



Αριστοτέλειο
Πανεπιστήμιο
Θεσσαλονίκης

Τεχνητή Νοημοσύνη

Γενετικοί Αλγόριθμοι

Ιώαννης Βλαχάβας

Τμήμα Πληροφορικής ΑΠΘ

Άδειες Χρήσης

Το παρόν εκπαιδευτικό υλικό υπόκειται σε άδειες χρήσης Creative Commons. Για εκπαιδευτικό υλικό, όπως εικόνες, που υπόκειται σε άλλου τύπου άδειας χρήσης, η άδεια χρήσης αναφέρεται ρητώς.



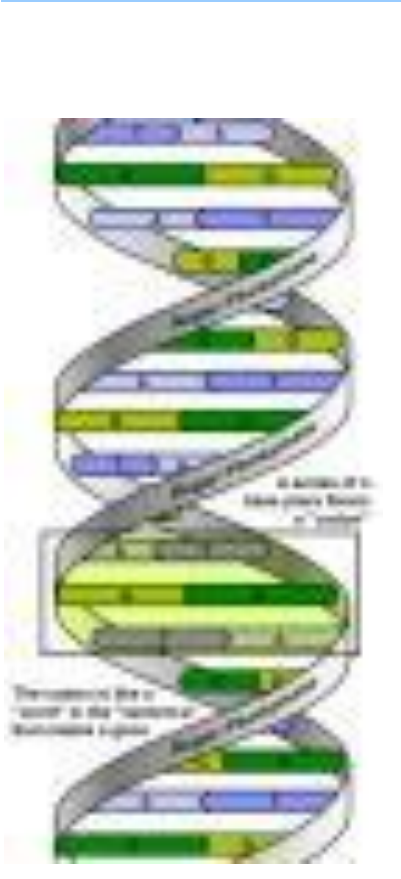
Χρηματοδότηση

Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα. Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης**» έχει χρηματοδοτήσει μόνο τη αναδιαμόρφωση του εκπαιδευτικού υλικού.



Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.





Γενετικοί Αλγόριθμοι

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.
- ❖ 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 (evolution) της φύσης.
- ❖ 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), γνωστική μοντελοποίηση (cognitive modeling), προβλήματα εφοδιαστικής (logistics), προβλήματα πλανόδιου πωλητή, προβλήματα βέλτιστου ελέγχου, βελτιστοποίηση ερωτήσεων σε βάσεις δεδομένων, κτλ.

Ιστορικά στοιχεία

- ❑ Το 1958 ο Friedberg, επιχείρησε να συνδυάσει μικρά προγράμματα FORTRAN, ωστόσο τα προγράμματα που προέκυψαν τις περισσότερες φορές δεν ήταν εκτελέσιμα.
- ❑ Το 1975 ο Holland έδωσε νέα ώθηση στο χώρο, χρησιμοποιώντας σειρές bits για να αναπαραστήσει λειτουργίες, με τρόπο τέτοιο ώστε κάθε συνδυασμός bits να είναι μια έγκυρη λειτουργία.

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 υπερσχύουν 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.

Γενική Μορφή Γενετικού Αλγόριθμου (1/2)

❖ **Βήματα:**

- 1) Δημιούργησε τυχαία έναν αρχικό πληθυσμό Π , με N υποψήφια (μη αποδεκτές, δηλαδή μη έγκυρες ή μη βέλτιστες, κλπ) λύσεις.
 - 2) Βαθμολόγησε (δηλ. πόσο κοντά σε μια αποδεκτή λύση είναι) κάθε υποψήφια λύση χρησιμοποιώντας μια *συνάρτηση καταλληλότητας (fitness function)*.
 - 3) Σχημάτισε $N/2$ ζευγάρια όχι απαραίτητα μοναδικών γονέων (δηλαδή μια λύση μπορεί να είναι γονέας περισσότερες φορές), δίνοντας μεγαλύτερη προτεραιότητα στις πλέον κατάλληλες λύσεις.
 - 4) Κάθε ζευγάρι *ζευγαρώνει* (mates), δίνοντας δύο νέες λύσεις, τους *απογόνους* (offsprings).
 - 5) 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 και στοιχεία του αρχικού πληθυσμού Π .
 - 6) Η διαδικασία επαναλαμβάνεται για το νέο πληθυσμό Π'
- ❖ Οι πιο συνηθισμένες συνθήκες τερματισμού της είναι η εύρεση μιας τέλει λύσης με βάση τη συνάρτηση καταλληλότητας ή η σύγκλιση όλων των λύσεων σε μια.
- ❖ Αναζήτηση στο χώρο των υποψηφίων λύσεων: Αναρρίχηση λόφου (hill climbing)

