



Machine Learning & Knowledge Extraction

DR KONSTANTINOS KARAMPIDIS

Πληροφορίες Μαθήματος

- ▶ Ωράριο:
 - ▶ Θεωρία: **Κάθε Τρίτη 09:00-13:00 – Αίθουσα 207**
 - ▶ Εργαστήριο: **13:00-14:00 ΕΡΓ6 (σύμφωνα με το πρόγραμμα που είναι αναρτημένο στο eclass)**
- ▶ Εργασίες
 - ▶ **1 project – Ομάδες έως 2 ατόμων – 80%**
 - ▶ **Εργαστηριακές ασκήσεις – 20%**
- ▶ Προαπαιτούμενα: Κανένα

Περιεχόμενο Μαθήματος

- ▶ Εισαγωγή στη Μηχανική Μάθηση - τι είναι, γιατί μας ενδιαφέρει, παραδείγματα προβλημάτων, η μηχανική μάθηση ως αναζήτηση, υπόθεση επαγωγικής μάθησης
- ▶ Επεξεργασία εισόδου – Μείωση διαστατικότητας- Αξιολόγηση
- ▶ Μέθοδοι επιβλεπόμενης μάθησης
- ▶ **Νευρωνικά Δίκτυα**
- ▶ **Εξελικτική Μάθηση – Γενετικοί Αλγόριθμοι**
- ▶ Μηχανική Μάθηση Βασιζόμενη σε Κανόνες
- ▶ Ενισχυτική Μάθηση
- ▶ Μάθηση Αναπαράστασης
- ▶ Εξόρυξη Δεδομένων

Ανάστροφη Μετάδοση Λάθους

- ▶ Γνωστή μέθοδος εκπαίδευσης ΤΝΔ πολλών επιπέδων
- ▶ Βασική ιδέα: εύρεση ποσοστού συνολικού σφάλματος που αναλογεί στα βάρη του κάθε νευρώνα
 - ▶ Διορθώσεις βαρών ανά νευρώνα – πολύπλοκη αν συμπεριλαμβάνονται κρυφά επίπεδα διότι η έξοδος τους επηρεάζει πολλούς νευρώνες
 - ▶ Βασίζεται στο γενικευμένο κανόνα Δέλτα

Ανάστροφη Μετάδοση Λάθους

Η εκπαίδευση νευρωνικών δικτύων με οπισθοδιάδοση (Backpropagation) περιλαμβάνει τα ακόλουθα βήματα:

- Ένα εμπρόσθιο πέρασμα, κάνοντας προβλέψεις σε δεδομένα εκπαίδευσης.
- Μια συνάρτηση απωλειών μετρά το σφάλμα των προβλέψεων του μοντέλου κατά τη διάρκεια αυτού του περάσματος προς τα εμπρός.
- Οπισθοδιάδοση του σφάλματος, ή ένα αντίστροφο πέρασμα, για τον υπολογισμό των μερικών παραγώγων της συνάρτησης απώλειας.
- Κάθοδος κλίσης (gradient descent), για την ενημέρωση των βαρών του μοντέλου.

Ανάστροφη Μετάδοση Λάθους

Forward pass

Τα νευρωνικά δίκτυα εξάγουν προβλέψεις μέσω εμπρόσθιας διάδοσης. Η προς τα εμπρός διάδοση είναι ουσιαστικά μια μακρά σειρά εμφωλευμένων εξισώσεων, με τις εξόδους των συναρτήσεων ενεργοποίησης από ένα επίπεδο νευρώνων να χρησιμεύουν ως είσοδοι στις συναρτήσεις ενεργοποίησης των νευρώνων του επόμενου επιπέδου.

Η εκπαίδευση του μοντέλου ξεκινά συνήθως με μια τυχαία αρχικοποίηση των βαρών. Οι υπερπαραμέτροι του μοντέλου, όπως ο αριθμός των κρυφών επιπέδων, ο αριθμός των νευρώνων σε κάθε επίπεδο και οι συναρτήσεις ενεργοποίησης για συγκεκριμένους νευρώνες, διαμορφώνονται χειροκίνητα και δεν υπόκεινται σε εκπαίδευση.

Ανάστροφη Μετάδοση Λάθους

Forward pass

Σε κάθε εμπρόσθιο πέρασμα, λαμβάνεται δείγμα εισόδου από το σύνολο δεδομένων εκπαίδευσης. Οι νευρώνες του επιπέδου εισόδου λαμβάνουν το διάνυσμα εισόδου και ο καθένας μεταβιβάζει την τιμή τους - πολλαπλασιασμένη με κάποιο τυχαίο αρχικό βάρος - στους κόμβους του πρώτου κρυμμένου επιπέδου. Οι νευρώνες σε αυτό το επίπεδο λαμβάνουν το σταθμισμένο άθροισμα αυτών των τιμών εξόδου ως είσοδο σε μια συνάρτηση ενεργοποίησης, της οποίας η τιμή εξόδου (υπό την προϋπόθεση ενός τυχαίου αρχικού βάρους) χρησιμεύει ως είσοδος στους νευρώνες του επόμενου επιπέδου. Αυτό συνεχίζεται μέχρι το επίπεδο εξόδου, όπου γίνεται η τελική πρόβλεψη.

Ανάστροφη Μετάδοση Λάθους

Loss function

Μετά από κάθε εμπρόσθιο πέρασμα, μια «συνάρτηση απώλειας» (loss function) μετρά τη διαφορά (ή «απώλεια») μεταξύ της προβλεπόμενης εξόδου του μοντέλου για μια δεδομένη είσοδο και των σωστών προβλέψεων (ground truth) για την εν λόγω είσοδο. Με άλλα λόγια, μετράει πόσο διαφέρει η πραγματική έξοδος του μοντέλου από την επιθυμητή έξοδο.

Στόχος αυτής της συνάρτησης απωλειών είναι να ποσοτικοποιήσει την ανακρίβεια με τρόπο που να αντικατοπτρίζει κατάλληλα τόσο τη φύση όσο και το μέγεθος του σφάλματος της εξόδου του μοντέλου για κάθε είσοδο.

Ανάστροφη Μετάδοση Λάθους

Loss function

Διαφορετικοί μαθηματικοί τύποι απώλειας είναι καταλληλότεροι για συγκεκριμένες εργασίες: για παράδειγμα, παραλλαγές του μέσου τετραγωνικού σφάλματος λειτουργούν καλά για προβλήματα παλινδρόμησης, ενώ παραλλαγές της απώλειας διασταυρούμενης εντροπίας λειτουργούν καλά για ταξινόμηση.

Επειδή η συνάρτηση απωλειών λαμβάνει την έξοδο ενός νευρωνικού δικτύου ως είσοδο και η έξοδος του νευρωνικού δικτύου είναι μια σύνθετη συνάρτηση που περιλαμβάνει πολλές ένθετες συναρτήσεις ενεργοποίησης μεμονωμένων νευρώνων, η διαφοροποίηση της συνάρτησης απωλειών συνεπάγεται διαφοροποίηση ολόκληρου του δικτύου. Για να γίνει αυτό, η οπισθοδιάδοση χρησιμοποιεί τον κανόνα της αλυσίδας (chain rule).

Ανάστροφη Μετάδοση Λάθους

Backward pass

Ξεκινώντας από το τελευταίο επίπεδο, ένα «αντίστροφο πέρασμα» διαφοροποιεί τη συνάρτηση απώλειας για να υπολογίσει πώς κάθε μεμονωμένη παράμετρος του δικτύου συμβάλλει στο συνολικό σφάλμα για μια μεμονωμένη είσοδο.

Ανάστροφη Μετάδοση Λάθους

Έστω L το τελικό επίπεδο ενός νευρωνικού δικτύου. Η τιμή softmax κάθε νευρώνα εξόδου αντιπροσωπεύει την πιθανότητα (0-1), ότι μια είσοδος ανήκει στην κατηγορία τους. Σε ένα τέλεια εκπαιδευμένο μοντέλο, ο νευρώνας που αντιπροσωπεύει τη σωστή ταξινόμηση θα έχει τιμή εξόδου κοντά στο 1 και οι άλλοι νευρώνες θα έχουν τιμή εξόδου κοντά στο 0.

Έστω L_c ο/οι νευρώνας/ες που αντιπροσωπεύει/ουν τη σωστή πρόβλεψη. Η συνάρτηση ενεργοποίησης του L_c είναι μια σύνθετη συνάρτηση, που περιέχει τις πολλές ένθετες συναρτήσεις ενεργοποίησης ολόκληρου του νευρωνικού δικτύου από το επίπεδο εισόδου έως το επίπεδο εξόδου. Η ελαχιστοποίηση της συνάρτησης απωλειών θα συνεπαγόταν την πραγματοποίηση προσαρμογών σε όλο το δίκτυο που φέρνουν την έξοδο της συνάρτησης ενεργοποίησης του L_c πιο κοντά στο 1.

Ανάστροφη Μετάδοση Λάθους

Για να το κάνουμε αυτό, θα πρέπει να γνωρίζουμε πώς οποιαδήποτε αλλαγή στα προηγούμενα στρώματα θα αλλάξει την ίδια την έξοδο του Lc. Με άλλα λόγια, θα πρέπει να βρούμε τις μερικές παραγώγους της συνάρτησης ενεργοποίησης του Lc.

