαρίσταται με 22 bit, **1110001110001010101010** ο οποίος αντιστοιχεί στον ακέραιο 3728042

- ❖ **Λεπτομέρειες στο βιβλίο**



Β) Κωδικοποίηση Πραγματικών Αριθμών

- ❖ Στην κωδικοποίηση πραγματικών τιμών (ή κινητής υποδιαστολής) οι παράμετροι του προβλήματος χρησιμοποιούνται στη φυσική τους εκδοχή
 - ☐ Δηλαδή ο γονότυπος ταυτίζεται με τον φαινότυπο
 - ☐ Είναι η καταλληλότερη επιλογή για προβλήματα βελτιστοποίησης που ορίζονται σε κάποιον συνεχή χώρο αναζήτησης γιατί δεν απαιτούνται τα ενδιάμεσα βήματα κωδικοποίησης και αποκωδικοποίησης των παραμέτρων που απαιτούνται στη δυαδική κωδικοποίηση πραγματικών αριθμών
 - ☐ Παρέχει γενικά καλύτερη ακρίβεια από τη δυαδική κωδικοποίηση
 - ☐ Τέλος, είναι διαισθητικά προτιμότερο να γίνεται αναζήτηση στο χώρο των πραγματικών αριθμών έχοντας και κωδικοποίηση πραγματικών αριθμών.
 - ☐ Για παράδειγμα, στο πρόβλημα τοποθέτησης δύο κεραιών κινητής τηλεφωνίας η λύση είναι δύο γεωγραφικές συντεταγμένες, έστω (Lat1, Lon1) και (Lat2, Lon2) με εύρος τιμών που καθορίζεται από τις παραμέτρους του προβλήματος.
 - ☐ Σε μια τέτοια περίπτωση, ο γονότυπος του ατόμου (λύση-χρωμόσωμα) θα περιέχει 4 γονίδια

Γονότυπος

Lat1	Lon1	Lat2	Lon2
------	------	------	------

Φαινότυπος

23.122	41.556	21.345	39.988
--------	--------	--------	--------



Κωδικοποίηση διάταξης

- ❖ Στην κωδικοποίηση διάταξης ή συνδυασμού (*order* ή *permutation encoding*) χρησιμοποιούνται ακέραιοι σε κάποια διάταξη, χωρίς όμως να επαναλαμβάνονται.
- ❖ Το πρόβλημα του πλανόδιου πωλητή (traveling salesman problem) είναι μια χαρακτηριστική περίπτωση προβλήματος βελτιστοποίησης
 - ☐ Υπάρχουν N πόλεις που ο πωλητής πρέπει να επισκεφτεί από μία φορά
 - ☐ Πρέπει να ελαχιστοποιηθεί κάποια συνάρτηση κόστους
 - Για παράδειγμα, η συνολική απόσταση που διανύθηκε ή ο χρόνος που απαιτήθηκε ή το καύσιμο που καταναλώθηκε ή και συνδυασμός παραγόντων
 - ☐ Για N πόλεις υπάρχουν $N!$ υποψήφιες λύσεις
- ❖ Κωδικοποίηση διάταξης για το πρόβλημα του πλανόδιου πωλητή με $N=14$

7	2	14	4	11	6	1	8	9	10	5	12	13	3
---	---	----	---	----	---	---	---	---	----	---	----	----	---


- ☐ Η διάταξη κωδικοποιεί μία διαδρομή που ξεκινά από την πόλη 7 και καταλήγει σε αυτή, αφού περάσει μία φορά από τις υπόλοιπες πόλεις.
- ☐ Πόλεις που δεν συνδέονται οδικώς μεταξύ τους δεν θα πρέπει προφανώς να τοποθετηθούν διαδοχικά στη διάταξη.
- ☐ Δεν είναι απαραίτητο να επαναληφθεί η αρχική πόλη (εδώ η 7) στο τέλος της κωδικοποίησης γιατί θα μας δυσκόλευε λίγο στο στάδιο της αναπαραγωγής
 - Εννοείται βέβαια ότι η τελική μετάβαση από την πόλη 3 πίσω στην αρχή (στην πόλη 7) θα συνυπολογιστεί στη συνάρτηση καταλληλότητας που αξιολογεί την ποιότητα της λύσης.



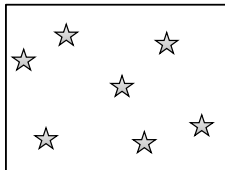
Αρχικοποίηση Πληθυσμού

- ❖ Ο μηχανισμός δημιουργίας του αρχικού πληθυσμού εξαρτάται από το εκάστοτε πρόβλημα και συνήθως είναι προφανής. Για παράδειγμα,
 - ❑ Στο πρόβλημα του πλανόδιου πωλητή, οποιαδήποτε διαδρομή δεν παραβιάζει τις απαιτήσεις του προβλήματος είναι μια λύση, φυσικά όχι η βέλτιστη, πιθανώς ούτε καν καλή.
 - ❑ Στο πρόβλημα της τοποθέτησης N κεραιών κινητής τηλεφωνίας, οποιοδήποτε σύνολο N σημείων αποτελεί λύση, πλην όμως όχι απαραίτητα τη βέλτιστη.
 - ❑ Στο πρόβλημα μεγιστοποίησης μιας μονοπαραμετρικής συνάρτησης $y=f(x)$, οποιοδήποτε x στο πεδίο ορισμού της συνάρτησης μπορεί να θεωρηθεί "λύση" κάποιας ποιότητας.
- ❖ Τυπικά μεγέθη πληθυσμών σε γενετικό αλγόριθμο είναι μερικές δεκάδες ως λίγες εκατοντάδες άτομα
 - ❑ Αν είναι πολύ μικρό, η ικανότητα αναζήτησης ανά γενιά περιορίζεται, η αναζήτηση θα πάρει πολλές γενιές και ίσως δεν δώσει και καλό αποτέλεσμα, με συχνότερο πρόβλημα τη σύγκλιση του πληθυσμού σε κάποιο τοπικό ακρότατο (και όχι το ολικό).
 - ❑ Αν το μέγεθος του αρχικού πληθυσμού είναι πολύ μεγάλο, θα υπάρχει υπολογιστική επιβάρυνση (χρόνος, μνήμη).
- ❖ Η αρχικοποίηση θα πρέπει να παράγει άτομα όσο περισσότερο διεσπαρμένα στο χώρο των λύσεων.

A



B



 - ❑ Ζητούμενο (αλλά όχι πάντα εύκολο να γίνει) είναι να υπάρχει επαρκής εκπροσώπηση λύσεων από όλες τις περιοχές του χώρου των λύσεων. Αλλιώς θα δυσκολέψει την αναζήτηση.



Συνάρτηση Καταλληλότητας

- ❖ Η *συνάρτηση καταλληλότητας* (*fitness function*) (ή συνάρτηση ευρωστίας) δέχεται ως είσοδο ένα χρωμόσωμα και επιστρέφει έναν αριθμό (συνήθως στο διάστημα $[0,1]$), που υποδηλώνει το πόσο κατάλληλο είναι.
- ❖ Η αξιολόγηση αυτή χρησιμοποιείται:
 - ☐ είτε από τη συνθήκη τερματισμού ή
 - ☐ από τη διαδικασία της πιθανοκρατικής επιλογής τους για να συμπεριληφθούν (ή όχι) στον πληθυσμό της επόμενης γενιάς.
- ❖ Η κατασκευή της μπορεί να είναι από απλή έως εξαιρετικά πολύπλοκη.
 - ☐ Η ιδανική συνάρτηση καταλληλότητας θα έπρεπε να είναι συνεχής και μονότονη.
- ❖ Προσεγγιστική συνάρτηση καταλληλότητας (*approximate fitness function*)
 - ☐ Συμβιβασμός μεταξύ επιθυμητής ακρίβειας και υπολογιστικού κόστους.
- ❖ Για παράδειγμα η συνάρτηση καταλληλότητας:
 - ☐ Στο πρόβλημα της τοποθέτησης κεραιών κινητής τηλεφωνίας, θα μπορούσε να υπολογίζει το ποσοστό του πληθυσμού που κατοικεί στην περιοχή και έχει ένταση σήματος συσκευής πάνω από κάποια τιμή κατωφλίου.
 - ☐ Σε ένα πρόβλημα μεγιστοποίησης μιας συνάρτησης, θα μπορούσε να είναι η ίδια η συνάρτηση
 - ☐ Σε προβλήματα διάταξης σε επιφάνεια θα μπορούσε να υπολογίζει το μέγεθος της απαιτούμενης επιφάνειας για την διάταξη των σχημάτων.
 - ☐ Στο TSP το μήκος της συνολικής διαδρομής