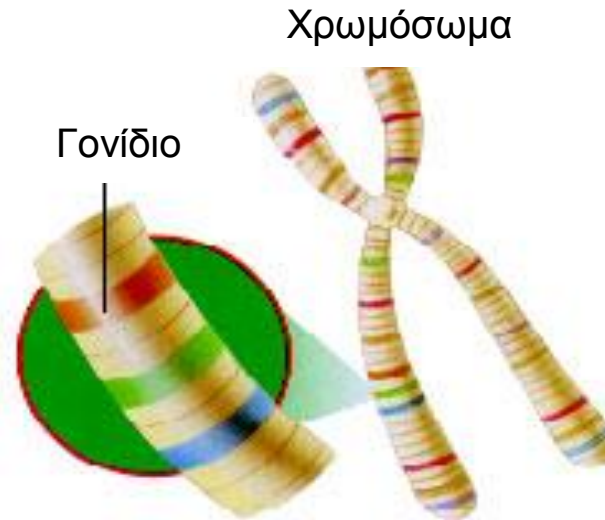
Γενική Μορφή Γενετικού Αλγόριθμου (2/2)

1. Δημιούργησε έναν αρχικό πληθυσμό Π , με N υποψήφια 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 $\Pi = \Pi'$

1. Ένας γενετικός αλγόριθμος για ένα συγκεκριμένο πρόβλημα περιλαμβάνει **πέντε συστατικά**:
 - Δημιουργία αρχικού πληθυσμού (συνήθως δημιουργείται τυχαία)
 - Αναπαράσταση λύσεων
 - Συνάρτηση καταλληλότητας
 - Επιλογή γονέων
 - Διαδικασία αναπαραγωγής

Αναπαράσταση Υποψήφιων Λύσεων (1/3)

- ❖ 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 χρωμο-σωμάτων.

Αναπαράσταση Υποψήφιων Λύσεων (2/3)

- ❖ 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.

Αναπαράσταση Υποψήφιων Λύσεων (3/3)

- ❖ 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.

Συνάρτηση Καταλληλότητας

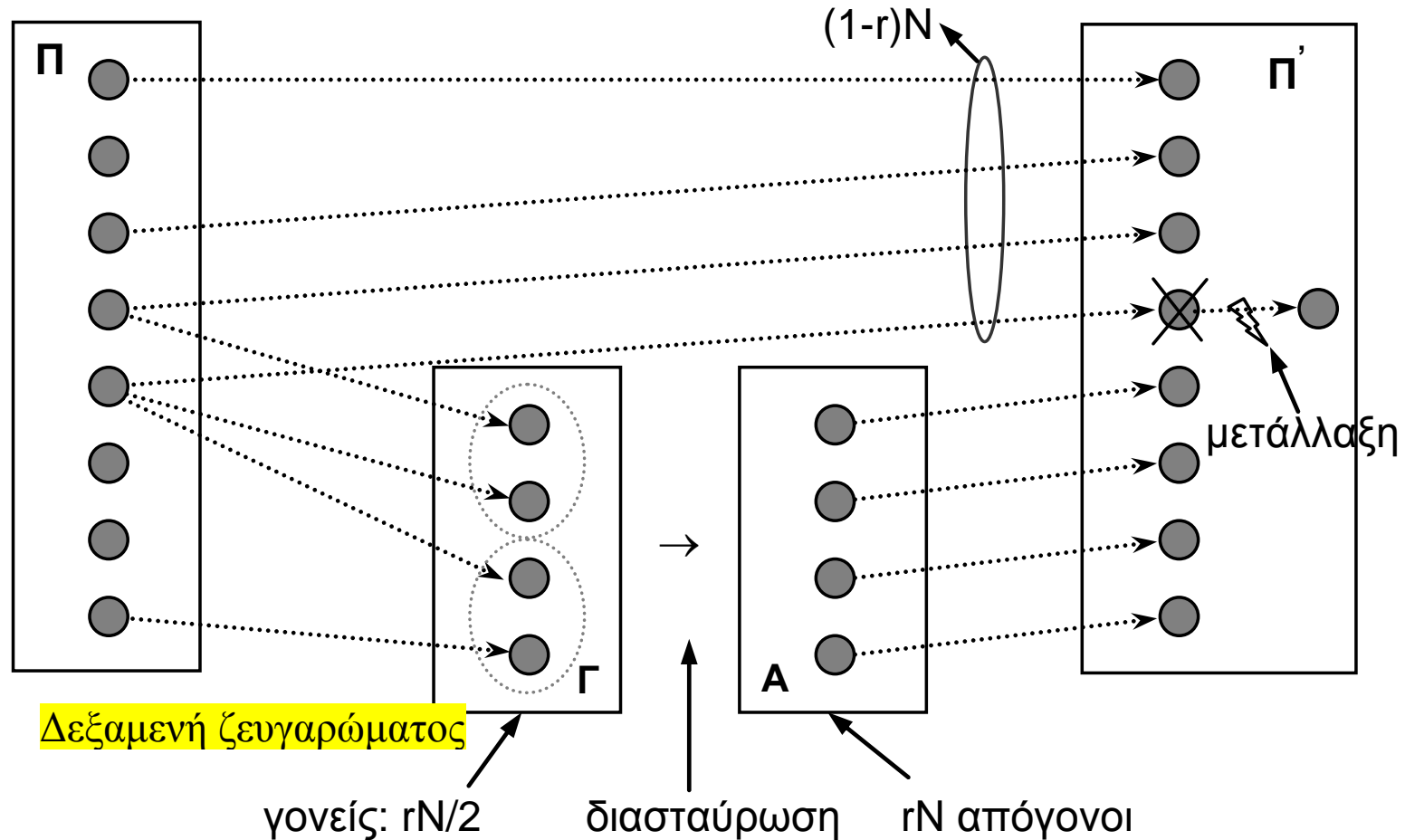
- ❖ Δέχεται ως είσοδο ένα χρωμόσωμα και επιστρέφει έναν αριθμό (συνήθως στο διάστημα $[0,1]$), που υποδηλώνει το πόσο κατάλληλο είναι.
- ❖ Η αξιολόγηση αυτή χρησιμοποιείται:
 - ❑ είτε από τη συνθήκη τερματισμού ή
 - ❑ από τη διαδικασία της πιθανοκρατικής επιλογής τους για να συμπεριληφθούν (ή όχι) στον 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.