Η έξοδος της συνάρτησης ενεργοποίησης του Lc εξαρτάται από τις συνεισφορές που λαμβάνει από τους νευρώνες του προτελευταίου στρώματος, το οποίο θα ονομάσουμε στρώμα L-1. Ένας τρόπος για να αλλάξουμε την έξοδο του Lc είναι να αλλάξουμε τα βάρη μεταξύ των νευρώνων στο L-1 και στο Lc. Υπολογίζοντας τη μερική παράγωγο κάθε βάρους του L-1 σε σχέση με τα άλλα βάρη, μπορούμε να δούμε πώς η αύξηση ή η μείωση οποιουδήποτε από αυτά θα φέρει την έξοδο του Lc πιο κοντά (ή πιο μακριά από το 1).

Ανάστροφη Μετάδοση Λάθους

Αλλά αυτός δεν είναι ο μόνος τρόπος για να αλλάξετε την έξοδο του Lc. Οι συνεισφορές που λαμβάνει ο Lc από τους νευρώνες L-1 καθορίζονται όχι μόνο από τα βάρη που εφαρμόζονται στις τιμές εξόδου του L-1, αλλά και από τις ίδιες τις πραγματικές τιμές εξόδου (πριν από τα βάρη). Οι τιμές εξόδου των νευρώνων L-1, με τη σειρά τους, επηρεάζονται από τα βάρη που εφαρμόζονται στις εισόδους που λαμβάνουν από το L-2. Έτσι, μπορούμε να διαφοροποιήσουμε τις συναρτήσεις ενεργοποίησης στον L-1 για να βρούμε τις μερικές παραγώγους των βαρών που εφαρμόζονται στις συνεισφορές του L-2.

Αυτές οι μερικές παράγωγοι μας δείχνουν πώς οποιαδήποτε αλλαγή σε ένα βάρος του L-2 θα επηρεάσει τις εξόδους στο L-1, οι οποίες στη συνέχεια θα επηρεάσουν την τιμή εξόδου του Lc και συνεπώς θα επηρεάσουν τη συνάρτηση απώλειας.

Ανάστροφη Μετάδοση Λάθους

Με την ίδια λογική, θα μπορούσαμε επίσης να επηρεάσουμε τις τιμές εξόδου που λαμβάνουν οι νευρώνες L-1 από τους νευρώνες L-2 προσαρμόζοντας τις συνεισφορές που λαμβάνουν οι νευρώνες L-2 από τους νευρώνες του L-3. Έτσι, βρίσκουμε τις μερικές παραγώγους στο L-3, και ούτω καθεξής, επαναλαμβάνοντας αναδρομικά αυτή τη διαδικασία μέχρι να φτάσουμε στο επίπεδο εισόδου. Όταν τελειώσουμε, έχουμε την κλίση της συνάρτησης απώλειας: ένα διάνυσμα της μερικής παραγώγου της για κάθε παράμετρο βάρους και bias στο δίκτυο.

Τώρα έχουμε ολοκληρώσει ένα εμπρόσθιο πέρασμα και ένα οπίσθιο πέρασμα για ένα μόνο παράδειγμα εκπαίδευσης. Ωστόσο, στόχος μας είναι να εκπαιδεύσουμε το μοντέλο να γενικεύει καλά σε νέες εισόδους. Για να γίνει αυτό απαιτείται εκπαίδευση σε μεγάλο αριθμό δειγμάτων που αντικατοπτρίζουν την ποικιλομορφία και το εύρος των εισόδων στις οποίες το μοντέλο θα αναλάβει να κάνει προβλέψεις μετά την εκπαίδευση.

Ανάστροφη Μετάδοση Λάθους

Gradient descent

Τώρα που έχουμε τις κλίσεις της συνάρτησης απώλειας σε σχέση με κάθε παράμετρο βάρους και bias στο δίκτυο, μπορούμε να ελαχιστοποιήσουμε τη συνάρτηση απώλειας - και συνεπώς να βελτιστοποιήσουμε το μοντέλο - χρησιμοποιώντας την κάθοδο κλίσης (gradient descent) για την ενημέρωση των παραμέτρων του μοντέλου.

Η μετακίνηση προς τα κάτω-κατεβαίνοντας-της κλίσης της συνάρτησης απώλειας θα μειώσει την απώλεια. Δεδομένου ότι η κλίση που υπολογίσαμε κατά την οπισθοδιάδοση περιέχει τις μερικές παραγώγους για κάθε παράμετρο του μοντέλου, γνωρίζουμε προς ποια κατεύθυνση να ενημερώσουμε κάθε παράμετρο μας για να μειώσουμε το σφάλμα.

Στόχος μας είναι να ενημερώσουμε επαναληπτικά τα βάρη μέχρι να φτάσουμε στην ελάχιστη κλίση.

Ανάστροφη Μετάδοση Λάθους

Learning rate

Το πόσο θα αλλάξουν οι τιμές είναι μια ρυθμίσιμη υπερπαράμετρος, που ονομάζεται ρυθμός μάθησης. Η επιλογή του σωστού ρυθμού μάθησης είναι σημαντική για αποτελεσματική και αποτελεσματική εκπαίδευση.

Θυμηθείτε ότι οι συναρτήσεις ενεργοποίησης σε ένα νευρωνικό δίκτυο είναι μη γραμμικές. Ορισμένες διαβαθμίσεις μπορεί να έχουν περίπου σχήμα U: το βήμα προς μια κατεύθυνση μετακινείται κάτω στην κλίση, αλλά αν συνεχίσετε να βαδίζετε προς αυτήν την κατεύθυνση, τελικά θα ανεβείτε στην κλίση.

Ένα χαμηλός ρυθμός μάθησης διασφαλίζει ότι βαδίζουμε πάντα προς τη σωστή κατεύθυνση, αλλά ο υπολογισμός τόσων πολλών αλλαγών είναι χρονοβόρος και υπολογιστικά ακριβός. Ένας υψηλός ρυθμός μάθησης είναι υπολογιστικά αποδοτικός, αλλά κινδυνεύει να υπερβεί το ελάχιστο.

Ανάστροφη Μετάδοση Λάθους

Batch size

Ένας άλλος παράγοντας που λαμβάνεται υπόψη στην κλίση κατάβασης είναι η συχνότητα ενημέρωσης των βαρών. Μια επιλογή είναι να υπολογίσετε τις κλίσεις για κάθε παράδειγμα στο σύνολο δεδομένων εκπαίδευσης και, στη συνέχεια, να λάβετε έναν μέσο όρο αυτών των κλίσεων και να τον χρησιμοποιήσετε για να ενημερώσετε τις παραμέτρους. Η διαδικασία επαναλαμβάνεται επαναλαμβανόμενα σε ένα αριθμό εποχών μέχρι να σταθεροποιηθεί το ποσοστό σφάλματος. Αυτή η μέθοδος ονομάζεται batch gradient descent.

Ανάστροφη Μετάδοση Λάθους

Batch size

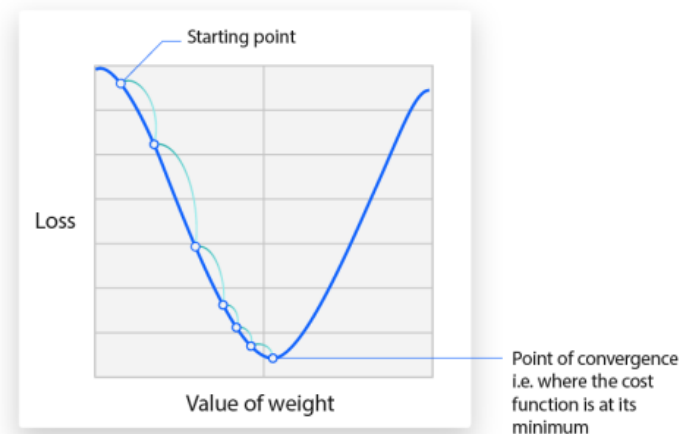
Όταν το σύνολο δεδομένων εκπαίδευσης είναι πολύ μεγάλο -όπως συμβαίνει συνήθως στη βαθιά μάθηση- η μέθοδος αυτή συνεπάγεται απαγορευτικά μεγάλους χρόνους επεξεργασίας.

Ο υπολογισμός των κλίσεων για εκατομμύρια παραδείγματα για κάθε επανάληψη ενημερώσεων βάρους γίνεται αναποτελεσματικός. Στην στοχαστική κλίση κάθοδου (Stochastic Gradient Descent), κάθε εποχή χρησιμοποιεί ένα μόνο παράδειγμα εκπαίδευσης για κάθε βήμα. Ενώ η απώλεια μπορεί να κυμαίνεται από εποχή σε εποχή, γρήγορα συγκλίνει στο ελάχιστο σε πολλές ενημερώσεις.

Ανάστροφη Μετάδοση Λάθους

Mini-batch gradient descent

Αντιπροσωπεύει μια μέσης οδό. Τα παραδείγματα εκπαίδευσης δειγματοληπτούνται τυχαία σε batches σταθερού μεγέθους και στη συνέχεια υπολογίζονται οι κλίσεις τους και υπολογίζεται ο μέσος όρος. Αυτό μετριάζει τις απαιτήσεις αποθήκευσης μνήμης κατά την εφαρμογή του batch gradient descent, ενώ παράλληλα μειώνει τη σχετική αστάθεια του SGD.



ΤΝΔ Kohonen

Ο αυτοοργανωμένος χάρτης (SOM) εφευρέθηκε από τον Dr. Teuvo Kohonen το 1982 και ήταν ευρέως γνωστός ως χάρτης Kohonen.

