Πρόγραμμα Μεταπτυχιακών Σπουδών
"Προηγμένες Τεχνολογίες Υπολογιστών και Πληροφορικής"

Υπολογιστική Νοημοσύνη

Μάθημα 4ο

Δρ. Βασίλειος Γ. Καμπουρλάζος
Δρ. Ανέστης Γ. Χατζημιχαηλίδης

Τμήμα Μηχανικών Πληροφορικής Τ.Ε.
ΤΕΙ Ανατολικής Μακεδονίας και Θράκης

2017-2018

Περιεχόμενα

Διευρυμένη Υπολογιστική Νοημοσύνη (ΥΝ)

- Επεκτάσεις της Κλασικής ΥΝ.
- Μεθοδολογίες Στατιστικές, Πιθανοτήτων, Δυνατοτήτων.
- Μεθοδολογίες με Γράφους.

Επεκτάσεις της Κλασικής ΥΝ

- Νευρο-Ασαφή Συστήματα.
- Δίκτυα Ακτινωτής Βάσης.
- Μοντέλα k Πλησιέστερων Γειτόνων.
- Μηχανές Διανυσμάτων Στήριξης.
- Σύγχρονες Τάσεις.

Η κλασική ΥΝ διευρύνθηκε κατ’ αρχήν για να ξεπεραστούν κάποια από τα επιμέρους μειονεκτήματα τεχνολογιών της κλασικής ΥΝ.

Νευρο-Ασαφή Συστήματα

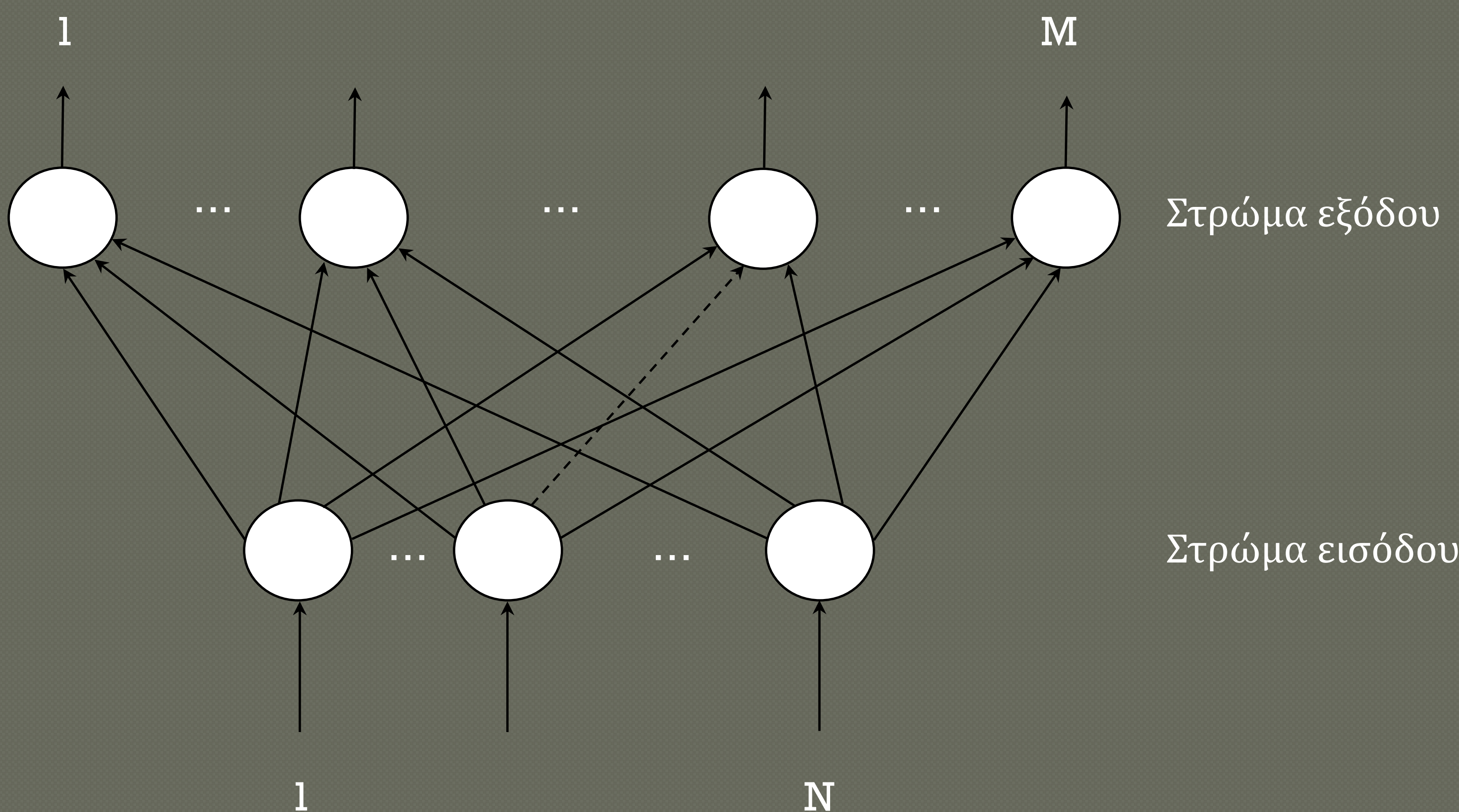
Από τη μια μεριά, τα κλασικά ΤΝΔ μπορούν να μαθαίνουν γρήγορα , αλλά λειτουργούν ως «μαύρα κουτιά» μέσα στα οποία δεν μπορούμε να δούμε ώστε να αιτιολογήσουμε τις απαντήσεις τους.