Διαδικασία Επιλογής Γονέων (1/3)

- ❖ Κατά τη διαδικασία επιλογής, αρχικά οι υποψήφιος 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**)
- ❖ **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 (2/3)

- ❖ Μια τεχνική είναι η επιλογή **αναλογικής καταλληλότητας** (*fitness proportionate selection*) όπου η πιθανότητα επιλογής ενός χρωμοσώματος x_i , υπολογίζεται από τη σχέση:

$$P(x_i) = \frac{\text{Καταλληλότητα}(x_i)}{\sum_{j=1}^N \text{Καταλληλότητα}(x_j)}$$

- ❖ Δηλαδή, η πιθανότητα να επιλεγεί ένα χρωμόσωμα είναι ευθέως ανάλογη της καταλληλότητάς του.
- ❖ 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 ότι οι την προηγούμενη τεχνική.

Διαδικασία Επιλογής Γονέων (3/3)

Τεχνική της ρουλέτας

- ❖ Ένας τρόπος υλοποίησης της αναλογικής καταλληλότητας.
 1. Παράγεται το άθροισμα S όλων των τιμών αξιολόγησης των υποψηφίων λύσεων.
 2. Επιλέγεται ένας τυχαίος αριθμός n , από το 0 μέχρι το S , χρησιμοποιώντας συνάρτηση ομοιόμορφης κατανομής για τη δημιουργία των τυχαίων αριθμών.
 3. Επαναληπτικά εξετάζεται κάθε υποψήφια λύση και η τιμή του προστίθεται σε έναν καταχωρητή K .
 4. Αν η τιμή του K γίνει μεγαλύτερη ή ίση του n , η λύση επιλέγεται και ο K μηδενίζεται. Στην αντίθετη περίπτωση εκτελείται πάλι το 3.
 5. Αν δεν έχει επιλεγεί ικανοποιητικός αριθμός υποψηφίων λύσεων εκτελείται το 2, αλλιώς ο αλγόριθμος τερματίζει.
- ❖ Οι υποψήφιες 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 υποψήφιες λύσεις μπορεί να επιλεγούν περισσότερο από μία φορές.