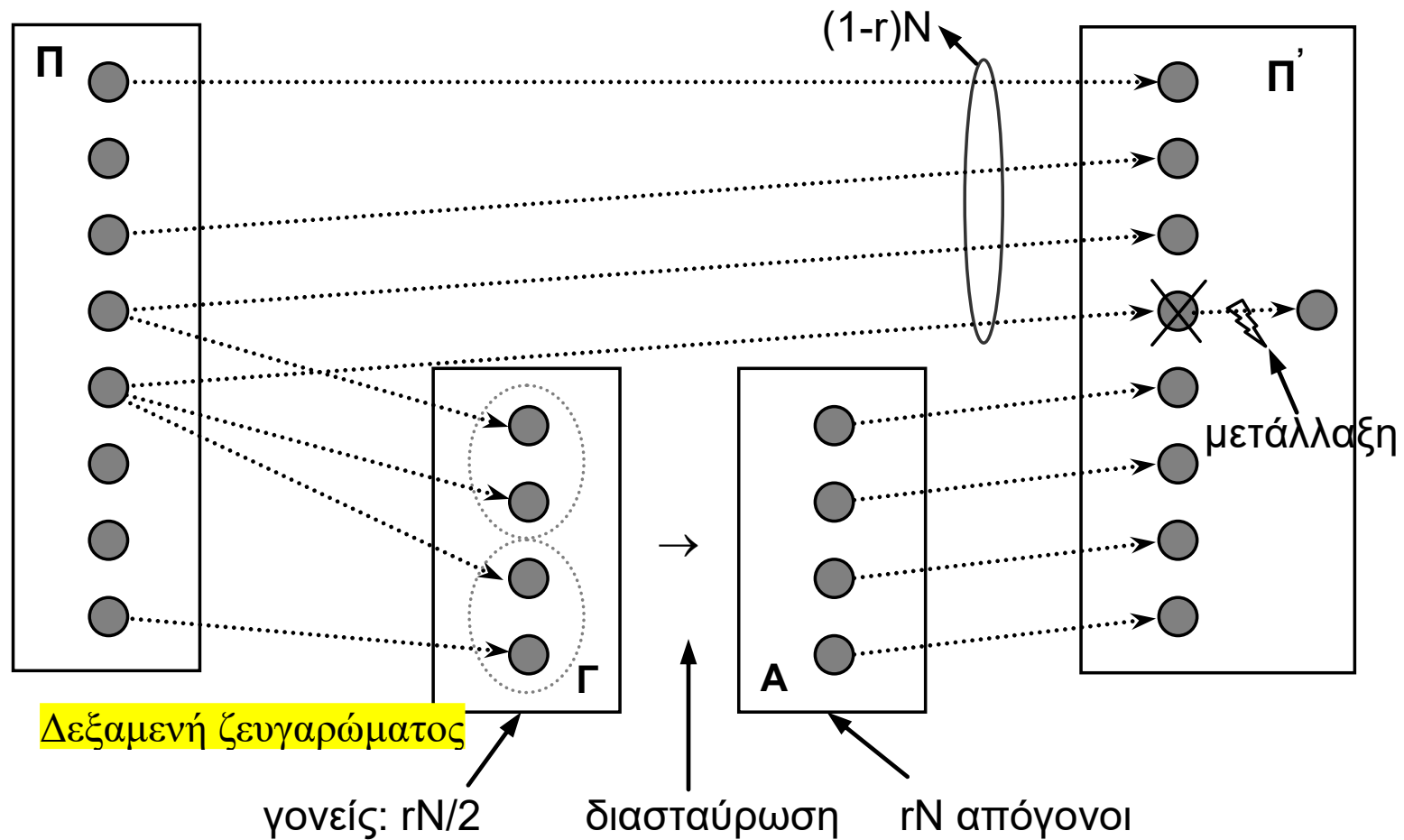
Διαδικασία Επιλογής Γονέων (1/3)

- ❖ Κατά τη διαδικασία επιλογής, αρχικά οι υποψήφιες λύσεις αντιγράφονται σε μια **δεξαμενή ζευγαρώματος** (*mating pool*).
 - ❑ Η δεξαμενή αυτή έχει μέγεθος ίσο με τον αρχικό πληθυσμό, εκτός από τις περιπτώσεις μερικής ανανέωσης, όπου είναι μικρότερη.
 - ❑ Σε αυτήν αντιγράφονται μέλη του αρχικού πληθυσμού, με πιθανότητα ανάλογη της καταλληλότητάς τους.
 - ❑ Κάποιοι γονείς με υψηλή τιμή στη συνάρτηση καταλληλότητας ενδέχεται να επιλεγούν προς αναπαραγωγή περισσότερες από μία φορές, ενώ κάποιοι γονείς με χαμηλή καταλληλότητα ενδέχεται να μην επιλεγούν καθόλου.
 - ❑ Για την επιλογή χρησιμοποιούνται αρκετές τεχνικές (π.χ. **Τεχνική της ρουλέτας**)
 - Κοινό χαρακτηριστικό όλων είναι ότι ευνοούν τα χρωμοσώματα (λύσεις) καλής ποιότητας ενώ οι διαφορές τους εστιάζονται στο πόσο έντονα ή όχι το κάνουν
- ❖ Ο τρόπος επιλογής των γονέων που θα ζευγαρώσουν επηρεάζει σημαντικά την απόδοση των γενετικών αλγορίθμων.



Γενικός γενετικός αλγόριθμος

Σχηματική λειτουργία (μερική ανανέωση)





Διαδικασία Επιλογής Γονέων (συνεχ.)

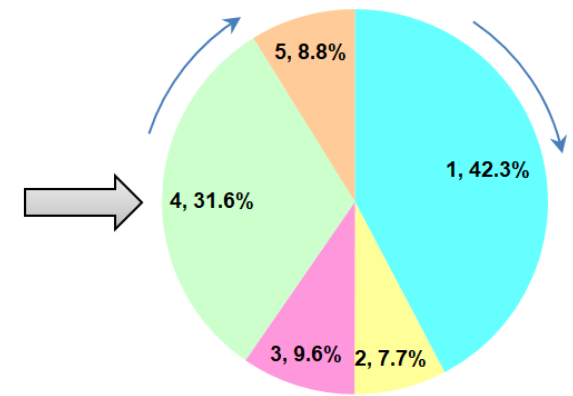
Τεχνική της ρουλέτας (roulette wheel selection)

- ❖ Είναι ίσως ο πιο συνηθισμένος μηχανισμός επιλογής γονέων
- ❖ Η πιθανότητα επιλογής ενός χρωμοσώματος x_i για να λειτουργήσει ως γονέας είναι ευθέως ανάλογη της τιμής καταλληλότητας του
 - ❑ Για αυτό ονομάζεται και μηχανισμός επιλογής **αναλογικής καταλληλότητας** (*fitness proportionate selection*) όπου η πιθανότητα επιλογής ενός χρωμοσώματος x_i , υπολογίζεται από τη σχέση:
$$P(x_i) = \frac{\text{Καταλληλότητα}(x_i)}{\sum_{j=1}^N \text{Καταλληλότητα}(x_j)}$$
 - ❑ Ο παρονομαστής ισούται με τη συνολική καταλληλότητα (ποιότητα) όλου του πληθυσμού των N λύσεων



Τεχνική της ρουλέτας (roulette wheel selection)

- ❖ Ένας τρόπος υλοποίησης της αναλογικής καταλληλότητας.
 1. Υπολογίζεται η τιμή καταλληλότητας κάθε μιας από τις N υποψήφιας λύσεις
 2. Παράγεται το άθροισμα S όλων των τιμών καταλληλότητας
 3. Επιλέγεται ένας τυχαίος αριθμός n , από το 0 μέχρι το S , χρησιμοποιώντας συνάρτηση ομοιόμορφης κατανομής για τη δημιουργία των τυχαίων αριθμών.
 4. Επαναληπτικά εξετάζεται κάθε υποψήφια λύση και η τιμή του προστίθεται σε έναν καταχωρητή K .
 5. Αν η τιμή του K γίνει μεγαλύτερη ή ίση του n , η λύση επιλέγεται και ο K μηδενίζεται. Στην αντίθετη περίπτωση εκτελείται πάλι το 4.
 6. Αν δεν έχει επιλεγεί ικανοποιητικός αριθμός υποψηφίων λύσεων εκτελείται το 3, αλλιώς ο αλγόριθμος τερματίζει.
- ❖ Στην πράξη, ο παραπάνω αλγόριθμος υλοποιεί ένα δίσκο ρουλέτας, χωρισμένο σε N τομείς, με κάθε τομέα να αντιστοιχεί σε μία λύση x_i έχοντας μέγεθος ανάλογο του $P(x_i)$
- ❖ Οι υποψήφιας λύσεις με μεγάλη τιμή καταλληλότητας έχουν μεγαλύτερη πιθανότητα να αυξήσουν την τιμή του καταχωρητή K ώστε να υπερβεί την τιμή n και συνεπώς να επιλεγούν.
- ❖ Κάποιες υποψήφιας λύσεις μπορεί να επιλεγούν περισσότερο από μία φορές.