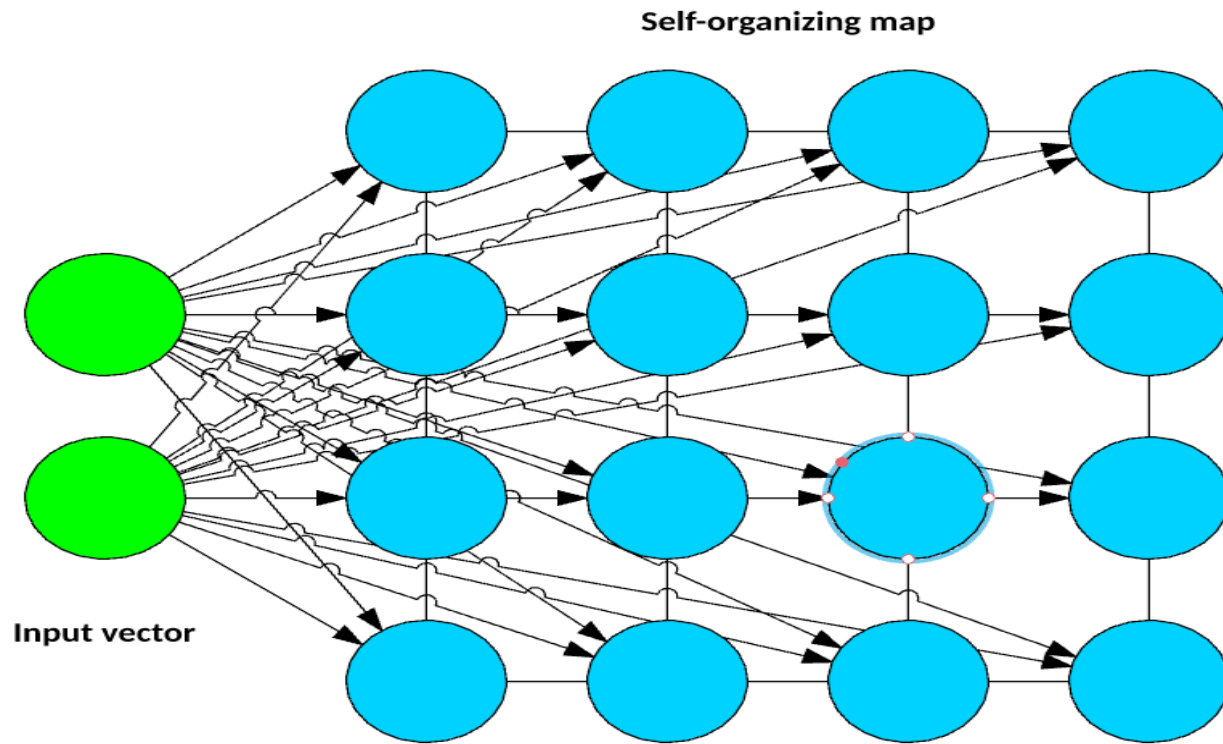
Ο SOM είναι ένα νευρωνικό δίκτυο χωρίς επίβλεψη που δημιουργεί συστάδες του συνόλου δεδομένων εισόδου μειώνοντας τη διαστατικότητα της εισόδου. Τα SOM διαφέρουν από τα παραδοσιακά τεχνητά νευρωνικά δίκτυα.

TNΔ Cohonen

- ▶ Είδος TNΔ ανταγωνισμού:

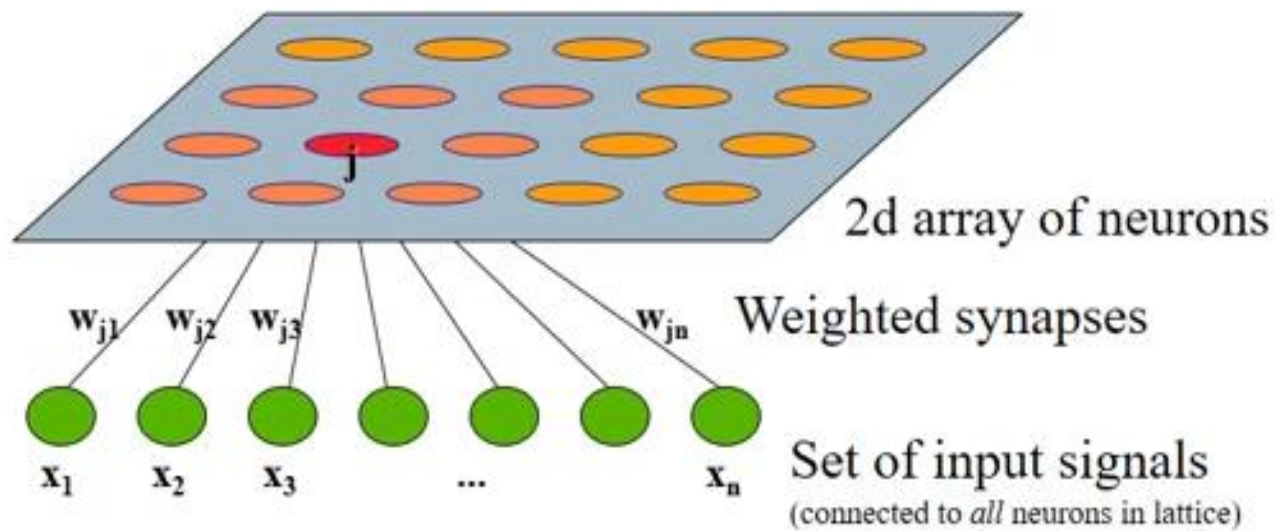
- ▶ Θετικός ή ακόμη και αρνητικός επηρεασμός άλλων νευρώνων από ένα ή περισσότερους νευρώνες
- ▶ Ποιος νευρώνας ανταποκρίνεται περισσότερο;
- ▶ Πχ. Ο νευρώνας με την μεγαλύτερη έξοδο βγαίνει νικητής και παράγει τα τελικά αποτελέσματα

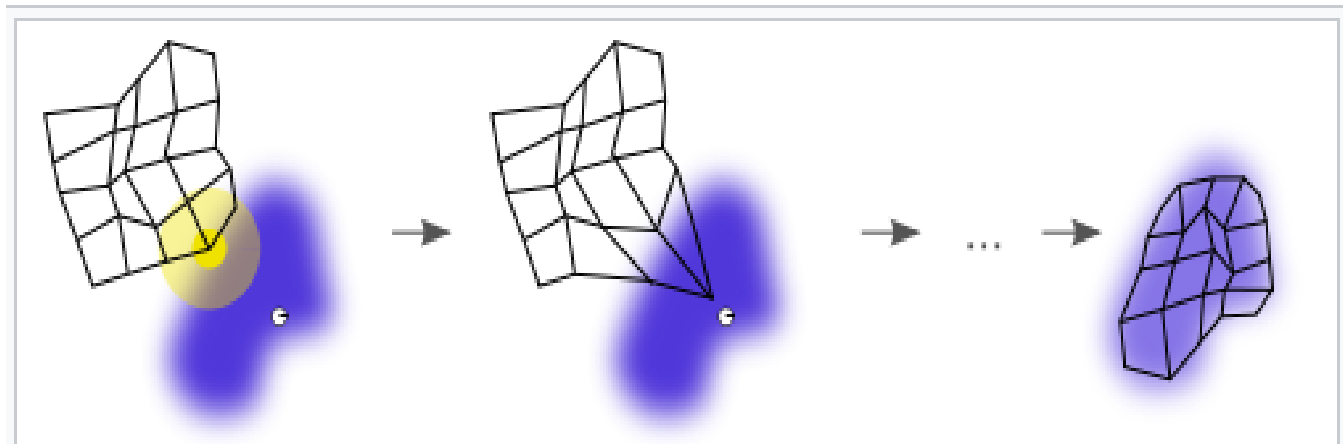
T Δ Kohonen



T Δ Kohonen

The neurons in the output layer are arranged on a map





An illustration of the training of a self-organizing map. The blue blob is the distribution of the training data, and the small white disc is the current training datum drawn from that distribution. At first (left) the SOM nodes are arbitrarily positioned in the data space. The node (highlighted in yellow) which is nearest to the training datum is selected. It is moved towards the training datum, as (to a lesser extent) are its neighbors on the grid. After many iterations the grid tends to approximate the data distribution (right).

