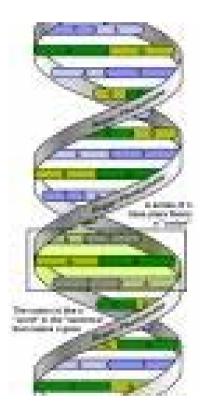
Κεφάλαιο 7



Γενετικοί Αλγόριθμοι

Τεχνητή Νοημοσύνη - Β' Έκδοση

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



Εισαγωγή

- * Σε αρκετές περιπτώσεις το μέγεθος ενός προβλήματος καθιστά απαγορευτική τη χρήση κλασικών μεθόδων αναζήτησης για την επίλυσή του.
- * 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.



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 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.



Γενική Μορφή Γενετικού Αλγόριθμου (1/2)

Βήματα:

- 1) Δημιούργησε τυχαία έναν αρχικό πληθυσμό Π, με Ν υποψήφιες (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 Ν/2 ζευγάρια όχι απαραίτητα μοναδικών γονέων, 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.
- 6) Η διαδικασία επαναλαμβάνεται για το νέο πληθυσμό Π'
- 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.



Γενική Μορφή Γενετικού Αλγόριθμου (2/2)

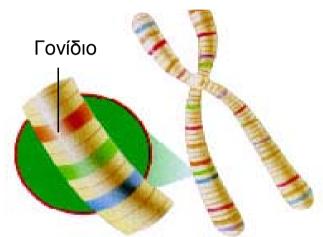
- 1. Δημιούργησε έναν αρχικό πληθυσμό Π, με Ν υποψήφιες λύσεις.
- 2. Υπολόγισε την καταλληλότητα κάθε λύσης.
- 3. Όσο δεν ισχύει κάποια συνθήκη τερματισμού:
 - α. Επανέλαβε Ν/2 φορές τα ακόλουθα βήματα:
 - ί. Επέλεξε δύο λύσεις από τον πληθυσμό Π.
 - ii. Συνδύασε τις δύο λύσεις για να βγάλεις δύο απογόνους.
 - 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 Π=Π'
- 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 (1/2)

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

Χρωμόσωμα



* 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).



Αναπαράσταση Υποψήφιων Λύσεων (2/2)

- * 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).**
 - 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.
- * 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).

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.



Διαδικασία Επιλογής Γονέων (1/2)

- Απόδοση πιθανοτήτων επιλογής προς αναπαραγωγή στα μέλη ενός πληθυσμού υποψηφίων λύσεων.
 - □ 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.
- * Κατά τη διαδικασία επιλογής, αρχικά οι υποψήφιες λύσεις αντιγράφονται σε μια δεξαμενή ζευγαρώματος (mating pool).
 - Σε αυτήν αντιγράφονται μέλη του αρχικού πληθυσμού, με πιθανότητα ανάλογη της καταλληλότητάς τους.
 - □ Για την επιλογή των χρωμοσωμάτων που θα αντιγραφούν στη δεξαμενή ζευγαρώματος χρησιμοποιούνται αρκετές τεχνικές.



Διαδικασία Επιλογής Γονέων (2/2)

Τεχνική της ρουλέτας

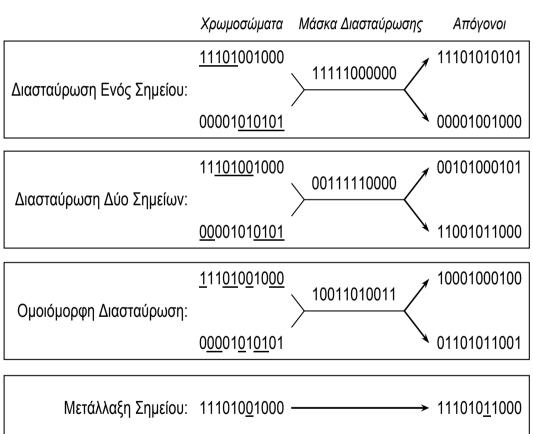
- 1. Παράγεται το άθροισμα S όλων των τιμών αξιολόγησης των υποψηφίων λύσεων.
- 2. Επιλέγεται ένας τυχαίος αριθμός η, από το 0 μέχρι το S, χρησιμοποιώντας συνάρτηση ομοιόμορφης κατανομής για τη δημιουργία των τυχαίων αριθμών.
- 3. Επαναληπτικά εξετάζεται κάθε υποψήφια λύση και η τιμή του προστίθεται σε έναν καταχωρητή Κ.
- 4. Αν η τιμή του Κ γίνει μεγαλύτερη ή ίση του n, η λύση επιλέγεται και ο Κ μηδενίζεται. Στην αντίθετη περίπτωση εκτελείται πάλι το 3.
- 5. Αν δεν έχει επιλεγεί ικανοποιητικός αριθμός υποψηφίων λύσεων εκτελείται το 2, αλλιώς ο αλγόριθμος τερματίζει.
- Αλλη τεχνική: επιλογή αναλογικής καταλληλότητας (fitness proportionate selection)
 - Πιθανότητα επιλογής χρωμοσώματος xi:

$$P(x_i) = \frac{K ατολληλότητα(x_i)}{\sum_{i=1}^{N} K ατολληλότητα(x_j)}$$



Αναπαραγωγή

- Αναπαραγωγή είναι η διαδικασία δημιουργίας απογόνων.
 - Σε αυτή εμπλέκονται ένα σύνολο από τελεστές οι οποίοι αντιστοιχούν σε διαδικασίες της βιολογικής εξέλιξης.
- **Φ** Οι πιο συνηθισμένοι τελεστές είναι:
 - Διασταύρωση ενός σημείου (singlepoint crossover)
 - Διασταύρωση δύο σημείων (two-point crossover)
 - Ομοιόμορφη διασταύρωση (uniform crossover)
 - Μετάλλαξη σημείου (point mutation)





Σύγκλιση Πληθυσμού

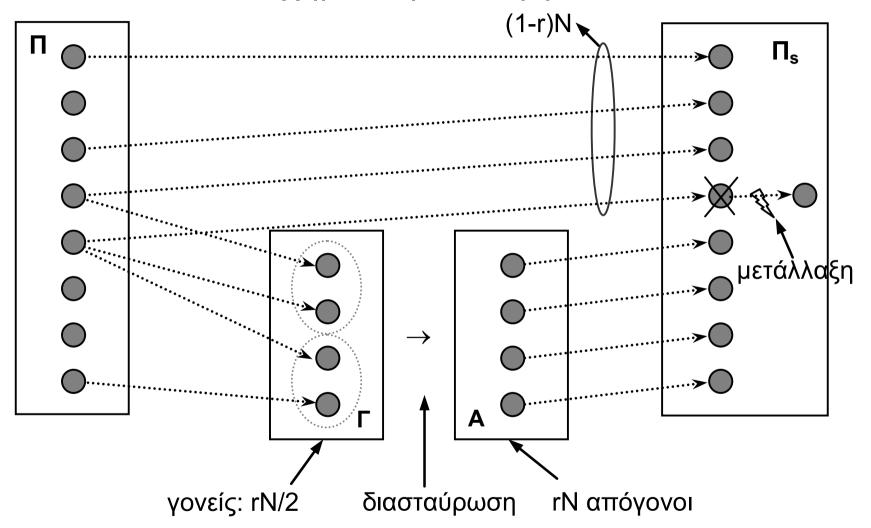
- **Σύγκλιση**: Η επικράτηση ενός χρωμοσώματος ή μικρών παραλλαγών του, σε μεγάλο ποσοστό στον πληθυσμό.
 - Με έναν αποδοτικό γενετικό αλγόριθμο, ο πληθυσμός θα πρέπει μετά από αρκετές επαναλήψεις να συγκλίνει προς το ολικό μέγιστο.
- Ένα γονίδιο συγκλίνει, όταν έχει την ίδια τιμή στο 95% των χρωμοσωμάτων.
- ❖ Ένας πληθυσμός συγκλίνει, όταν όλα τα γονίδιά του έχουν συγκλίνει.
- ❖ Εναλλακτική προσέγγιση στο θέμα της εξέλιξης του πληθυσμού είναι η μη ανανέωση ολόκληρου του πληθυσμού σε κάθε γενιά, αλλά ενός μέρους αυτού.

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.



Σχηματική λειτουργία





Περιγραφή (1/2)

- * Στον αλγόριθμο αυτόν, σε κάθε επανάληψη παράγεται μια νέα γενιά χρωμοσωμάτων Πε που προκύπτει μέσω δεδομένης διαδικασίας από τον αρχικό πληθυσμό Π.
- Αρχικός και τελικός πληθυσμός έχουν μέγεθος Ν αλλά ποσοστό r του αρχικού πληθυσμού αντικαθίσταται σε κάθε κύκλο.
- Τρώτα 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, Γ είναι η δεξαμενή ζευγαρώματος από όπου, με κάποια διαδικασία διασταύρωσης, προκύπτει το σύνολο Α των απογόνων.



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.
- **Φ** Τα βήματα του αλγορίθμου δίνονται στη συνέχεια:



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

- 1. Αρχικοποίησε τον Πληθυσμό: Π←Τυχαία παραγωγή Ν χρωμοσωμάτων
- 2. Αξιολόγησε: ∀x∈Π υπολόγισε την ποσότητα Καταλληλότητα(x)
- 3. Επανέλαβε: Όσο \max (Καταλληλότητα (x) , $\forall x \in \Pi$) < Όριο Καταλληλότητας
 - i. Επέλεξε πιθανοκρατικά (1-r)N μέλη του Π και πρόσθεσέ τα στο Πs. Η πιθανότητα $P(x_i)$ να επιλεγεί ένα χρωμόσωμα x από το Π δίνεται από τη σχέση:

$$P(x_i) = \frac{K\alpha \tau \alpha \lambda \lambda \eta \lambda \dot{\eta} \tau \alpha(x_i)}{\sum_{j=1}^{N} K\alpha \tau \alpha \lambda \lambda \eta \lambda \dot{\eta} \tau \alpha(x_j)}$$