Επιλογή τουρνουά

- ❖ Η επιλογή ρουλέτας ευνοεί τα χρωμοσώματα υψηλής καταλληλότητας αλλά το κάνει σχετικά έντονα και σε βάρος των λιγότερο ποιοτικών χρωμοσωμάτων (λύσεων).
- ❖ Αυτή η προτίμηση είναι φυσικά επιθυμητή αλλά όχι σε τόσο έντονο βαθμό, ώστε να αποτραπούν φαινόμενα πρόωρης σύγκλισης του πληθυσμού των λύσεων προς τις ποιοτικές λύσεις κάτι που περιορίζει την αναζήτηση
- ❖ Σε μια άλλη τεχνική, που ονομάζεται **επιλογή τουρνουά** (*tournament selection*) επιλέγονται τα καταλληλότερα χρωμοσώματα με μια προκαθορισμένη πιθανότητα q και τα λιγότερα κατάλληλα με πιθανότητα $1-q$.
- ❖ Αυτό οδηγεί σε πληθυσμούς που παρουσιάζουν μεγαλύτερη ποικιλία από την προηγούμενη τεχνική.



Αναπαραγωγή (ή Διασταύρωση) (1/3)

Για δυαδική κωδικοποίηση

- ❖ Αναπαραγωγή είναι η διαδικασία δημιουργίας απογόνων.
 - ❑ Σε αυτή εμπλέκονται ένα σύνολο από τελεστές τελεστών οι οποίοι αντιστοιχούν σε διαδικασίες της βιολογικής εξέλιξης
- ❖ Οι πιο συνηθισμένοι τελεστές είναι η **διασταύρωση** (crossover) και η **μετάλλαξη** (mutation)
 - ❑ Ο τελεστής διασταύρωσης παράγει 2 απογόνους από 2 ακολουθίες-γονείς, αντιγράφοντας επιλεγμένα bit από κάθε γονέα με τρόπο τέτοιο ώστε **το i-οστό bit του απογόνου να είναι το i-οστό bit ενός εκ των γονέων του.**
 - Το ποιος γονέας θα συνεισφέρει το κάθε bit αποφασίζεται βάσει ενός μηχανισμού που ονομάζεται **μάσκα διασταύρωσης** (crossover mask).
 - ❑ Ο τελεστής μετάλλαξης αλλοιώνει ένα χρωμόσωμα του νέου πληθυσμού, μεταβάλλοντας την τιμή κάποιου bit.
- ❖ Οι πιο συνηθισμένοι τελεστές διασταύρωσης και μετάλλαξης είναι:
 - ❑ Διασταύρωση ενός σημείου (single-point crossover)
 - ❑ Διασταύρωση δύο σημείων (two-point crossover)
 - ❑ Ομοιόμορφη διασταύρωση (uniform crossover)
 - ❑ Μετάλλαξη σημείου (point mutation)

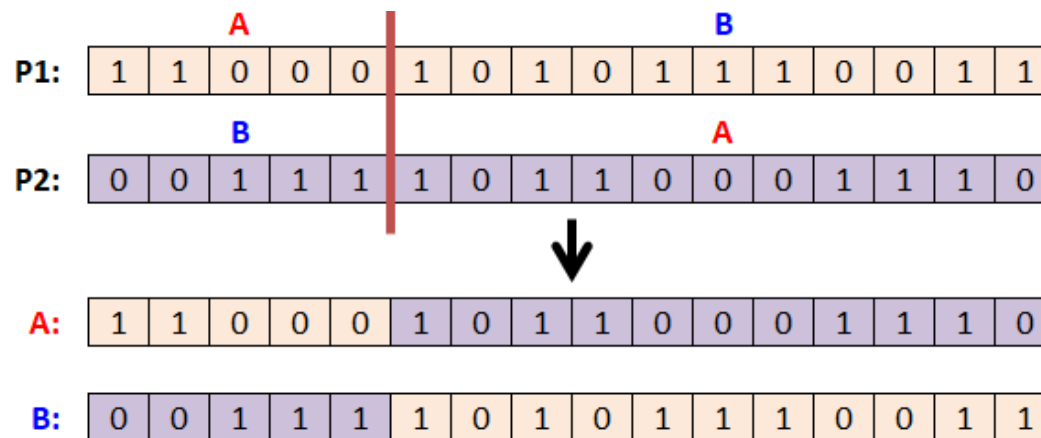


Αναπαραγωγή (2/3)

- ❖ Στη διασταύρωση ενός σημείου, η μάσκα διασταύρωσης ξεκινά με έναν αριθμό συνεχόμενων άσων ("1") που ακολουθούνται από τόσα μηδενικά, όσα χρειάζονται για να συμπληρωθεί η ακολουθία.
 - ❑ Ο ένας απόγονος παίρνει τα bit (το γενετικό υλικό) που δε χρησιμοποιήθηκε στη δημιουργία του άλλου.
- ❖ Για παράδειγμα, η μάσκα διασταύρωσης "11111000000" δείχνει ότι ο ένας απόγονος θα πάρει τα 5 πρώτα bit από τον πρώτο γονέα και τα επόμενα 6 από το δεύτερο γονέα, ενώ ο δεύτερος απόγονος το αντίστροφο.
- ❖ Ιδιότητα της διασταύρωσης ενός σημείου:

Η μέση τιμή των πραγματικών τιμών των δύο γονέων ισούται με τη μέση τιμή των πραγματικών τιμών των δύο απογόνων.

 - ❑ Για παράδειγμα, στο Σχήμα, οι γονείς P1 και P2 αντιστοιχούν στους ακραίους 50547 και 15758 αντίστοιχα και έχουν μέση τιμή 33152.
 - ❑ Οι δύο απόγονοι A και B αντιστοιχούν στους ακραίους 50574 και 15731 αντίστοιχα και έχουν και αυτοί την ίδια μέση τιμή.





Αναπαγωγή (3/3)

Χρωμοσώματα

Μάσκα Διασταύρωσης

Απόγονοι

Διασταύρωση Ενός Σημείου:

11101001000

11111000000

11101010101

00001010101

00001001000

Διασταύρωση Δύο Σημείων:

11101001000

00111110000

00101000101

00001010101

11001011000

Ομοιόμορφη Διασταύρωση:

11101001000

10011010011

10001000100

00001010101

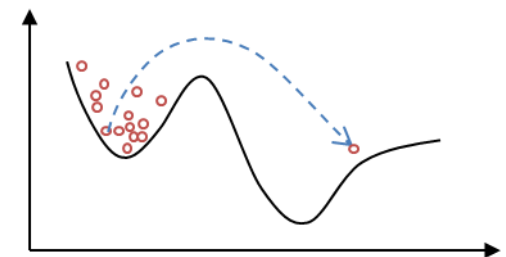
01101011001

Μετάλλαξη Σημείου: 11101001000 → 1110101000



Μετάλλαξη (για δυαδική κωδικοποίηση)

- ❖ Δεν παράγει νέο χρωμόσωμα (λύση) αλλά μεταβάλλει ένα υπάρχον
- ❖ **Αναστροφή ενός σημείου**: στο χρωμόσωμα που αποφασίστηκε (με κάποια πιθανότητα) να γίνει μετάλλαξη, επιλέγεται τυχαία ένα bit και αλλάζει τιμή από 0 σε 1 και αντίστροφα
- ❖ Οι λύσεις που κωδικοποιούν μπορεί να διαφέρουν.
 - ❑ Για παράδειγμα, έστω το χρωμόσωμα 00000001 που αντιστοιχεί στον ακέραιο αριθμό 1 και λόγω μετάλλαξης το πρώτο bit από 0 γίνεται 1.
 - ❑ Το παραγόμενο χρωμόσωμα 10000001 μπορεί να διαφέρει μόνο κατά ένα bit από το αρχικό, αντιστοιχεί όμως σε πολύ διαφορετικό ακέραιο (129)
- ❖ Ο λόγος ύπαρξης του τελεστή μετάλλαξης είναι η εισαγωγή *ποικιλομορφίας (diversity)*.
 - ❑ Οι διάφοροι μηχανισμοί αναπαραγωγής, έχουν την τάση αργά ή γρήγορα να προκαλούν σύγκλιση των λύσεων προς τις καλύτερες αξιοποιώντας κάποιο καλό χρωμόσωμα που υπάρχει εκεί (**exploitation**).
 - ❑ Η μετάλλαξη, διαφοροποιώντας ένα χρωμόσωμα, πιθανώς να απομακρύνει την αναζήτηση από τυχόν περιοχή τοπικού ακρότατου δημιουργώντας έτσι εστίες αναζήτησης σε άλλες περιοχές του χώρου των λύσεων (**exploration**)
- ❖ Άλλοι μηχανισμοί Μετάλλαξης για δυαδική κωδικοποίηση:
 - ❑ Πολλαπλή αναστροφή, Μετάλλαξη ανταλλαγής