Source: https://en.wikipedia.org/wiki/Self-organizing_map

TNΔ Cohonen

- ▶ Χρησιμοποιεί μη επαγωγική μάθηση
- ▶ Αντιστοιχεί σε γεωμετρική τοπολογία (πχ. επίπεδο, σφαίρα, κλπ.)
- ▶ Κάθε νευρώνας είναι συνδεδεμένος με την είσοδο k στοιχείων και λαμβάνει ένα πλήρες αντίγραφο του διανύσματος εισόδου τροποποιημένου με βάρη
- ▶ Υπάρχει σύνδεση νευρώνων στο ίδιο επίπεδο έτσι ώστε κοντινοί νευρώνες να επηρεάζονται θετικά και μακρινοί ουδέτερα ή αρνητικά

TNΔ Cohonen

- ▶ Αναπροσαρμογή βαρών:
 - ▶ Για μια συγκεκριμένη είσοδο επιλέγεται ο νευρώνας που είναι πιο κοντά σε αυτή και μεταβάλλει τα βάρη του ώστε να τη πλησιάσει περισσότερο
 - ▶ Λόγω της συνδεσμολογίας, και οι κοντινοί του νευρώνες αλλάζουν τα βάρη τους
- ▶ Κατά τη διαδικασία μάθησης μειώνεται τόσο η αναπροσαρμογή των βαρών όσο και το μέγεθος γειτνίασης μέχρι ένα κατώφλι κάτω από το οποίο θεωρούμε ότι το TNΔ είναι εκπαιδευμένο
- ▶ Μέσω των διαδικασιών ανταγωνισμού και σχέσεων γειτνίασης, δημιουργείται ένας χάρτης πάνω στη γεωμετρική τοπολογία που αντιστοιχεί στην κατηγοριοποίηση των εισόδων του TNΔ
 - ▶ σύνολα από νευρώνες αντικατοπτρίζουν μια συγκεκριμένη κλάση / κατηγορία

Επιλογή Νικητή Νευρώνα

► 3 εναλλακτικοί τρόποι:

1. Συνάρτηση μεγίστου χρησιμοποιείται για την εύρεση του νευρώνα με τη μεγαλύτερη απόκριση στην είσοδο. Οπότε αποδίδεται έξοδος +1 για αυτόν και 0 για τους υπόλοιπους
2. Επιλέγεται ο νευρώνας με το μεγαλύτερο εσωτερικό γινόμενο μεταξύ του διανύσματος εισόδου και αυτού των βαρών
3. Επιλέγεται ο νευρώνας με τη μικρότερη Ευκλείδεια απόσταση μεταξύ των διανυσμάτων εισόδου και βαρών

Διαδικασία Εκπαίδευσης ΤΝΔ Cohonen

- ▶ Γίνεται αρχικά μια ανάθεση μικρών τιμών στα βάρη εισόδου
 - ▶ Τόσο τα αρχικά βάρη όσο και τα διανύσματα εισόδου μπορούν να κανονικοποιηθούν ως προς το 1:

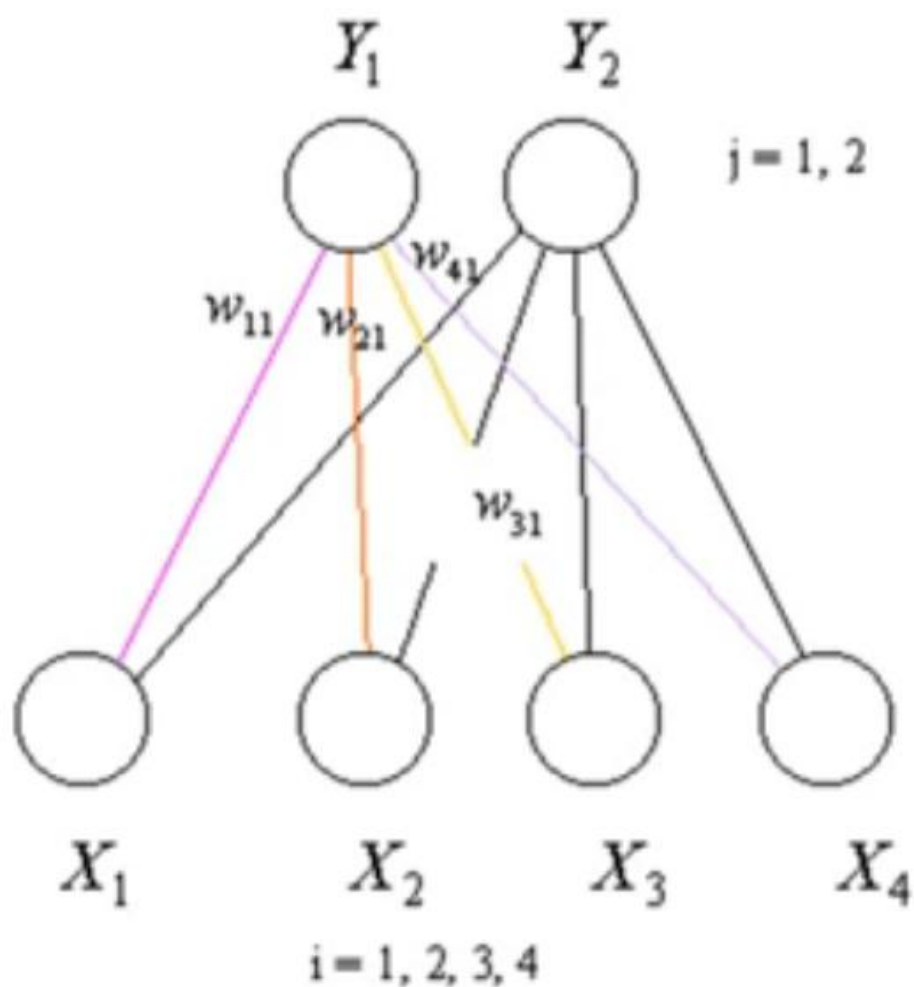
$$s'_i = \frac{s_i}{\left[s_1^2 + s_2^2 + \mathbf{K} + s_k^2 \right]^{1/2}}, w'_{ij} = \frac{w_{ij}}{\left[w_{i1}^2 + w_{i2}^2 + \mathbf{K} + w_{ik}^2 \right]^{1/2}}$$

- ▶ Για κάθε νευρώνα υπολογίζεται η απόσταση των βαρών του από το διάνυσμα εκπαίδευσης:

$$d_i = \sqrt{\sum_{k=1}^n (s_k - w_{ik})^2}$$

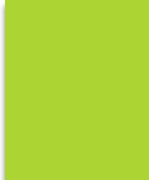
- ▶ Νικητής είναι ο νευρώνας με την μικρότερη απόσταση
 - ▶ Τόσο ο νευρώνας όσο και η γειτονιά του προσαρμόζουν τα βάρη τους
- ▶ Στον επόμενο κύκλο ο ρυθμός εκπαίδευσης και το μέγεθος γειτνίασης μικραίνουν

Step	Action
0	Initialize weights. Set max value for R , set learning rate α .
1	While stopping condition false do steps 2 to 8
2	For each input vector \underline{x} do steps 3 to 5
3	For each j neuron, compute the Euclidean distance $D(j) = \sqrt{\sum_{i=1}^n (x_i - w_{ij})^2}$
4	Find the index J such that $D(J)$ is a minimum
5	For all neurons j within a specified neighbourhood of J and for all i $w_{ij}(\text{new}) = w_{ij}(\text{old}) + \alpha(x_i - w_{ij}(\text{old}))$
6	Update learning rate α . It is a decreasing function of the number of epochs.
7	Reduce radius of topological neighbourhood at specified times
8	Test stopping condition. Typically this is a small value of the learning rate with which the weight updates are insignificant.



Let the initial weight matrix be

$$\begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \\ w_{31} & w_{32} \\ w_{41} & w_{42} \end{bmatrix} = \begin{bmatrix} 0.2 & 0.8 \\ 0.6 & 0.4 \\ 0.5 & 0.7 \\ 0.9 & 0.3 \end{bmatrix}$$



Consider a simple example in which there are only **4 input training patterns**.

x_1	x_2	x_3	x_4
1	1	0	0
0	0	0	1
1	0	0	0
0	0	1	1

Let the learning rate at time $t + 1$ be given by
and suppose $\alpha(t = 0) = 0.6$

$$\alpha(t + 1) = \frac{\alpha(t)}{2},$$

Let topological radius $R = 0$.

Following the algorithm presented in the previous algorithm:

For vector 1100 (We are using the Euclidean distance squared for convenience)

$$D(1) = (1-0.2)^2 + (1-0.6)^2 + (0-0.5)^2 + (0-0.9)^2 = 1.86$$

$$D(1) = 1.86, D(2) = 0.98$$

Hence $J = 2$. Note that $R = 0$, so we need not update the weights of any neighboring neurons.

Using $w_{ij}(\text{new}) = w_{ij}(\text{old}) + \alpha(x_i - w_{ij}(\text{old}))$, the new weight matrix is

$$\begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \\ w_{31} & w_{32} \\ w_{41} & w_{42} \end{bmatrix} = \begin{bmatrix} 0.2 & 0.92 \\ 0.6 & 0.76 \\ 0.5 & 0.28 \\ 0.9 & 0.12 \end{bmatrix} \rightarrow \begin{matrix} w_{12}(\text{new}) = w_{12}(\text{old}) + \alpha(x_1 - w_{12}(\text{old})) \\ = 0.8 + 0.6(1 - 0.8) = 0.92 \end{matrix}$$

Likewise for remains training patterns

For vector 0001

$$D(1) = 0.66, D(2) = 2.2768$$

Hence $J = 1$.

$$\begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \\ w_{31} & w_{32} \\ w_{41} & w_{42} \end{bmatrix} = \begin{bmatrix} 0.08 & 0.92 \\ 0.24 & 0.76 \\ 0.20 & 0.28 \\ 0.96 & 0.12 \end{bmatrix}$$

For vector 0011

$$D(1) = 0.7056, D(2) = 2.724$$

Hence $J = 1$

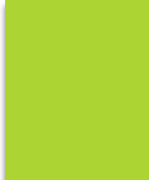
$$\begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \\ w_{31} & w_{32} \\ w_{41} & w_{42} \end{bmatrix} = \begin{bmatrix} 0.032 & 0.968 \\ 0.096 & 0.304 \\ 0.680 & 0.112 \\ 0.984 & 0.048 \end{bmatrix}$$

For vector 1000

$$D(1) = 1.8656, D(2) = 0.6768$$

Hence $J = 2$

$$\begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \\ w_{31} & w_{32} \\ w_{41} & w_{42} \end{bmatrix} = \begin{bmatrix} 0.08 & 0.968 \\ 0.24 & 0.304 \\ 0.20 & 0.112 \\ 0.96 & 0.048 \end{bmatrix}$$



$$\alpha(1) = \frac{\alpha(0)}{2} = \frac{0.6}{2} = 0.3$$

Now reduce learning rate (step 6):