Αναπαραγωγή (1/3)

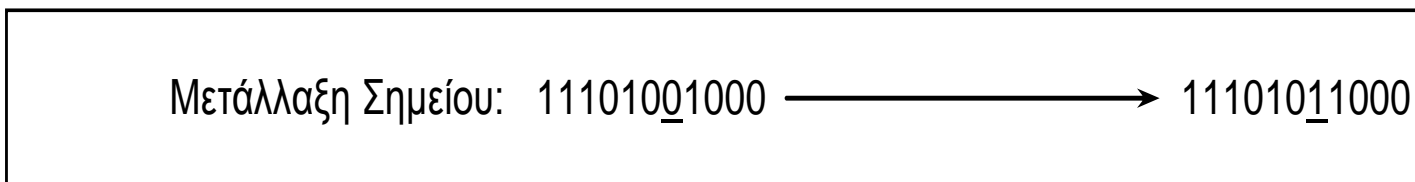
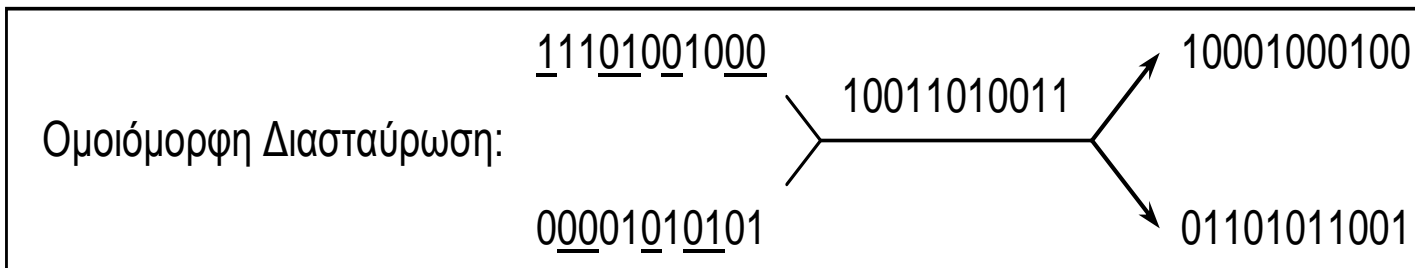
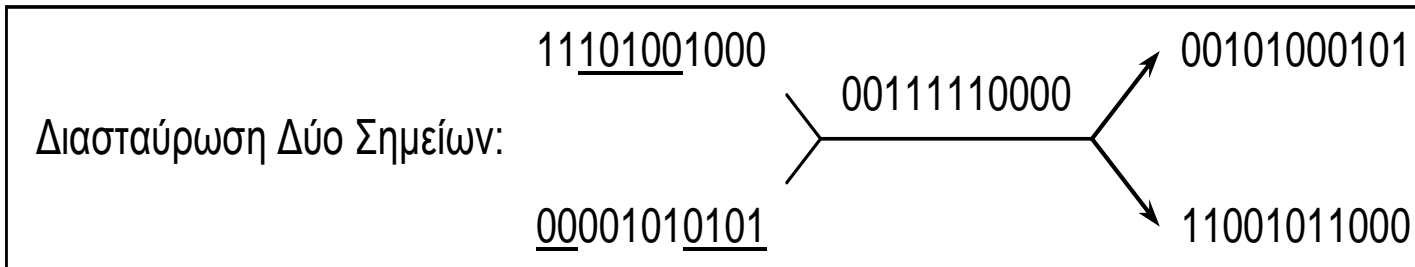
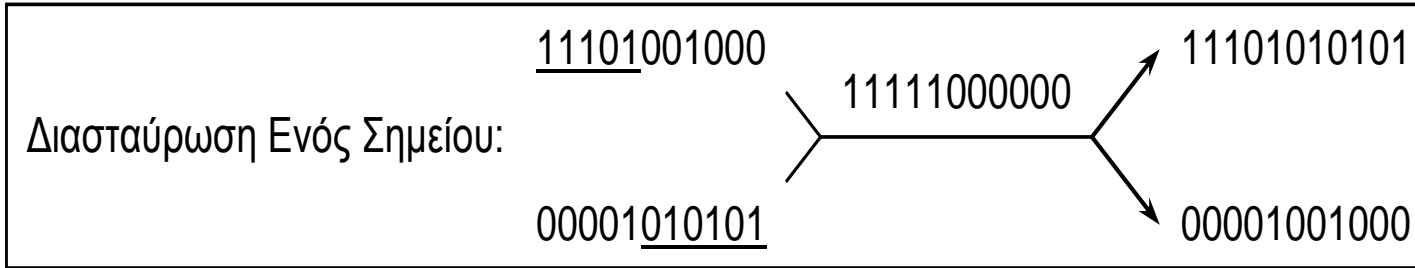
- ❖ Αναπαραγωγή είναι η διαδικασία δημιουργίας απογόνων.
 - ❑ Σε αυτή εμπλέκονται ένα σύνολο από τελεστές τελεστών οι οποίοι αντιστοιχούν σε διαδικασίες της βιολογικής εξέλιξης.
- ❖ Οι πιο συνηθισμένοι τελεστές είναι η **διασταύρωση** (**crossover**) και η **μετάλλαξη** (**mutation**).
 - ❑ Ο τελεστής 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 άσπων ("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.