Αναπαραγωγή πραγματικών αριθμών

- ❖ Στην κωδικοποίηση πραγματικών αριθμών κάθε χρωμόσωμα γονέας αποτελείται από έναν ή περισσότερους πραγματικούς αριθμούς, ανάλογα με το πρόβλημα.
- ❖ Θα πρέπει να γίνει μετατροπή των πραγματικών αριθμών σε δυαδικούς κατάλληλου μήκους, καθώς και συνεχείς μετατροπές μεταξύ των δύο συστημάτων αρίθμησης στους κύκλους λειτουργίας του αλγόριθμου.
 - ☐ Άρα υπάρχει αυξημένο υπολογιστικό κόστος.
- ❖ Υπάρχουν μηχανισμοί διασταύρωσης που λειτουργούν απευθείας με πραγματικούς αριθμούς
 - ☐ Αριθμητική διασταύρωση
 - ☐ BLX-a (Blend crossover a)
 - ☐ Γραμμική διασταύρωση
 - ☐ Προσομοιωμένη δυαδική διασταύρωση
- ❖ **Λεπτομέρειες στο βιβλίο**



Κωδικοποίηση ακεραίων αριθμών

- ❖ Αν η κωδικοποίηση των χρωμοσωμάτων περιλαμβάνει ακέραιους αριθμούς και όχι πραγματικούς, μπορεί να χρησιμοποιηθούν οι τελεστές διασταύρωσης της δυαδικής κωδικοποίησης αφού πρώτα καθοριστεί ένα πλαίσιο μετατροπής μεταξύ ακεραίων και δυαδικών.
- ❖ Οι τελεστές διασταύρωσης της κωδικοποίησης πραγματικών αριθμών δεν μπορούν να χρησιμοποιηθούν καθώς δεν παράγουν ακέραιους αριθμούς.
- ❖ **Λεπτομέρειες στο βιβλίο**

Κωδικοποίηση διάταξης

- ❖ Στην κωδικοποίηση διάταξης, οι λύσεις κωδικοποιούνται ως μια ακολουθία (διάταξη) συγκεκριμένων ακεραίων που δεν επαναλαμβάνονται.
 - ❑ π.χ. traveling salesman problem
- ❖ Οι μηχανισμοί αναπαραγωγής της κωδικοποίησης πραγματικών αριθμών δεν είναι κατάλληλοι γιατί παράγουν άλλους αριθμούς, τη στιγμή που στην κωδικοποίηση διάταξης μας ενδιαφέρει η διάταξη συγκεκριμένου συνόλου ακεραίων που έχει καθοριστεί εξ' αρχής.
- ❖ Μπορεί να εφαρμοστεί η αναπαραγωγή τύπου διασταύρωσης (crossover) αρκεί να αντιμετωπιστεί το πρόβλημα επανάληψης των ίδιων αριθμών
- ❖ **Λεπτομέρειες στο βιβλίο**



Δημιουργία Επόμενης Γενιάς

- ❖ Είναι το τελευταίο βήμα στον κύκλο λειτουργίας ενός Γενετικού Αλγορίθμου (ΓΑ)
- ❖ Η μία ακραία προσέγγιση είναι η αντικατάσταση όλων των ατόμων με άτομα-απογόνους.
- ❖ Εναλλακτική προσέγγιση στο θέμα της εξέλιξης του πληθυσμού είναι η μερική ανανέωση (*partial replacement*) ατόμων μεταξύ διαδοχικών γενεών.
- ❖ **Χάσμα γενεών** (generation gap): Το ποσοστό των λύσεων (χρωμοσωμάτων) κάθε γενιάς που ανανεώθηκε προς το σύνολο του πληθυσμού (N).
 - ☐ Στους τυπικούς ΓΑ ο συντελεστής αυτός ισούται με τη μονάδα. ($G=1$)
 - ☐ Τάση: **μερική ανανέωση** (*steady-state replacement*). ($G=1/N$)
- ❖ Η μέθοδος της μερικής ανανέωσης προσεγγίζει πιο πολύ στην πραγματικότητα, αφού εκεί συνυπάρχουν πάντα σε κάποιο βαθμό οι διαφορετικές γενεές.
 - ☐ Μάλιστα δίνεται η δυνατότητα στους απογόνους να ανταγωνιστούν τους γονείς τους, επικρατώντας και πάλι ο καλύτερος.



Σύγκλιση Πληθυσμού

- ❖ **Σύγκλιση:** Η επικράτηση ενός χρωμοσώματος ή μικρών παραλλαγών του σε μεγάλο ποσοστό στον πληθυσμό.
 - ❑ Με έναν αποδοτικό γενετικό αλγόριθμο, ο πληθυσμός θα πρέπει μετά από αρκετές επαναλήψεις να συγκλίνει προς το ολικό μέγιστο.
- ❖ Προβλήματα:
 - ❑ **Πρόωρη σύγκλιση** (*premature convergence*) του πληθυσμού γύρω από κάποιο χρωμόσωμα, το οποίο όμως αποτελεί τοπικό μέγιστο και δε μπορεί να ξεφύγει
 - Η αναζήτηση εστιάζει σε λύσεις με παρόμοιο γενετικό υλικό με τις τρέχουσες καλές λύσεις, η ποικιλομορφία του πληθυσμού των υποψήφιων λύσεων μειώνεται γρήγορα
 - ❑ **Αργή σύγκλιση** (*slow convergence*) όπου μετά από ένα μεγάλο αριθμό επαναλήψεων ο πληθυσμός εξακολουθεί να μην έχει συγκλίνει.
 - Δεν ευνοούνται επαρκώς τα καλά χρωμοσώματα με αποτέλεσμα να μην υπάρχει γρήγορη πρόοδος στη βελτίωση της καλύτερης τρέχουσας λύσης
- ❖ Στην πράξη ο γενετικός αλγόριθμος πρέπει να κινηθεί κάπου μεταξύ αυτών των δύο ακραίων περιπτώσεων.
 - ❑ Καθορίζεται από τον τρόπο επιλογής γονέων



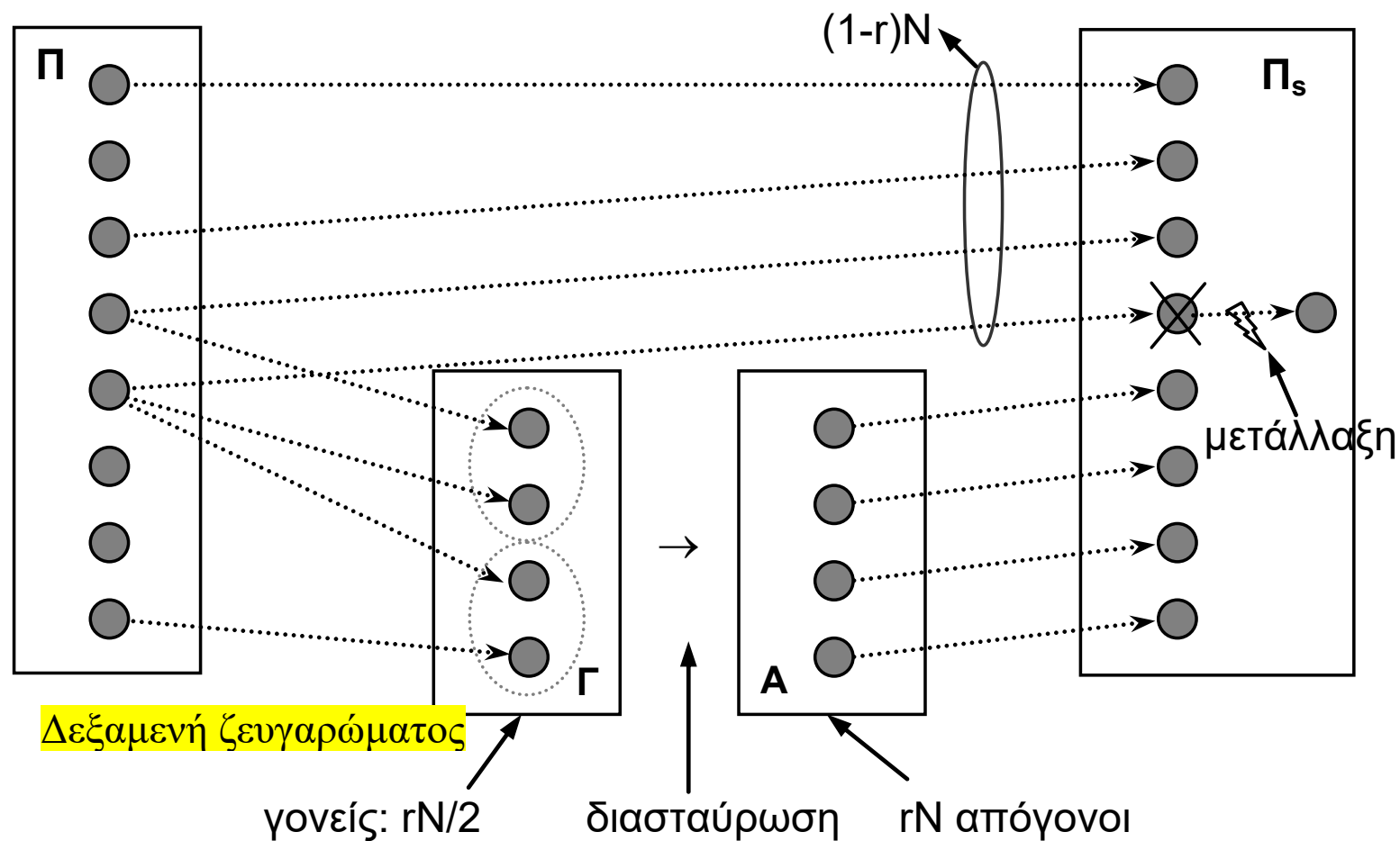
Γενικός γενετικός αλγόριθμος

Περιγραφή (1/2)