It can be shown that after **100 presentations** of all the input vector, the final weight matrix is

$$\begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \\ w_{31} & w_{32} \\ w_{41} & w_{42} \end{bmatrix} = \begin{bmatrix} 6.7 \times 10^{-17} & 1 \\ 2 \times 10^{-16} & 0.49 \\ 0.51 & 2.3 \times 10^{-16} \\ 1 & 1 \times 10^{-16} \end{bmatrix}$$

This matrix seems to converge to

$$\begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \\ w_{31} & w_{32} \\ w_{41} & w_{42} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0.5 \\ 0.5 & 0 \\ 1 & 0 \end{bmatrix}$$

Cluster 1



Cluster 2



TEST NETWORK

Suppose the input pattern is 1 1 0 0. This matrix seems to c
Then

$$D(j) = (w_{1j} - x_1)^2 + (w_{2j} - x_2)^2 + (w_{3j} - x_3)^2 + (w_{4j} - x_4)^2$$

$$D(1) = (0 - 1)^2 + (0 - 1)^2 + (0.5 - 0)^2 + (1 - 0)^2 = 3.25$$

$$D(2) = (1 - 1)^2 + (0.5 - 1)^2 + (0 - 0)^2 + (0 - 0)^2 = 0.25$$

$$\begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \\ w_{31} & w_{32} \\ w_{41} & w_{42} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0.5 \\ 0.5 & 0 \\ 1 & 0 \end{bmatrix}$$

Cluster 1

Thus neuron 2 is the "**winner**",
and is the localized active region
of the SOM. Notice that we may
label this input pattern to belong
to cluster 2.

For all the other patterns, we find
the clusters are as listed below.

x_1	x_2	x_3	x_4	Cluster
1	1	0	0	2
0	0	0	1	1
1	0	0	0	2
0	0	1	1	1

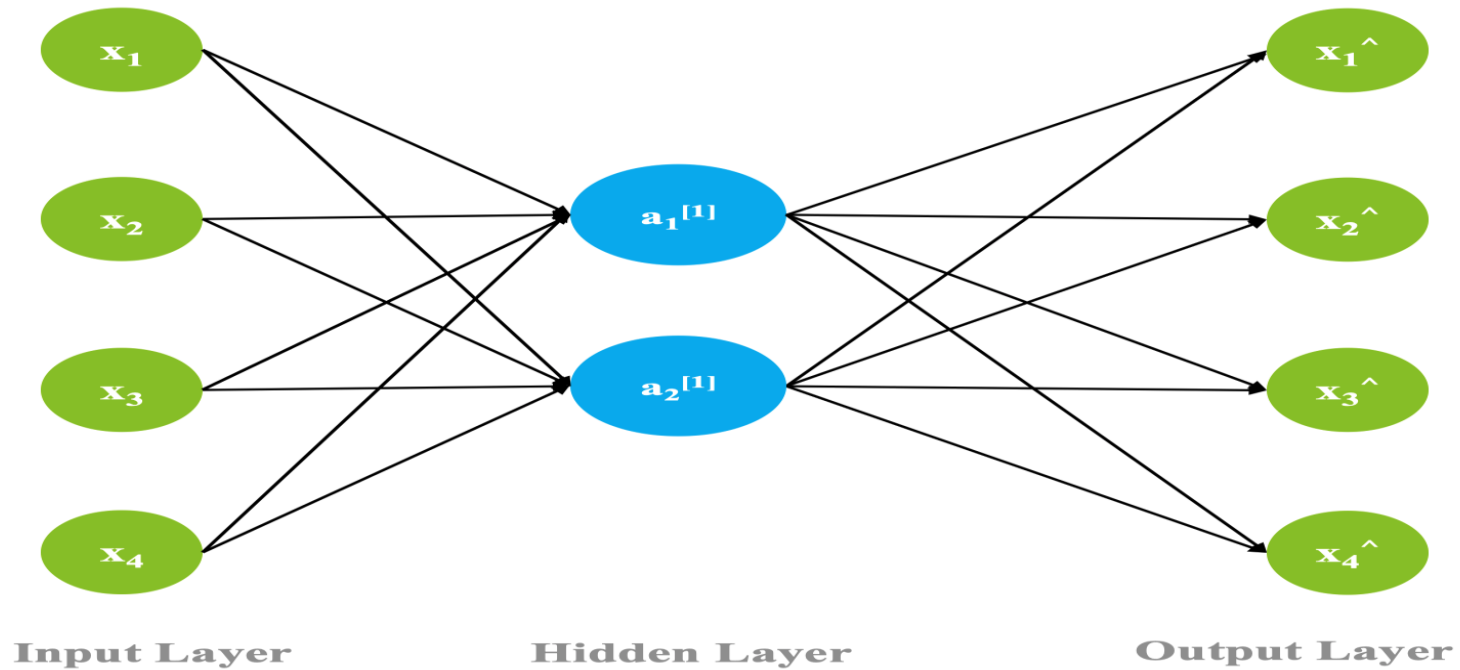
Autoencoders

Η πρώτη γνωστή χρήση των autoencoders έγινε από τον LeCun το 1987. Αυτή η παραλλαγή ενός ANN απαρτίζεται από 3 επίπεδα: επίπεδο εισόδου, κρυφό και επίπεδο εξόδου.

Αρχικά, το επίπεδο εισόδου κωδικοποιείται στο κρυφό επίπεδο χρησιμοποιώντας μια κατάλληλη συνάρτηση κωδικοποίησης (encoder). Ο αριθμός των κόμβων στο κρυφό επίπεδο είναι πολύ μικρότερος από τον αριθμό των κόμβων στο επίπεδο εισόδου. Αυτό το κρυφό επίπεδο περιέχει τη συμπιεσμένη αναπαράσταση της αρχικής εισόδου. Το επίπεδο εξόδου στοχεύει στην ανακατασκευή του επιπέδου εισόδου χρησιμοποιώντας μια λειτουργία αποκωδικοποιητή (decoder).

Autoencoders

Autoencoders



Source: <https://developer.ibm.com/articles/cc-machine-learning-deep-learning-architectures/>

Γενετικοί Αλγόριθμοι – Εισαγωγή

Οι γενετικοί αλγόριθμοι είναι μια κατηγορία πιθανοκρατικών αλγορίθμων που βασίζονται στη Δαρβινική θεωρία της εξέλιξης

Γενετικοί Αλγόριθμοι – Εισαγωγή

- ▶ Περίληψη μηχανισμού εξέλιξης:
 - ▶ Μόνο οι οργανισμοί που επιβιώνουν στο περιβάλλον τους πολλαπλασιάζονται μέσω της διαδικασίας αναπαραγωγής
 - ▶ Οι απόγονοι παρουσιάζουν κάποια (μικρή) διαφοροποίηση σε σχέση με τους προγόνους τους
 - ▶ Υπερισχύουν οι απόγονοι που περιέχουν τα καλύτερα χαρακτηριστικά
 - ▶ Όταν το περιβάλλον μεταλλάσσεται αργά, τότε οι οργανισμοί προσαρμόζονται σταδιακά σε αυτό. Διαφορετικά, αρκετά είδη οργανισμών μπορούν να εξαφανιστούν.
 - ▶ Μπορεί να συντελεστεί τυχαία και η διαδικασία της μετάλλαξης η οποία μπορεί να οδηγήσει είτε σε θάνατο είτε στην δημιουργία καλύτερων οργανισμών

Βασική Φιλοσοφία – Ιδέα

- ▶ Δημιούργησε ένα πληθυσμό Π από N υποψήφιας λύσεις
- ▶ Βαθμολόγησε τις λύσεις με βάση μια συνάρτηση καταλληλότητας (fitness function)
- ▶ Σχημάτισε από τον αρχικό πληθυσμό N , 2 ζευγάρια (όχι απαραίτητα από μοναδικούς γονείς) δίνοντας μεγαλύτερη προτεραιότητα στις καλύτερες λύσεις

Βασική Φιλοσοφία – Ιδέα

- ▶ Από κάθε ζευγάρι προκύπτουν 2 λύσεις / απογόνους (offsprings)
- ▶ Έτσι, προκύπτει ένας νέος πληθυσμός Π' που είναι συνήθως καλύτερος από τον προηγούμενο. Η διαδικασία επαναλαμβάνεται για τον νέο πληθυσμό
- ▶ Συνηθισμένες συνθήκες τερματισμού:
 - ▶ Εύρεση μιας τέλειας λύσης (με βάση την συνάρτηση καταλληλότητας)
 - ▶ Σύγκλιση όλων των λύσεων σε 1 (σύγκλιση σημαίνει επικράτηση ενός χρωμοσώματος ή μικρών διαφοροποιήσεων του σε μεγάλο ποσοστό του πληθυσμού)

Περιγραμμά Λειτουργίας

- ▶ Εκτελεί αναζήτηση στο χώρο υποψήφιας λύσεων
- ▶ Στόχος η εύρεση μιας λύσης που μεγιστοποιεί την συνάρτηση καταλληλότητας
- ▶ Η αναζήτηση είναι παράλληλη διότι μπορεί να πραγματοποιηθεί ξεχωριστά για κάθε υποψήφια λύση
- ▶ Η αναζήτηση εστιάζει τόσο στις πιο καλές λύσεις αλλά και στις υπόλοιπες για να μην εγκλωβιστεί σε ένα τοπικό μέγιστο