Αναπαραγωγή (3/3)

Χρωμοσώματα Μάσκα Διασταύρωσης Απόγονοι



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.
- ❖ Ένα γονίδιο συγκλίνει, όταν έχει την ίδια τιμή στο 95% των χρωμοσωμάτων.
- ❖ Ένας πληθυσμός συγκλίνει, όταν όλα τα γονίδιά του έχουν συγκλίνει.
- ❖ Προβλήματα:
 - ❑ Πρόωρη σύγκλιση (premature convergence) του πληθυσμού γύρω από κάποιο χρωμόσωμα, το οποίο όμως αποτελεί τοπικό μέγιστο και δε μπορεί να ξεφύγει.
 - ❑ Αργή σύγκλιση (slow convergence) όπου μετά από ένα μεγάλο αριθμό επαναλήψεων, ο πληθυσμός εξακολουθεί να μην έχει συγκλίνει.
- ❖ Εναλλακτική προσέγγιση στο θέμα της εξέλιξης του πληθυσμού είναι η μη ανανέωση ολόκληρου του πληθυσμού σε κάθε γενιά, αλλά ενός μέρους αυτού.
- ❖ **Χάσμα γενεών** (generation gap): Το ποσοστό των χρωμοσωμάτων κάθε γενιάς που ανανεώθηκε, προς το σύνολο των χρωμοσωμάτων.
 - ❑ Στους τυπικούς ΓΑ ο συντελεστής αυτός ισούται με τη μονάδα.
 - ❑ Τάση: μέθοδος μερικής ανανέωσης (*steady-state replacement*).

Μερική Ανανέωση Πληθυσμού

- ❖ 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.
- ❖ Υπάρχουν δύο θέματα στην προσέγγιση της μερικής ανανέωσης: πώς θα επιλεγούν οι γονείς που θα ζευγαρώσουν και πώς θα επιλεγεί ισάριθμος αριθμός γονέων που θα αποχωρήσουν, ώστε να αφήσουν χώρο για τους απογόνους.
- ❖ Οι προσεγγίσεις που ακολουθούνται είναι οι ακόλουθες τρεις:
 - ❑ Επιλογή των γονέων προς αναπαραγωγή με πιθανότητα ανάλογη προς την καταλληλότητά τους και τυχαία επιλογή των γονέων που θα αποχωρήσουν.
 - ❑ Επιλογή των γονέων προς αναπαραγωγή τυχαία και επιλογή των γονέων που θα αποχωρήσουν με πιθανότητα αντιστρόφως ανάλογη προς την καταλληλότητά τους.
 - ❑ Επιλογή των γονέων προς αναπαραγωγή με πιθανότητα ανάλογη προς την καταλληλότητά τους και επιλογή των γονέων που θα αποχωρήσουν με πιθανότητα αντιστρόφως ανάλογη προς την καταλληλότητά τους.