- ❖ Στον αλγόριθμο αυτόν, σε κάθε επανάληψη παράγεται μια νέα γενιά χρωμοσωμάτων (υποψ. Λύσεων) P' που προκύπτει μέσω δεδομένης διαδικασίας από τον αρχικό πληθυσμό P .
- ❖ Αρχικός και τελικός πληθυσμός έχουν μέγεθος N αλλά ένα ποσοστό r του αρχικού πληθυσμού αντικαθίσταται σε κάθε κύκλο.
- ❖ Πρώτα επιλέγεται ένας συγκεκριμένος αριθμός χρωμοσωμάτων του υπάρχοντα πληθυσμού για να συμπεριληφθεί απευθείας στην επόμενη γενιά (**μερική ανανέωση**).
- ❖ Στη συνέχεια, παράγονται επιπλέον μέλη μέσω μιας διαδικασίας διασταύρωσης.
- ❖ Τόσο η επιλογή των μελών της υπάρχουσας γενιάς που θα συνεχίσουν και στην επόμενη όσο και των χρωμοσωμάτων-γονέων γίνεται πιθανοκρατικά.
- ❖ Στο σημείο αυτό επιλέγεται τυχαία ένα ποσοστό m του πληθυσμού αυτού στο οποίο θα συμβούν τυχαίες μεταλλάξεις.

Γενικός γενετικός αλγόριθμος

Σχηματική λειτουργία





Γενικός γενετικός αλγόριθμος

Περιγραφή (2/2)

- ❖ Άρα σε ένα γενετικό αλγόριθμο πρέπει να ορισθούν οι ακόλουθες παράμετροι:
 - ☐ Καταλληλότητα: η συνάρτηση καταλληλότητας.
 - ☐ Όριο_Καταλληλότητας: η ελάχιστη τιμή καταλληλότητας που πρέπει να επιτευχθεί από ένα χρωμόσωμα, ώστε να τερματίσει η διαδικασία.
 - ☐ N: ο αριθμός των χρωμοσωμάτων του πληθυσμού.
 - ☐ r: το ποσοστό του πληθυσμού που αντικαθίσταται σε κάθε κύκλο.
 - ☐ m: το ποσοστό μετάλλαξης.
- ❖ Τα βήματα του αλγορίθμου δίνονται στη συνέχεια:



Γενικός γενετικός αλγόριθμος

Γενετικός_Αλγόριθμος (Καταλληλότητα, Όριο_Καταλληλότητας, N , r , m)

1. Αρχικοποίησε τον Πληθυσμό: $\Pi \leftarrow$ Τυχαία παραγωγή N χρωμοσωμάτων
2. Αξιολόγησε: $\forall x \in \Pi$ υπολόγισε την ποσότητα Καταλληλότητα(x)
3. Επανέλαβε: Όσο $\max(\text{Καταλληλότητα}(x), \forall x \in \Pi) < \text{Όριο_Καταλληλότητας}$
 - i. Επέλεξε πιθανοκρατικά $(1-r)N$ μέλη του Π και πρόσθεσέ τα στο Π' .
Η πιθανότητα $P(x_i)$ να επιλεγεί ένα χρωμόσωμα x από το Π δίνεται από τη σχέση:
$$P(x_i) = \frac{\text{Καταλληλότητα}(x_i)}{\sum_{j=1}^N \text{Καταλληλότητα}(x_j)}$$
 - ii. Διασταύρωση: Πιθανοκρατικά επέλεξε $rN/2$ ζεύγη χρωμοσωμάτων από το Π σύμφωνα με την πιθανότητα $P(x_i)$.
 \forall ζεύγος (x_1, x_2) , παρήγαγε δύο απογόνους με εφαρμογή του τελεστή διασταύρωσης. Πρόσθεσε τους απογόνους στο Π' .
 - iii. Μετάλλαξη: Επέλεξε $m\%$ από τα μέλη του Π' με ομοιόμορφη πιθανότητα. Για κάθε ένα, αντέστρεψε ένα τυχαία επιλεγμένο bit.
 - iv. Ενημέρωσε: $\Pi \leftarrow \Pi'$.
 - v. $\forall x \in \Pi$ υπολόγισε την ποσότητα Καταλληλότητα(x)
4. Επέστρεψε το χρωμόσωμα με τη μεγαλύτερη Καταλληλότητα



Παράδειγμα (1/2)

- ❖ Έστω ένας πληθυσμός αποτελείται από 4 χρωμοσώματα.
- ❖ Στην 3^η στήλη φαίνονται οι τιμές της συνάρτησης καταλληλότητας, ενώ στην 4^η οι πιθανότητες επιλογής προς αναπαραγωγή κάθε χρωμοσώματος, ανηγμένες στο διάστημα $[0, 1]$.
 - ❑ Οι τιμές της 4^{ης} στήλης προκύπτουν από το λόγο της καταλληλότητας του χρωμοσώματος προς το άθροισμα των καταλληλοτήτων όλων των χρωμοσωμάτων του πληθυσμού.

Υποψήφιος λύσεις	bit-string	Καταλληλότητα	Πιθανότητα επιλογής
A	000110010111	8	0.32
B	111010101100	6	0.24
Γ	001110101001	6	0.24
Δ	111011011100	5	0.20

- ❖ Έστω ότι το αποτέλεσμα της διαδικασίας επιλογής είναι το **[A,B,B,Γ]** και
 - ❑ τα ζευγάρια που σχηματίζονται τυχαία είναι τα **(B,A)** και **(B,Γ)** και
 - ❑ ως τελεστής αναπαραγωγής του πρώτου ζευγαριού επιλέγεται η διασταύρωση ενός σημείου με μάσκα διασταύρωσης "111100000000" ενώ το για το δεύτερο ο ίδιος τελεστής με μάσκα διασταύρωσης "11111111000"



Παράδειγμα (2/2)

Πρώτο ζευγάρι: μάσκα διασταύρωσης "111100000000"

Δεύτερο ζευγάρι: μάσκα διασταύρωσης "111111111000"

Γονέας 1	(B) 1110-10101100	(B) 1110-10101100	(B) 111010101-100	(B) 111010101-100
Γονέας 2	(A) 0001-10010111	(A) 0001-10010111	(Γ) 001110101-001	(Γ) 001110101-001
Απόγονος	1 ^{ος}	2 ^{ος}	3 ^{ος}	4 ^{ος}
Διασταύρωση	1110-10010111	0001-10101100	111010101-001	001110101-100
Μετάλλαξη	1110-10010111	0001-10101100	111 <u>1</u> 0101-001	001110101-10 <u>1</u>