- ii. Δ ιασταύρωση: Πιθανοκρατικά επέλεξε rN/2 ζεύγη χρωμοσωμάτων από το Π σύμφωνα με την πιθανότητα $P(\mathbf{x}_i)$.
 - \forall ζεύγος $(\mathbf{x}_1,\mathbf{x}_2)$, παρήγαγε δύο απογόνους με εφαρμογή του τελεστή διασταύρωσης. Πρόσθεσε τους απογόνους στο Π s.
- iii.Μετάλλαξη: Επέλεξε m% από τα μέλη του Πs με ομοιόμορφη πιθανότητα. Για κάθε ένα, αντέστρεψε ένα τυχαία επιλεγμένο bit.
- iv. Ενημέρωσε: Π←Πs.
- v. ∀x∈Π υπολόγισε την ποσότητα Καταλληλότητα <math>(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
Α	000110010111	8	0.32
В	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
 - **α** 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)

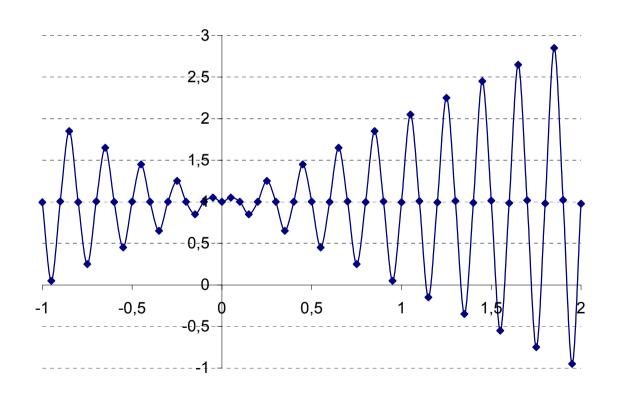
Γονέας 1	(B) 1110-10101100	(B) 1110-10101100	(B) 111010101-100	(B) 111010101-100
Γονέας 2	(A) 0001-10010111	(A) 0001-10010111	(Γ) 001110101-001	(Γ) 001110101-001
Απόγονος	1 ^{oç}	2 ° ⁵	3 ^{oç}	4 ^{oç}
Διασταύρωση	<i>1110</i> -10010111	0001-10101100	111010101-001	001110101- <i>100</i>



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 $v=\langle i_1i_2...i_n\rangle$, όπου $i_1,\,i_2,\,...,\,i_n\in 1..n$ 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.



Εφαρμογές (1/2)

- **Φ** 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.
- **Φ** 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 ενίστε 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.

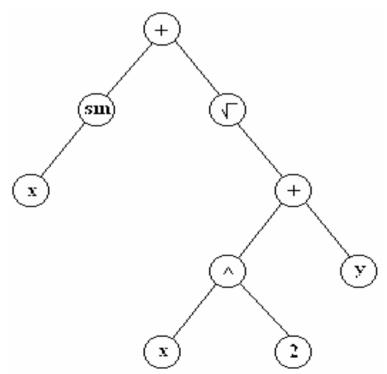


Γενετικός Προγραμματισμός

Αναπαράσταση

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

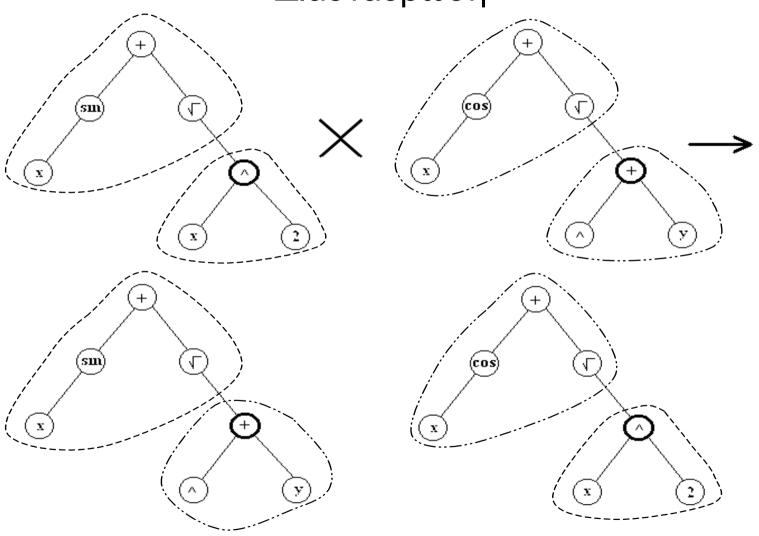
$$\sin(x) + \sqrt{x^2 + y}$$





Γενετικός Προγραμματισμός

Διασταύρωση



Τεχνητή Νοημοσύνη, Β' Έκδοση

23