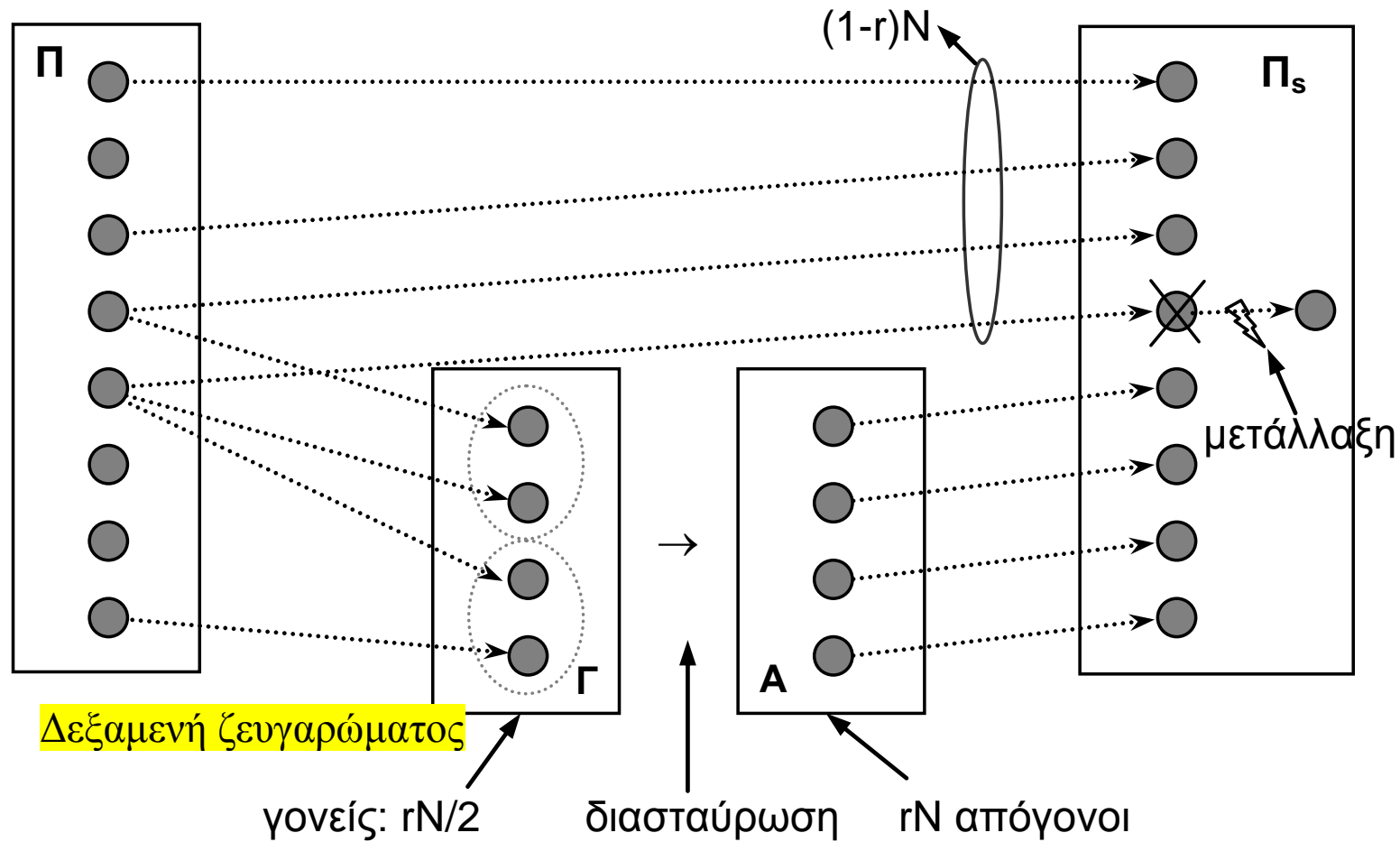
Γενικός γενετικός αλγόριθμος

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 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.
- ❖ Στο Σχήμα, το Γ είναι η δεξαμενή ζευγαρώματος από όπου, με κάποια διαδικασία διασταύρωσης, προκύπτει το σύνολο A των απογόνων.

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. Αρχικοποίηση του Πληθυσμού: $\Pi \leftarrow$ Τυχαία παραγωγή N χρωμοσωμάτων
2. Αξιολόγηση: $\forall x \in \Pi$ υπολόγισε την ποσότητα Καταλληλότητα(x)
3. Επανάλαβε: 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. Ενημέρωση: $\Pi \leftarrow \Pi'$.
 - 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.

Υποψήφιες 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 "111111111000"

Παράδειγμα (2/2)

Πρώτο ζευγάρι: μάσκα διασταύρωσης "111100000000"