Εφαρμογές

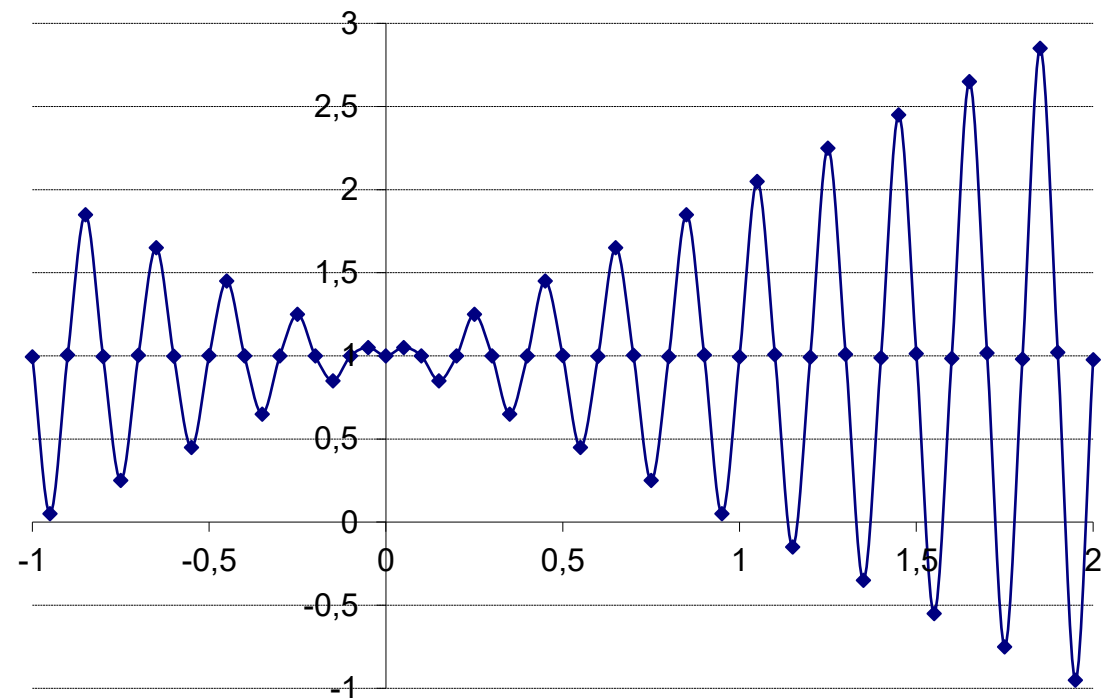
- ❖ Οι γενετικοί αλγόριθμοι, ο γενετικός προγραμματισμός αλλά και άλλες εξελικτικές τεχνικές, είναι κατάλληλες μέθοδοι επίλυσης προβλημάτων σε τομείς που έχουν κάποιο ή κάποια από τα ακόλουθα χαρακτηριστικά:
 - ☐ Ο χώρος των λύσεων είναι ιδιαίτερα μεγάλος (συνήθως εκθετικής πολυπλοκότητας) για να ελεγχθεί εξαντλητικά.
 - ☐ Είναι εύκολο να παραχθούν λύσεις, όχι απαραίτητα καλής ποιότητας.
 - ☐ Ο τρόπος αλληλεπίδρασης των παραμέτρων του προβλήματος είναι άγνωστος ή ελάχιστα γνωστός.
 - ☐ Το πρόβλημα δεν είναι εύκολο να μοντελοποιηθεί μαθηματικά ώστε να λυθεί με αναλυτικό τρόπο.
 - ☐ Είναι αποδεκτή μια προσεγγιστική λύση.
 - ☐ Μικρές βελτιώσεις έναντι λύσεων από άλλες μεθοδολογίες επίλυσης αποφέρουν μεγάλο όφελος (π.χ. οικονομικό).
 - ☐ Υπάρχουν εξομοιωτές που μπορούν να δοκιμάσουν την επίδοση μιας δυνητικά λύσης αλλά όχι τρόποι για την απευθείας δημιουργία τέτοιων λύσεων.
 - Για παράδειγμα, υπάρχουν μαθηματικά μοντέλα που υπολογίζουν τις επιδόσεις μιας κεραίας με δεδομένα χαρακτηριστικά αλλά όχι το αντίθετο, που να μας λένε δηλαδή πώς ακριβώς πρέπει να είναι μια κεραία για να έχει επιθυμητή συμπεριφορά.



Παραδείγματα

Εύρεση μεγίστου συνάρτησης μιας μεταβλητής

- ❖ $f(x)=x \cdot \sin(10\pi \cdot x)+1.0$, στο διάστημα $[-1, 2]$
- ❖ Η αντιμετώπιση του προβλήματος αυτού αναλυτικά (δηλαδή μηδενισμός πρώτης παραγώγου, κλπ) είναι δύσκολη.
 - ❑ Το πρόβλημα μπορεί να αντιμετωπισθεί με χρήση γενετικών αλγορίθμων.
- ❖ Χρήση δυαδικής αναπαράστασης





Το πρόβλημα του πλανόδιου πωλητή

- ❖ Εύρεση της σειράς με την οποία ένας πωλητής πρέπει να περάσει από όλες τις πόλεις ενός συνόλου πόλεων και να επιστρέψει στην αρχική, ώστε να έχει το μικρότερο δυνατό κόστος, όπως αυτό εκφράζεται κάθε φορά (π.χ. χρόνος, χρήματα κλπ).
- ❖ Αναπαράσταση με διανύσματα ακεραίων αριθμών μήκους ίσου με το πλήθος των πόλεων. Έτσι, για παράδειγμα, ένα χρωμόσωμα θα έχει τη μορφή $v = \langle i_1 i_2 \dots i_n \rangle$, όπου $i_1, i_2, \dots, i_n \in 1..n$ και $i_j \neq i_k$ για $j \neq k$.
- ❖ Δημιουργία αρχικού πληθυσμού με τυχαίο τρόπο.
- ❖ Απλή συνάρτηση καταλληλότητας.
- ❖ Αναπαραγωγή: Μπορεί να προκύψουν μη-έγκυρα χρωμοσώματα !
 - ☐ Έλεγχος των απογόνων και επιδιόρθωση των μη-έγκυρων χρωμοσωμάτων.
 - ☐ Τροποποιημένες τεχνικές διασταύρωσης και μετάλλαξης, οι οποίες δίνουν πάντα έγκυρα χρωμοσώματα.



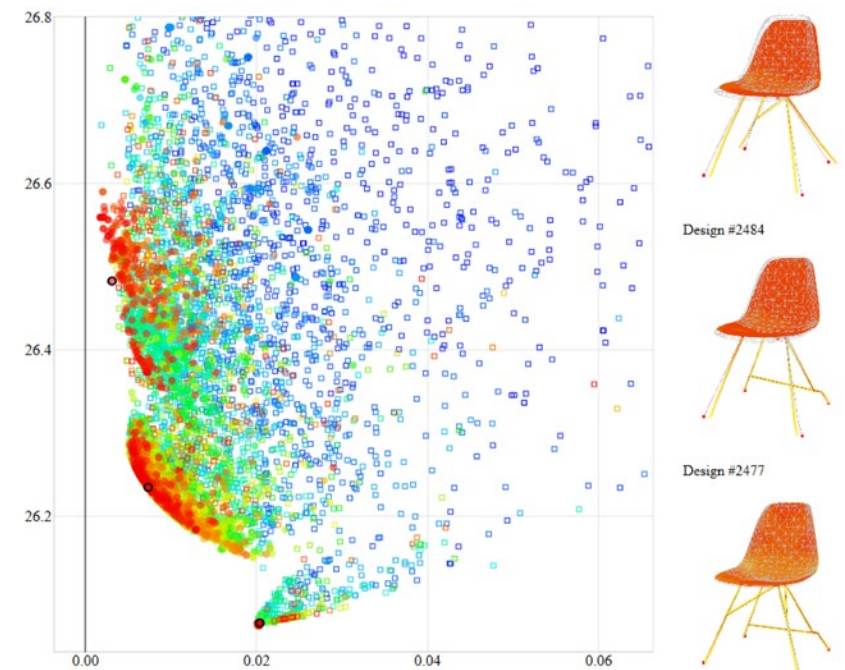
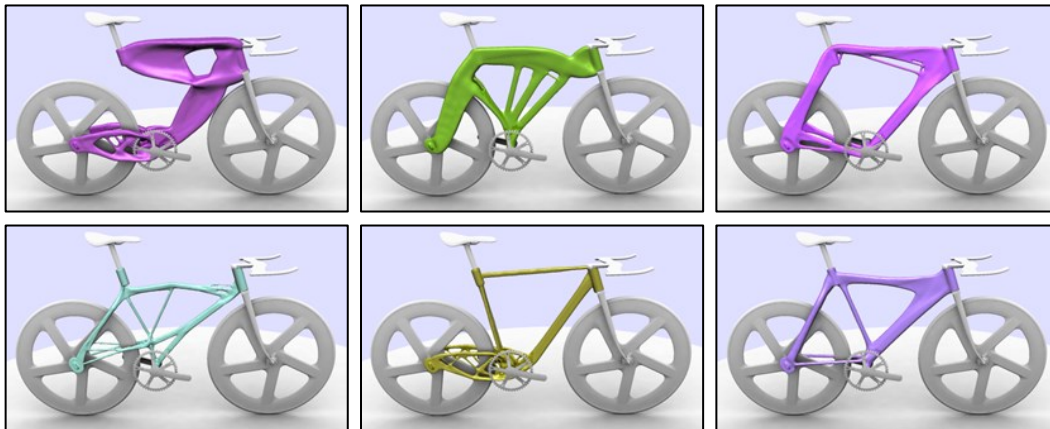
Το πρόβλημα του ταξιδιωτικού σάκου

- ❖ Στην ίδια κατηγορία ανήκουν και τα προβλήματα σακιδίου (*knapsack problems*) όπου δοθέντος ενός συνόλου από αντικείμενα με βάρος και αξία, πρέπει να καθοριστεί το πλήθος αντικειμένων που πρέπει να συμπεριληφθεί στο σακίδιο ώστε το βάρος να μην ξεπεράσει κάποιο όριο και η αξία να μεγιστοποιηθεί.
- ❖ Ένα άλλο πρόβλημα με μεγάλη πρακτική σημασία στη βιομηχανία είναι η αποθήκευση σε κιβώτια (*bin packing*) και αφορά την εύρεση του βέλτιστου τρόπου αποθήκευσης ενός αριθμού αντικειμένων διάφορων όγκων σε κιβώτια (containers) όγκου V , έτσι ώστε να ελαχιστοποιηθεί το πλήθος των απαιτούμενων κιβωτίων.
 - ❑ Το πρόβλημα του σακιδίου μπορεί να θεωρηθεί ως ειδική περίπτωση αυτής της κατηγορίας (μεγιστοποίηση της αξίας των αντικειμένων που χωρούν σε δεδομένο χώρο).
- ❖ Τέλος στην κατηγορία αυτών των εφαρμογών εντάσσονται και τα προβλήματα καταμερισμού εργασιών (*job-shop scheduling*) και κατασκευής ωρολογίων προγραμμάτων (*timetabling*).