Συστατικά Μέρη

- ▶ Δημιουργία αρχικού τυχαίου πληθυσμού
- ▶ Αναπαράσταση λύσεων
- ▶ Συνάρτηση καταλληλότητας
- ▶ Επιλογή ζευγαριών
- ▶ Διαδικασία αναπαραγωγής

Αναπαράσταση λύσεων

- ▶ Μέσω μιας συμβολοσειράς (χρωμόσωμα) πεπερασμένου αλφαβήτου. Κάθε μέρος της συμβολοσειράς αναφέρεται ως γονίδιο.
 - ▶ Παράδειγμα: Το DNA αντιστοιχεί σε αλφάβητο 4 στοιχείων A, G, T, C
- ▶ Συνήθως, χρησιμοποιείται δυαδικό αλφάβητο. Συνεπώς, οι συμβολοσειρές είναι bit-strings.
- ▶ Όμως, υπάρχουν περιπτώσεις πιο προχωρημένων αναπαραστάσεων λύσεων
- ▶ Οι λύσεις τις πιο πολλές φορές αντιστοιχούν σε μεταβλητές δεδομένων διαφορετικού τύπου. Οπότε οι μεταβλητές αυτές θα πρέπει να μετατραπούν σε δυαδικές
 - ▶ Παράδειγμα: μια boolean και μια short integer που θα χρειαστούν για την αναπαράστασή τους $8 + 1 = 9$ bits. Οπότε, αν η πρώτη είναι ίση με true και η δεύτερη με 73, τότε η αναπαράσταση της λύσης θα ήταν η: 101001001

Συνάρτηση Καταλληλότητας

- ▶ Λαμβάνει ως είσοδο ένα χρωμόσωμα και επιστρέφει μια τιμή που υποδηλώνει την καταλληλότητά του. Η τιμή αντιστοιχεί συνήθως στο εύρος πραγματικών τιμών $[0.0, 1.0]$ με θετική κατεύθυνση (1.0 -> τέλεια λύση)
- ▶ Χρησιμοποιείται τόσο στη/στις συνθήκες τερματισμού όσο και στη διαδικασία πιθανοκρατικής επιλογής των λύσεων (ως προς τον πληθυσμό της νέας γενιάς)
- ▶ Η κατασκευή της συνάρτησης εξαρτάται από το εκάστοτε πρόβλημα που πρέπει να επιλυθεί

Συνάρτηση Καταλληλότητας

- ▶ Ανάλογα με το πρόβλημα, η κατασκευή της συνάρτησης μπορεί να είναι από πολύ απλή ως εξαιρετικά πολύπλοκη
- ▶ Ιδανικά η συνάρτηση θα πρέπει να είναι συνεχής και μονότονη. Αν δεν είναι εφικτό αυτό (πράγμα σύνηθες), τότε δεν πρέπει να έχει πολλά τοπικά μέγιστα ή ένα απομονωμένο ολικό μέγιστο
- ▶ Πρέπει να αντικατοπτρίζει ρεαλιστικά την αξία του χρωμοσώματος – λύσης. Σε περίπτωση μη εφικτότητας έγκυρης λύσης, τότε θα πρέπει να αντικατοπτρίζει την απόσταση από την ιδανική έγκυρη λύση (πχ. αριθμός περιορισμών που έχουν παραβιαστεί).
- ▶ Το κόστος υπολογισμού καταλληλότητας λύσης είναι σημαντικό. Οπότε, μπορεί να επιλεχθεί μια προσεγγιστική συνάρτηση καταλληλότητας από μια ιδανική αν η πρώτη προσεγγίζει σε ικανοποιητικό βαθμό τον ιδανικό υπολογισμό της αξίας λύσης ενώ η δεύτερη αντιστοιχεί σε ένα μεγάλο υπολογιστικό κόστος

Διαδικασία Επιλογής Ζευγαριών – Γονέων

- ▶ Αντιστοιχεί στη πιθανοκρατική επιλογή μελών του πληθυσμού προς αναπαραγωγή
- ▶ Μπορεί κάποιοι γονείς να επιλεγούν περισσότερες από 1 φορές (πχ. καλύτερες λύσεις) και κάποιοι άλλοι καθόλου
- ▶ Αρχικά οι υποψήφιος λύσεις αντιγράφονται σε μια δεξαμενή ζευγαρώματος (mating pool)
 - ▶ Μέγεθος δεξαμενής: ίσο με αρχικό πληθυσμό
 - ▶ Μέλη: αρχικού πληθυσμού με πιθανότητα επιλογής ανάλογη με την καταλληλότητά τους
- ▶ Διάφορες τεχνικές επιλογής χρωματοσωμάτων μπορούν να χρησιμοποιηθούν

Τεχνική Ρουλέτας

► Αλγόριθμος:

1. Υπολόγισε το άθροισμα S όλων των τιμών καταλληλότητας των υποψήφιων λύσεων
2. Επέλεξε ένα τυχαίο αριθμό n μεταξύ 0 και S χρησιμοποιώντας ομοιόμορφη συνάρτηση κατανομής
3. Επαναληπτικά εξέτασε κάθε υποψήφια λύση και πρόσθεσε την τιμή σε ένα καταχωρητή K
4. Αν η τιμή του K γίνει ίση ή μεγαλύτερη του n , η λύση επιλέγεται και το K μηδενίζεται. Διαφορετικά πάμε στο βήμα 3
5. Αν δεν έχει επιλεγεί ο ζητούμενος αριθμός από λύσεις, τότε πάμε στο βήμα 2. Διαφορετικά, τερματίζουμε

Διαδικασία Επιλογής Γονέων

- ▶ Αφού δημιουργηθεί η δεξαμενή, παράγονται τα ζευγάρια τυχαία από τα οποία θα προκύψουν οι απόγονοι για το νέο πληθυσμό
- ▶ Ο τρόπος επιλογής γονέων επηρεάζει την απόδοση του γενετικού αλγορίθμου. Συνήθως, μπορεί να παρουσιαστεί το πρόβλημα της πρόωρης σύγκλισης (σε τοπικό μέγιστο) ενώ το ζητούμενο είναι η σύγκλιση προς το ολικό μέγιστο
- ▶ Ένα γονίδιο συγκλίνει όταν έχει ίδια τιμή στο 95% των χρωμοσωμάτων
- ▶ Ένας πληθυσμός έχει συγκλίνει όταν όλα τα γονίδιά του έχουν συγκλίνει

Πρόωρη Σύγκλιση

- ▶ Το φαινόμενο πρόωρης σύγκλισης μπορεί να παρουσιαστεί όταν η συνάρτηση καταλληλότητας έχει πολύ απότομες μεταβολές ή πολλά τοπικά μέγιστα
- ▶ Τρόποι αντιμετώπισης:
 - ▶ Απεικόνιση συνάρτησης καταλληλότητας σε μια νέα συνάρτηση (fitness remapping)
 - ▶ Καθορισμός ελάχιστων και μέγιστων ορίων ως προς το πόσες φορές επιλέγεται ένα χρωμόσωμα προς αναπαραγωγή σε κάθε κύκλο ανανέωσης πληθυσμού

Αργή Σύγκλιση

- ▶ Αντίθετο φαινόμενο
- ▶ Μη σύγκλιση πληθυσμού έπειτα από ένα μεγάλο αριθμό επαναλήψεων
- ▶ Παρουσιάζεται όταν η συνάρτηση καταλληλότητας έχει μικρές κλίσεις, δηλ. η διαφορά μεταξύ των μεγίστων και των ελαχίστων της είναι μικρή
- ▶ Λύση: απεικόνιση συνάρτησης καταλληλότητας

Εναλλακτική Προσέγγιση Επιλογής Γονέων

- ▶ Μέθοδος μερικής ανανέωσης
- ▶ Ανανέωση μόνο ποσοστού του πληθυσμού σε κάθε κύκλο ανανέωσης
- ▶ Χάσμα γενεών: ποσοστό χρωματοσωμάτων που ανανεώθηκε
- ▶ Τάση: ποσοστό < 1 ή πιο ακραία μόνο 2 μέλη επιλέγονται
- ▶ Η μέθοδος αυτή προσεγγίζει την πραγματικότητα διότι συνήθως συνυπάρχουν διαφορετικές γενεές που ανταγωνίζονται μεταξύ τους

Μέθοδος Μερικής Ανανέωσης

- ▶ 2 θέματα στην μέθοδο αυτή:
 - ▶ Επιλογή γονέων και επιλογή ισάριθμων γονέων που θα αποχωρήσουν.
- ▶ Λύσεις:
 - ▶ Επιλογή γονέων με πιθανότητα ανάλογη της καταλληλότητάς τους:

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

Τυχαία επιλογή γονέων προς αποχώρηση.