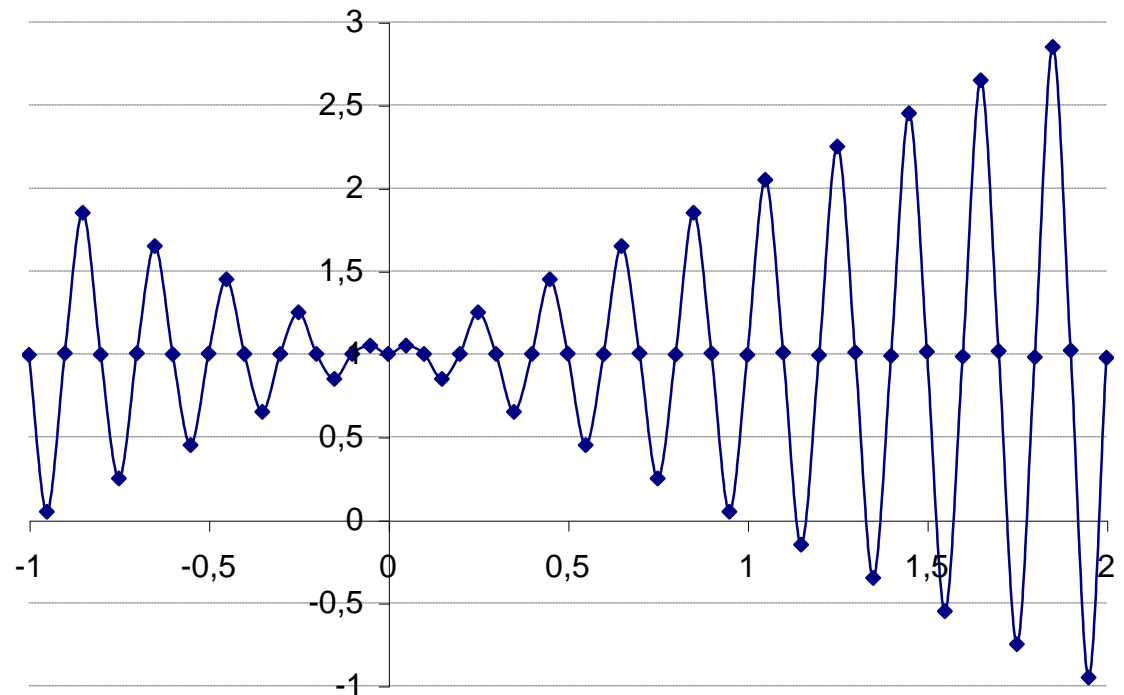
Δεύτερο ζευγάρι: μάσκα διασταύρωσης "111111111000"

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	<i>1110-10010111</i>	<i>0001-10101100</i>	<i>111010101-001</i>	<i>001110101-100</i>
Μετάλλαξη	<i>1110-10010111</i>	<i>0001-10101100</i>	<i>111<u>1</u>10101-001</i>	<i>001110101-10<u>1</u></i>

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.

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

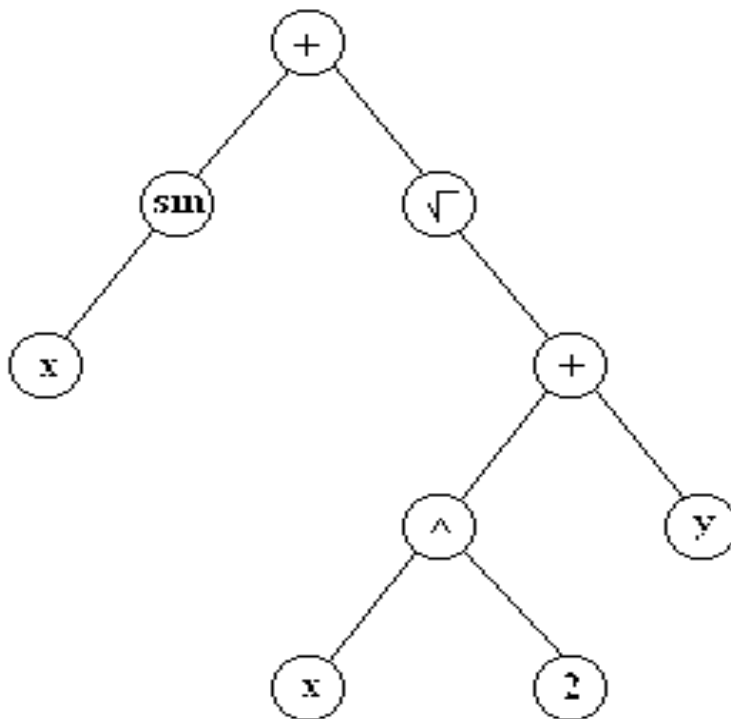
- ❑ Στα συστήματα μηχανικής μάθησης οι γενετικοί αλγόριθμοι μπορεί να χρησιμοποιηθούν για την προσέγγιση συναρτήσεων.
- ❑ Η πιο γνωστή εφαρμογή είναι αυτή των συστημάτων ταξινόμησης (classifier systems), ωστόσο οι γενετικοί αλγόριθμοι έχουν χρησιμοποιηθεί και σε παιχνίδια, επίλυση λαβυρίνθων, καθώς και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