Δημιουργική Σχεδίαση (generative design)

- ❖ Η διαδικασία σχεδίασης αντικειμένων όπου ο δημιουργός θέτει κάποιους γενικούς περιορισμούς (σταθερά σημεία, σημεία πρόσδεσης, περιορισμούς όγκου ή έκτασης, αντοχή σε δυνάμεις, κτλ.) και μια διαδικασία δημιουργεί "αντικείμενα" που συμφωνούν με τους περιορισμούς
- ❖ Οι ΓΑ, μπορεί να χρησιμοποιηθούν στη σχεδίαση κατασκευών και εξαρτημάτων (κεραίες, αγωγοί με ιδιαίτερη διατομή, ακροφύσια κινητήρων, αεροδυναμικές επιφάνειες, γέφυρες, μηχανολογικά εξαρτήματα, έπιπλα, κ.α.) όπου ζητούμενο μπορεί να είναι τόσο η εύρεση μιας λύσης, όσο και η βελτιστοποίησή της.





Αποτελεσματικότητα και Αποδοτικότητα

❖ Θεώρημα των σχημάτων

- ❑ Ένα *σχήμα* (*scheme*) είναι ένα χρωμόσωμα που δημιουργείται από το αλφάβητο $\{0, 1, \#\}$, όπου το σύμβολο $\#$ υποδηλώνει οποιαδήποτε τιμή.
- ❑ Ένα χρωμόσωμα λέγεται ότι περιέχει ένα σχήμα, εάν οι δύο συμβολοσειρές ταυτίζονται για όλα τα διάφορα από το $\#$ στοιχεία του σχήματος.
- ❑ *Τάξη* (*order*) ενός σχήματος το πλήθος των συμβόλων που είναι διάφορα από το $\#$.
- ❑ *Μήκος* (*length*) ενός σχήματος η απόσταση των δύο πιο εξωτερικών μη- $\#$ συμβόλων.

❖ Η υψηλή τιμή καταλληλότητας ενός χρωμοσώματος οφείλεται στο γεγονός ότι αυτό περιέχει "καλά" σχήματα.

❖ Αποδίδοντας πιθανότητα αναπαραγωγής στα χρωμοσώματα ανάλογη προς την καταλληλότητά τους, τα "καλά" σχήματα τυγχάνουν εκθετικά αυξανόμενου αριθμού αναπαραγωγών στις επόμενες γενεές.

❖ Ο αριθμός των σχημάτων τα οποία μεταφέρονται από γενεά σε γενεά είναι της τάξης του N^3 , όπου N το μέγεθος του πληθυσμού (*έμμεσος παραλληλισμός, implicit parallelism*).

❖ **Σχετικά γονίδια** είναι αυτά τα οποία συνδυασμένα επηρεάζουν τη συνάρτηση καταλληλότητας.



Εφαρμογές (1/2)

- ❖ Εύρεση μέγιστης τιμής αριθμητικών συναρτήσεων.
 - ❑ Η εύρεση του μέγιστου μιας συνάρτησης δεν είναι καθόλου εύκολη υπόθεση για συναρτήσεις πολλών μεταβλητών, οι οποίες εμφανίζουν ασυνέχειες, θόρυβο, κλπ.
 - ❑ Το πλεονέκτημα που εμφανίζει η εφαρμογή τους σε αυτά τα προβλήματα είναι ότι η συνάρτηση καταλληλότητας είναι δεδομένη.
- ❖ Επεξεργασία εικόνων
 - ❑ Αναγνώριση προτύπων, όπως ακμές, επιφάνειες, ακόμη και αντικείμενα, σε ψηφιοποιημένες εικόνες.
- ❖ Συνδυαστική βελτιστοποίηση.
 - ❑ Το κλασσικό πρόβλημα κατανομής πόρων σε δραστηριότητες, με σκοπό τη μεγιστοποίηση του οφέλους ή την ελάττωση του κόστους.
 - ❑ Ο έλεγχος όλων των υποψήφιων λύσεων να είναι αδύνατος (συνδυαστική έκρηξη)
 - ❑ Γνωστά προβλήματα αυτής της κατηγορίας: του *πλανόδιου πωλητή*, η *αποθήκευση κιβωτίων*, σχεδίαση VLSI κυκλωμάτων, *καταμερισμός εργασιών*, *ωρολόγιο πρόγραμμα*



Εφαρμογές (2/2)

❖ Σχεδίαση

- ☐ Κατασκευών και εξαρτημάτων, με ζητούμενο τόσο την εύρεση μιας λύσης, όσο και τη βελτιστοποίησή της.
- ☐ Οι αλγόριθμοι μπορούν να δοκιμάσουν συνδυασμούς και ιδέες που ο ανθρώπινος νους δε θα δοκίμαζε ποτέ, δίνοντας ενίοτε πρωτότυπα αποτελέσματα.

❖ Μηχανική μάθηση

- ☐ Στα συστήματα μηχανικής μάθησης οι γενετικοί αλγόριθμοι μπορεί να χρησιμοποιηθούν για την προσέγγιση συναρτήσεων.
- ☐ Η πιο γνωστή εφαρμογή είναι αυτή των συστημάτων ταξινόμησης (classifier systems), ωστόσο οι γενετικοί αλγόριθμοι έχουν χρησιμοποιηθεί και σε παιχνίδια, επίλυση λαβυρίνθων, καθώς και για πολιτικές και οικονομικές αναλύσεις.
- ☐ Η μηχανική μάθηση περιγράφεται σε σχετικό κεφάλαιο.

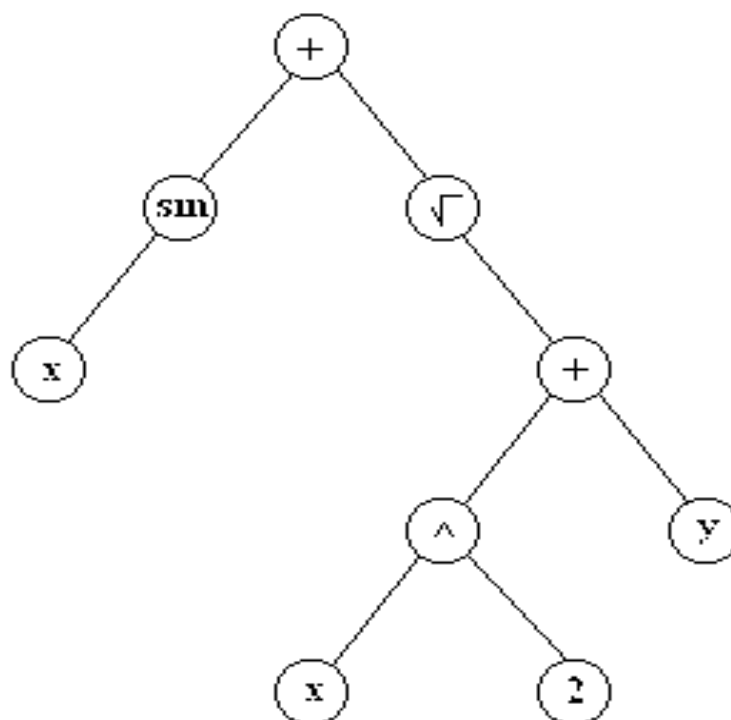


Γενετικός Προγραμματισμός

Αναπαράσταση

- ❖ Αυτόματη κατασκευή προγραμμάτων υπολογιστών.
- ❖ Οι υποψήφιες λύσεις είναι προγράμματα υπολογιστών.