- ▶ Τυχαία επιλογή γονέων. Επιλογή γονέων προς αποχώρηση με πιθανότητα αντιστρόφως ανάλογη της καταλληλότητάς τους:

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

Διαδικασία Αναπαραγωγής

- ▶ Διαδικασία δημιουργίας απογόνων από 2 γονείς
- ▶ Υλοποίηση διαδικασίας μέσω διαφορετικών τεχνικών
 - ▶ Διασταύρωση (Crossover)
 - ▶ Ενός σημείου
 - ▶ Δύο σημείων
 - ▶ Ομοιόμορφη
 - ▶ Μετάλλαξη (Mutation)
- ▶ Συνήθως πραγματοποιείται τόσο διασταύρωση (1 από 3 είδη) όσο και μετάλλαξη για την παραγωγή των απογόνων

Διασταύρωση Ενός Σημείου

- ▶ Επιλογή τυχαίου σημείου διασταύρωσης για τους 2 γονείς που αντιστοιχεί στον αριθμό N μεταξύ 1 και L , όπου L είναι το μέγεθος των bit-strings
- ▶ Παραγωγή απογόνων από γονείς A και B :
 - ▶ Πρώτος απόγονος παράγεται από τα πρώτα N γονίδια του A και τα $L-N$ τελευταία του B
 - ▶ Δεύτερος απόγονος παράγεται από τα πρώτα N γονίδια του B και τα $L-N$ τελευταία του A

Παράδειγμα

- ▶ Έστω $A=011001$ και $B=110010$ ενώ $N=4$. Τότε θα έχουμε 2 απογόνους με την εξής μορφή:
 - ▶ 1^{ος}: **011010**
 - ▶ 2^{ος}: **110001**

Διασταύρωση 2 Σημείων

- ▶ Επιλογή τυχαία 2 σημείων μεταξύ 1 και L.
- ▶ Οι απόγονοι δημιουργούνται εναλλάξ με βάση τα μέρη των προγόνων τους που διαμερίζονται από τα 2 αυτά σημεία
- ▶ Παράδειγμα:
 - ▶ A = **00001010101**
 - ▶ B = **11101001000**
 - ▶ N1 = 3, N2 = 8

Διασταύρωση 2 Σημείων

τμήμα1 = A[:2] = θέσεις 1–2 = 00

τμήμα2 = B[2:7] = θέσεις 3–7 = 10100

τμήμα3 = A[7:] = θέσεις 8–11 = 101

Σύνθεση απογόνων (εναλλαγή τμημάτων: A–B–A για τον πρώτο, B–A–B για τον δεύτερο):

1ος απόγονος = τμήμα1 + τμήμα2 + τμήμα3
= "00" + "10100" + "101" = **00101000101**

2ος απόγονος = αντίστροφα (B–A–B):

τμήμα1' = B[:2] = 11

τμήμα2' = A[2:7] = 00101

τμήμα3' = B[7:] = 000

2ος απόγονος = "11" + "00101" + "000" = **11001011000**

Ομοιόμορφη Διασταύρωση

- ▶ Τυχαία επιλογή πολλαπλών σημείων (περίπου $L/2$)
- ▶ Παραγωγή εναλλάξ των απογόνων με βάση τα μέρη των προγόνων που διαμερίζονται από τα σημεία αυτά
- ▶ Παράδειγμα
 - ▶ $A = \mathbf{00001010101}$
 - ▶ $B = \mathbf{11101001000}$
 - ▶ $N1 = 1, N2 = 4, N3 = 6, N4 = 7, N5 = 10$

Ομοιόμορφη Διασταύρωση

▶ Παράδειγμα

▶ $A = \mathbf{00001010101}$

▶ $B = \mathbf{11101001000}$

▶ $N_1 = 1, N_2 = 4, N_3 = 6, N_4 = 7, N_5 = 10$

Αυτά δημιουργούν 6 τμήματα (όσα τα σημεία + 1). Τα τμήματα εναλλάσσονται ως εξής:

Τμήμα 1: από τον A

Τμήμα 2: από τον B

Τμήμα 3: από τον A

Τμήμα 4: από τον B

Τμήμα 5: από τον A

Τμήμα 6: από τον B

Ο απόγονος 2 παίρνει τα αντίστροφα.

Ομοιόμορφη Διασταύρωση

Διαμόρφωση των τμημάτων

Με τα σημεία αυτά, τα όρια είναι:

Τμήμα1: θέσεις **1 έως 1**

Τμήμα2: θέσεις **2 έως 4**

Τμήμα3: θέσεις **5 έως 6**

Τμήμα4: θέσεις **7 έως 7**

Τμήμα5: θέσεις **8 έως 10**

Τμήμα6: θέσεις **11 έως 11**

Ομοιόμορφη Διασταύρωση

Εξάγουμε τα τμήματα από τους γονείς:

A	B
$A[1:1] = 0$	$B[1:1] = 1$
$A[2:4] = 000$	$B[2:4] = 110$
$A[5:6] = 10$	$B[5:6] = 10$
$A[7:7] = 1$	$B[7:7] = 0$
$A[8:10] = 010$	$B[8:10] = 100$
$A[11:11] = 1$	$B[11:11] = 0$

Ομοιόμορφη Διασταύρωση

Σύνθεση Απογόνων (εναλλαγή)

1ος απόγονος (A-B-A-B-A-B):

Από A: 0

Από B: 110

Από A: 10

Από B: 0

Από A: 010

Από B: 0

0 | 110 | 10 | 0 | 010 | 0
= **01101011001**

Ομοιόμορφη Διασταύρωση

2ος απόγονος (B-A-B-A-B-A):

Από B: 1

Από A: 000

Από B: 10

Από A: 1

Από B: 100

Από A: 1

1 | 000 | 10 | 1 | 100 | 1
= **10001000100**

Μετάλλαξη

- ▶ Εκτελείται στους απογόνους που παράγονται από τη διασταύρωση
- ▶ Με βάση μια μικρή πιθανότητα, μπορούμε να αλλάξουμε το σύμβολο / τιμή σε κάθε γονίδιο του bit-string (συνήθως μόνο μια τιμή αλλάζει)
- ▶ Παράδειγμα: 011**0**10 -> 011**1**10

1^ο Παράδειγμα

- ▶ Πληθυσμός με 4 χρωματοσώματα και bit-strings μεγέθους 12 bit

Λύση	Αναπαράσταση	Καταλληλότητα	Πιθανότητα Επιλογής
A	000110010111	8	0.32 (8/25)
B	111010101100	6	0.24 (6/25)
Γ	001110101001	6	0.24 (6/25)
Δ	111011011100	5	0.20 (5/25)

- Αποτέλεσμα διαδικασίας επιλογής: [A,B,B,Γ]
- Τυχαία ζευγάρια είναι (B,A) και (B,Γ)
- Χρησιμοποιούμε διασταύρωση ενός σημείου και μετάλλαξη

1^ο Παράδειγμα

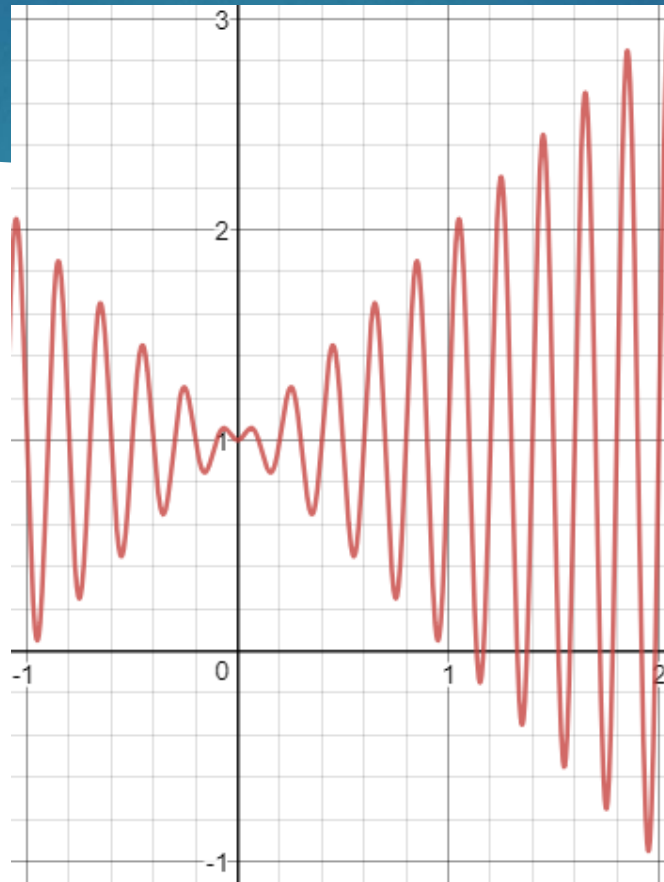
- ▶ Αποτέλεσμα πρώτου κύκλου για πρώτο ζευγάρι:

	1 ^{ος} Απόγονος	2 ^{ος} απόγονος
B	1110 10101100	1110 10101100
A	0001 10010111	0001 10010111
Διασταύρωση	111010010111	000110101100
Μετάλλαξη	111010010111	000110101100

- Αποτέλεσμα πρώτου κύκλου για δεύτερο ζευγάρι:

	1 ^{ος} Απόγονος	2 ^{ος} απόγονος
B	111010101 100	111010101 100
Γ	001110101001	001110101001
Διασταύρωση	111010101001	001110101100
Μετάλλαξη	111 <u>1</u> 10101001	00111010110<u>1</u>

2^ο Παράδειγμα



- ▶ Έστω η συνάρτηση: $f(x) = x \cdot \sin(10\pi x) + 1.0$
- ▶ Πρόβλημα: εύρεση μεγίστου στο διάστημα $[-1, 2]$

2^ο Παράδειγμα

- ▶ Η αντιμετώπιση του προβλήματος αναλυτικά είναι δύσκολη – χρήση γενετικών αλγορίθμων
- ▶ Απλή δυαδική κωδικοποίηση για τους αριθμούς με ακρίβεια 6 δεκαδικών ψηφίων
- ▶ Αντιστοίχιση μεταξύ δυαδικών και πραγματικών αριθμών στο εν λόγω διάστημα:

όπου x' είναι ο δεκαδικός αριθμός

$$x = -1 + x' \cdot \frac{3}{2^{22} - 1}$$

2^ο Παράδειγμα

- ▶ Η συνάρτηση καταλληλότητας είναι η $f(x)$
- ▶ Οι τεχνικές διασταύρωσης και μετάλλαξης είναι άμεσα εφαρμόσιμες
- ▶ Η συνάρτηση παρουσιάζει μέγιστο για $x \sim 1.85$ που αντιστοιχεί στην τιμή $f(1.85)=2.85$

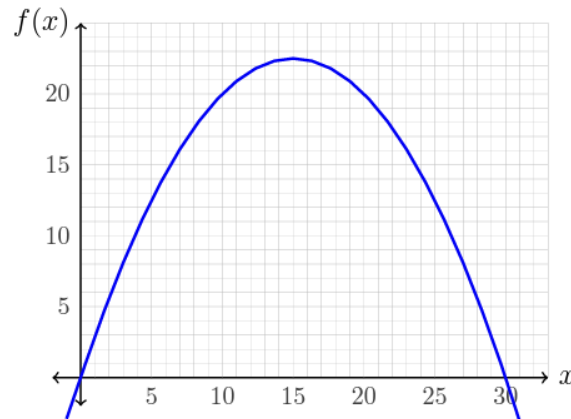
3^ο Παράδειγμα

Θεωρήστε το πρόβλημα της μεγιστοποίησης της συνάρτησης

$$f(x) = \frac{-x^2}{10} + 3x$$

όπου το x κυμαίνεται μεταξύ 0 και 31.