Γενετικός Προγραμματισμός

Αναπαράσταση

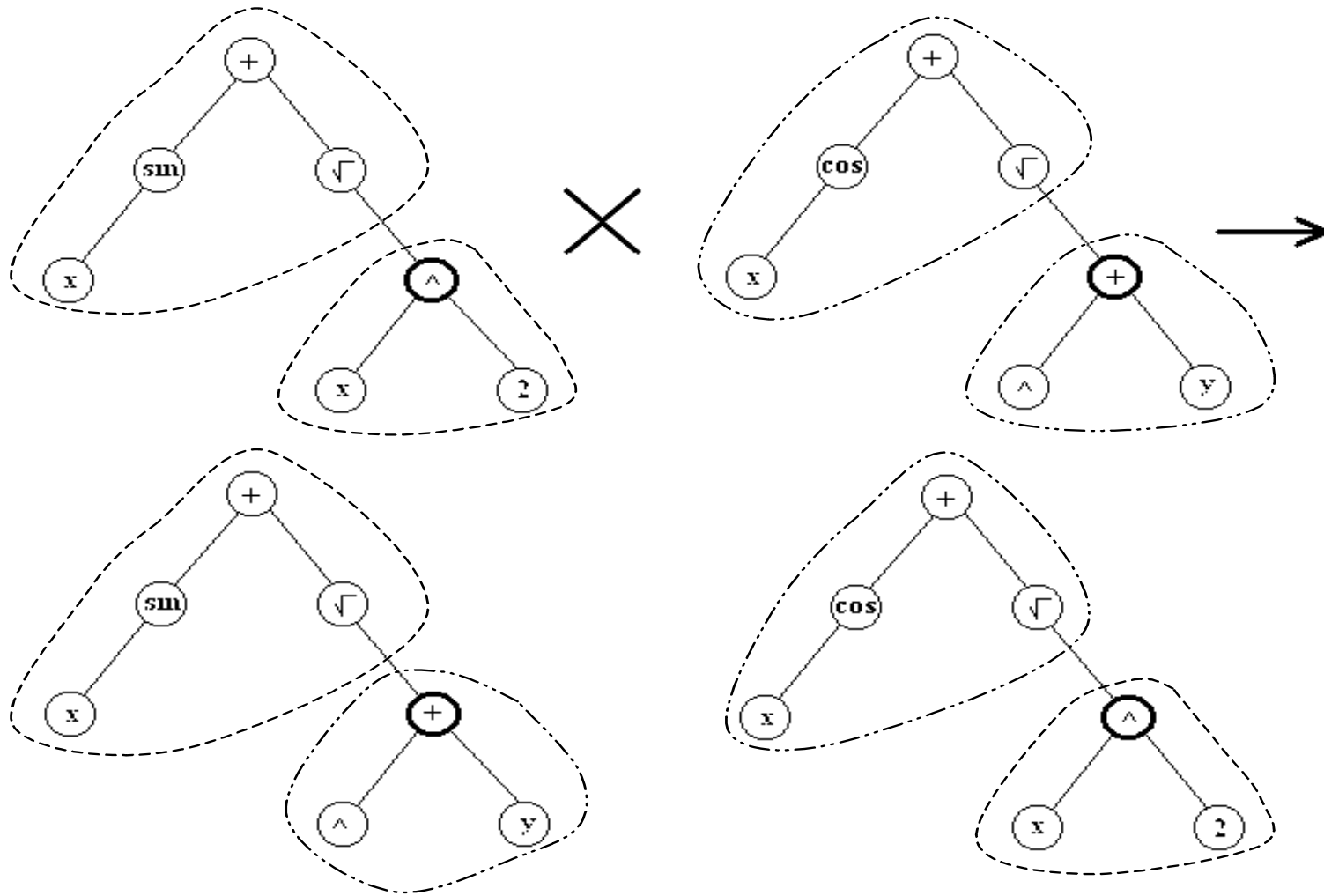
- ❖ Αυτόματη κατασκευή προγραμμάτων υπολογιστών.
- ❖ Οι υποψήφιες λύσεις είναι προγράμματα υπολογιστών.

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



Γενετικός Προγραμματισμός

Διασταύρωση



Ασκήσεις

- ❖ 6.1. Έστω σε ένα γενετικό αλγόριθμο, οι υποψήφιες 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)
- ❖ 6.2 Έστω σε ένα γενετικό αλγόριθμο, οι υποψήφιες 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)
- ❖ 6.3. 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).
- ❖ 6.4. Σε έναν γενετικό αλγόριθμο, δίνονται οι γονείς $A=10110111$ και $B=00011101$. Ορίστε μια μάσκα διασταύρωσης 2 σημείων και παράγετε τους απογόνους.