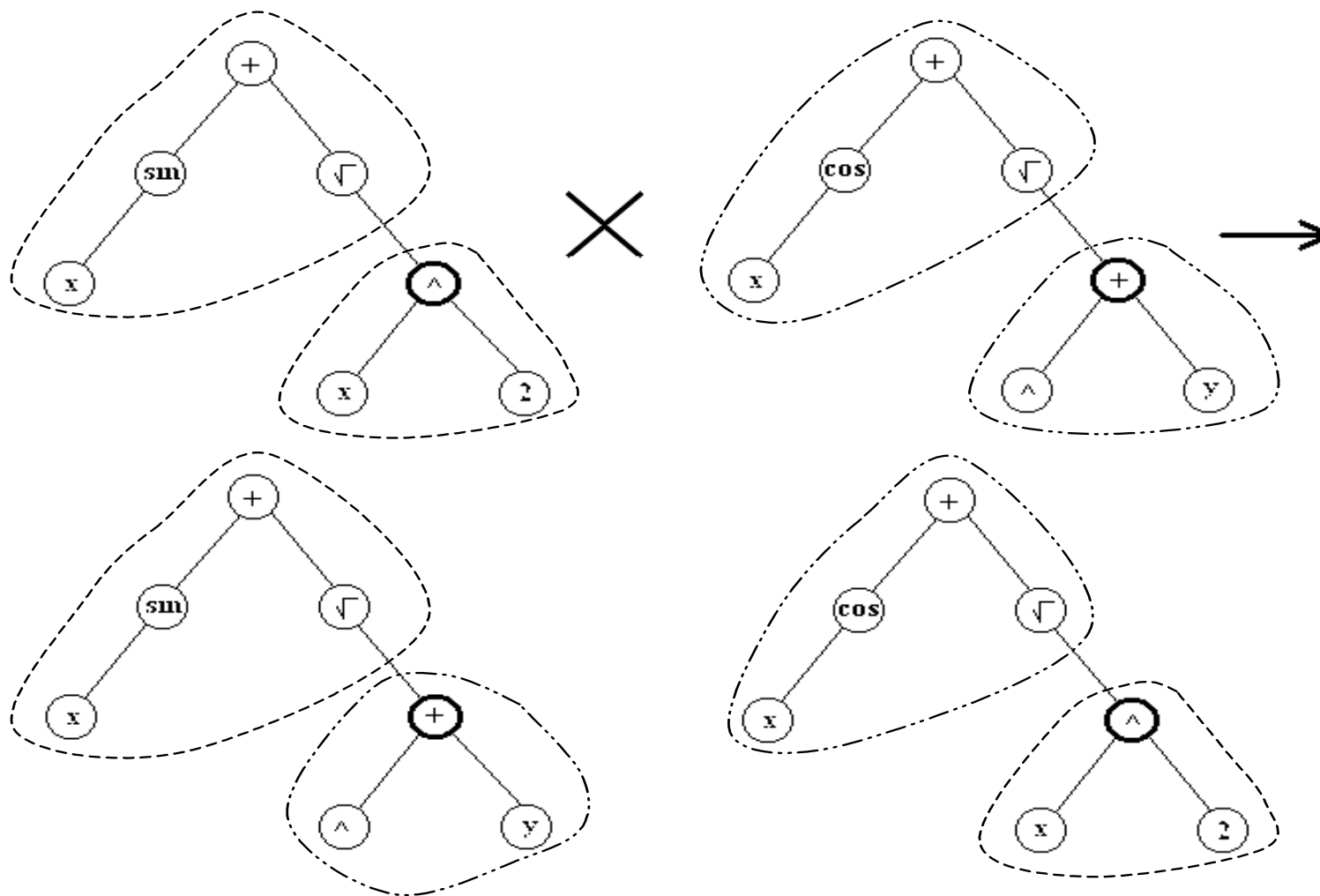
$$\sin(x) + \sqrt{x^2 + y}$$





Γενετικός Προγραμματισμός

Διασταύρωση





Ερωτήσεις

- ❖ Για ποιες περιπτώσεις ενδείκνυνται οι γενετικοί αλγόριθμοι;
- ❖ Οι υποψήφιες λύσεις που παράγονται στους γενετικούς αλγορίθμους, αξιολογούνται από τη συνάρτηση
- ❖ Περιγράψτε τη γενική μορφή ενός γενετικού αλγόριθμου.
- ❖ Ποιοι είναι οι πιο συνηθισμένοι τελεστές διασταύρωσης και μετάλλαξης
- ❖ Σύγκλιση πληθυσμού
- ❖ Τι είναι η μερική ανανέωση
- ❖

Κυκλώστε το αντίστοιχο γράμμα Σ (ωστό) – Λ (άθος) στις επόμενες ερωτήσεις: (Προσοχή Σε αυτό το θέμα υπάρχει αρνητική βαθμολογία)

Κάθε χρωμόσωμα στους γενετικούς αλγορίθμους αποτελεί μια πιθανή λύση του προβλήματος	Σ	Λ
Τα σύμβολα του αλφάβητου των γενετικών αλγορίθμων είναι τα A,G,T,C	Σ	Λ
Τα σύμβολα του αλφάβητου των γενετικών αλγορίθμων είναι τα 0,1	Σ	Λ
Στους γενετικούς αλγορίθμους μόνο ένα μέρος του αρχικού πληθυσμού συμμετέχει στην αναπαραγωγή	Σ	Λ
Η συνάρτηση καταλληλότητας στους Γενετικούς Αλγορίθμους επιστρέφει πραγματικούς αριθμούς	Σ	Λ



Ασκήσεις

- ❖ Έστω σε ένα γενετικό αλγόριθμο, οι υποψήφιος λύσεις x_1, x_2, x_3, x_4 με τιμές καταλληλότητας: $K(x_1) = 0.7, K(x_2) = 0.6, K(x_3) = 0.3$ και $K(x_4) = 0.8$. Να υπολογιστεί η πιθανότητα $p(x_4)$ επιλογής του χρωμοσώματος x_4 χρησιμοποιώντας επιλογή αναλογικής καταλληλότητας (fitness proportionate selection)
- ❖ Απάντηση:
$$p(x_4) = K(x_4) / \sum(K(x_i)) = 0.8 / (0.7 + 0.6 + 0.3 + 0.8) = 0.8 / 2.4 = 1/3$$
- ❖ Έστω σε ένα γενετικό αλγόριθμο, οι υποψήφιος λύσεις x_1, x_2, x_3, x_4 με τιμές καταλληλότητας: $K(x_1) = 0.3, K(x_2) = 0.5, K(x_3) = 0.4$ και $K(x_4) = 0.3$. Να υπολογιστεί η πιθανότητα $p(x_2)$ επιλογής του χρωμοσώματος x_2 χρησιμοποιώντας επιλογή αναλογικής καταλληλότητας (fitness proportionate selection)

- ❖ Σε έναν γενετικό αλγόριθμο, δίνονται οι γονείς A=01111001 και B=00010110. Να επιλέξετε μια μάσκα διασταύρωσης ενός σημείου και να παράγετε τον (τους) απόγονο (απογόνους).
- ❖ Απάντηση:
Έστω η μάσκα διασταύρωσης: 11100000
Οι απόγονοι είναι α) 01111001 x 00010110 = 01110110
 β) 01111001 x 00010110 = 00011001
- ❖ Σε έναν γενετικό αλγόριθμο, δίνονται οι γονείς A=10110111 και B=00011101. Ορίστε μια μάσκα διασταύρωσης 2 σημείων και παράγετε τους απογόνους.
- ❖ Σε έναν γενετικό αλγόριθμο δίνονται τρία μέλη του αρχικού πληθυσμού A=01001001, B=10011100 και Γ =10001001 και δύο νέα μέλη Δ=01011001 και Ε=10001100 που προήλθαν από τη διασταύρωση δύο εκ των τριών μελών του αρχικού πληθυσμού (χωρίς μετάλλαξη). Ποιοι ήταν οι δύο γονείς και με ποια μάσκα διασταύρωσης προέκυψαν τα Δ και Ε?



Ασκήσεις

- ❖ Έστω το πρόβλημα N-βασιλισσών στο σκάκι. Να επιλυθεί με χρήση των γενετικών αλγορίθμων. Συγκεκριμένα, να κωδικοποιηθεί το πρόβλημα, δηλαδή να περιγραφούν οι μεταβλητές του προβλήματος και η συνάρτηση καταλληλότητας. Στη συνέχεια, να γίνει υλοποίηση σε κάποια γλώσσα προγραμματισμού και να λυθεί. Να πειραματιστείτε με διάφορες κωδικοποιήσεις και παραμέτρους του αλγορίθμου επίλυσης.
- ❖ Να κατασκευαστεί πρόγραμμα σε οποιαδήποτε γλώσσα προγραμματισμού για την εύρεση με εφαρμογή γενετικών αλγορίθμων, του μεγίστου της συνάρτησης $f(x) = x \cdot \sin(10 \cdot \pi \cdot x) + 1$ στο διάστημα πραγματικών αριθμών $[-1, 2]$. Το πρόγραμμα να είναι παραμετροποιήσιμο ως προς το μήκος των χρωμοσωμάτων, το μέγεθος του αρχικού πληθυσμού, την πιθανότητα μετάλλαξης και τον αριθμό των επαναλήψεων που εκτελεί. Να εκτελεστεί για διάφορες τιμές των παραμέτρων και να σχολιαστούν τα αποτελέσματα.
- ❖ Να κατασκευαστεί πρόγραμμα σε οποιαδήποτε γλώσσα προγραμματισμού για την επίλυση με εφαρμογή γενετικών αλγορίθμων του προβλήματος του πλανόδιου πωλητή. Να γίνει εφαρμογή σε πρόβλημα με 10 πόλεις.