Figure 1: Graph of $f(x) = \frac{-x^2}{10} + 3x$



3^ο Παράδειγμα

Για λύσουμε αυτό πρόβλημα χρησιμοποιώντας έναν γενετικό αλγόριθμο, πρέπει να κωδικοποιήσουμε τις πιθανές τιμές του x ως χρωμοσώματα.

Για αυτό το παράδειγμα, θα κωδικοποιήσουμε το x ως δυαδικό ακέραιο αριθμό μήκους 5. Έτσι, τα χρωμοσώματα για τον γενετικό μας αλγόριθμο θα είναι ακολουθίες από 0 και 1 με μήκος 5 bits και θα έχουν εύρος από 0 (00000) έως 31 (11111). Για να ξεκινήσουμε τον αλγόριθμο, επιλέγουμε τυχαία έναν αρχικό πληθυσμό 10 χρωμοσωμάτων.

Ο αρχικός πληθυσμός χρωμοσωμάτων που προκύπτει παρουσιάζεται στον πίνακα 1. Στη συνέχεια, παίρνουμε την τιμή x που αντιπροσωπεύει κάθε χρωμόσωμα και ελέγχουμε την καταλληλότητά του με τη συνάρτηση καταλληλότητας.

3^ο Παράδειγμα

Table 1: Initial Population

Chromosome Number	Initial Population	x Value	Fitness Value $f(x)$	Selection Probability
1	01011	11	20.9	0.1416
2	11010	26	10.4	0.0705
3	00010	2	5.6	0.0379
4	01110	14	22.4	0.1518
5	01100	12	21.6	0.1463
6	11110	30	0	0
7	10110	22	17.6	0.1192
8	01001	9	18.9	0.1280
9	00011	3	8.1	0.0549
10	10001	17	22.1	0.1497
Sum			147.6	
Average			14.76	
Max			22.4	

3^ο Παράδειγμα

Επιλέγουμε τα χρωμοσώματα που θα αναπαραχθούν με βάση τις τιμές καταλληλότητάς τους, χρησιμοποιώντας την ακόλουθη πιθανότητα:

$$P(\text{chromosome } i \text{ reproduces}) = \frac{f(x_i)}{\sum_{k=1}^{10} f(x_k)}$$

Ο Γκόλντμπεργκ παρομοιάζει αυτή τη διαδικασία με την περιστροφή ενός σταθμισμένου τροχού ρουλέτας. Δεδομένου ότι ο πληθυσμός μας έχει 10 χρωμοσώματα και κάθε «ζευγάρι» παράγει 2 απογόνους, χρειαζόμαστε 5 ζευγαρώματα για να παράγουμε μια νέα γενιά με 10 χρωμοσώματα. Τα επιλεγμένα χρωμοσώματα εμφανίζονται στον πίνακα 2. Για τη δημιουργία των απογόνων τους, επιλέγεται τυχαία ένα σημείο διασταύρωσης, το οποίο εμφανίζεται στον πίνακα ως κάθετη γραμμή. Σημειώστε ότι είναι δυνατόν να μην πραγματοποιηθεί διασταύρωση, οπότε οι απόγονοι είναι ακριβή αντίγραφα των γονέων τους.

3^ο Παράδειγμα

Table 2: Reproduction & Second Generation

Chromosome Number	Mating Pairs	New Population	x Value	Fitness Value $f(x)$
5	01 100	01010	10	20
2	11 010	11100	28	5.6
4	0111 0	01111	15	22.5
8	0100 1	01000	8	17.6
9	0001 1	01010	10	20
2	1101 0	11011	27	8.1
7	10110	10110	22	17.6
4	01110	01110	14	22.4
10	100 01	10001	17	22.1
8	010 01	01001	9	18.9
			Sum	174.8
			Average	17.48
			Max	22.5

3^ο Παράδειγμα

Τέλος, κάθε κομμάτι των νέων χρωμοσωμάτων μεταλλάσσεται με χαμηλή πιθανότητα. Για αυτό το παράδειγμα, αφήνουμε την πιθανότητα μετάλλαξης να είναι 0,001. Με 50 συνολικά μεταφερόμενες θέσεις bit, αναμένουμε να μεταλλαχθούν $50 \times 0,001 = 0,05$ bit. Συνεπώς, είναι πιθανό να μην μεταλλαχθεί κανένα bit στη δεύτερη γενιά. Για λόγους επεξήγησης, έχουμε μεταλλάξει ένα bit στο νέο πληθυσμό, το οποίο εμφανίζεται με έντονη γραφή στον Πίνακα 2.

3^ο Παράδειγμα

Αφού ολοκληρωθούν η επιλογή, η διασταύρωση και η μετάλλαξη, ο νέος πληθυσμός δοκιμάζεται με τη συνάρτηση καταλληλότητας. Οι τιμές καταλληλότητας αυτής της δεύτερης γενιάς παρατίθενται στον πίνακα 2. Παρόλο που η εξαγωγή συμπερασμάτων από μία μόνο δοκιμή ενός αλγορίθμου με βάση την πιθανότητα είναι, «στην καλύτερη περίπτωση, μια επικίνδυνη υπόθεση», το παράδειγμα αυτό δείχνει πώς οι γενετικοί αλγόριθμοι «εξελισσονται» προς καταλληλότερες υποψήφια λύσεις. Συγκρίνοντας τον Πίνακα 2 με τον Πίνακα 1, βλέπουμε ότι τόσο η μέγιστη όσο και η μέση καταλληλότητα του πληθυσμού έχουν αυξηθεί μετά από μία μόνο γενιά.

Εφαρμογές

- ▶ Εύρεση μέγιστης τιμής για αριθμητικές συναρτήσεις
 - ▶ Γενικά δύσκολη για συναρτήσεις πολλαπλών μεταβλητών με ασυνέχειες και θόρυβο
 - ▶ Πλεονέκτημα: η συνάρτηση καταλληλότητας δίδεται
- ▶ Επεξεργασία εικόνων:
 - ▶ Αναγνώριση προτύπων (πχ. αντικείμενα) σε εικόνες
- ▶ Συνδυαστική βελτιστοποίηση
 - ▶ Χρήση σε κλασικά προβλήματα βελτιστοποίησης για την κατανομή πόρων σε δραστηριότητες (πχ. πλανόδιου πωλητή)
 - ▶ Συνδυαστική έκρηξη του χώρου αναζήτησης με βάση το μέγεθος του προβλήματος

Εφαρμογές

- ▶ Μηχανική
 - ▶ Σχεδίαση κατασκευών (πχ. γέφυρες) και εξαρτημάτων
- ▶ Μηχανική μάθηση
 - ▶ Προσέγγιση συναρτήσεων
 - ▶ Ανακάλυψη κανόνων συσχέτισης
 - ▶ Συστήματα κατηγοριοποίησης
- ▶ Παιχνίδια, επίλυση λαβυρίνθων, πολιτικές και οικονομικές αναλύσεις

Μηχανική Μάθηση & Εξόρυξη Γνώσης

Ερωτήσεις
?

Βιβλιογραφία

- ▶ Ι. Βλαχάβας, Π. Κεφαλάς, Ν. Βασιλειάδης, Φ. Κόκκορας, Η. Σακελλαρίου, Τεχνητή Νοημοσύνη - Γ' Έκδοση, ISBN: 978-960-8396-64-7, Έκδοση/Διάθεση: Εκδόσεις Πανεπιστημίου Μακεδονίας, 2011
- ▶ Ian H. Witten and Eibe Frank. 2005. Data Mining: Practical Machine Learning Tools and Techniques, Second Edition (Morgan Kaufmann Series in Data Management Systems). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.
- ▶ Κ. Διαμαντάρας, Ι. Μπότσης, Μηχανική Μάθηση – Α' Έκδοση, ISBN: 978-960-461-955-5, Εκδόσεις Κλειδάριθμος, 2019
- ▶ P.-N. Tan, M.Steinbach, V. Kumar, «Introduction to Data Mining», Addison Wesley, 2006
- ▶ An Introduction to Genetic Algorithms Jenna Carr
<https://www.whitman.edu/Documents/Academics/Mathematics/2014/carrjk.pdf>