Από τη άλλη μεριά, τα κλασικά ασαφή συστήματα μπορούν μεν να εξηγήσουν ικανοποιητικά τις απαντήσεις τους, αλλά δεν μπορούν να μαθαίνουν.

Συνδυασμός τεχνητών νευρωνικών δικτύων και ασαφών συστημάτων ώστε να ξεπεραστούν μειονεκτήματα της κάθε τεχνολογίας αποτελεί η νευρωνική υλοποίηση ενός ασαφούς συστήματος συμπερασμού που συνδυάζει πλεονεκτήματα των δύο.

Έτσι προέκυψαν τα **νευρο-ασαφή συστήματα** (**ΝΑΣ**) με την πλέον δημοφιλή μορφή ΝΑΣ να είναι τα **Προσαρμοστικά Νευρο-Ασαφή Συστήματα Συμπερασμού** που αποτελούν μια νευρωνική υλοποίηση ενός ασαφούς συστήματος συμπερασμού.

Νευρο-Ασαφές Σύστημα τύπου Mamdani



Στόχος

Δοθέντος ενός συνόλου ζευγών $(x_1,f(x_1)),\ldots,(x_n,f(x_n))$ μιας άγνωστης συνάρτησης f: $R^N \rightarrow R^M$, το τυπικό πρόβλημα είναι να υπολογιστεί μια ικανοποιητική προσέγγιση f': $R^N \rightarrow R^M$ της συνάρτησης f, ώστε για $x_0 \in R^N$, με $x_0 \neq x_i$, $i \in \{1,\ldots,n\}$, το εκτιμώμενο $f'(x_0)$ να είναι όσο το δυνατόν πιο κοντά στο πραγματικό $f(x_0)$, δηλ. η υπολογισθείσα συνάρτηση f' να έχει καλή ικανότητα γενίκευσης.

Η μέθοδος

Συναρτήσεις συμμετοχής είναι αποθηκευμένες:
(1) Στα βάρη που συνδέουν τους νευρώνες εισόδου - εξόδου, και
(2) Στους νευρώνες εξόδου.

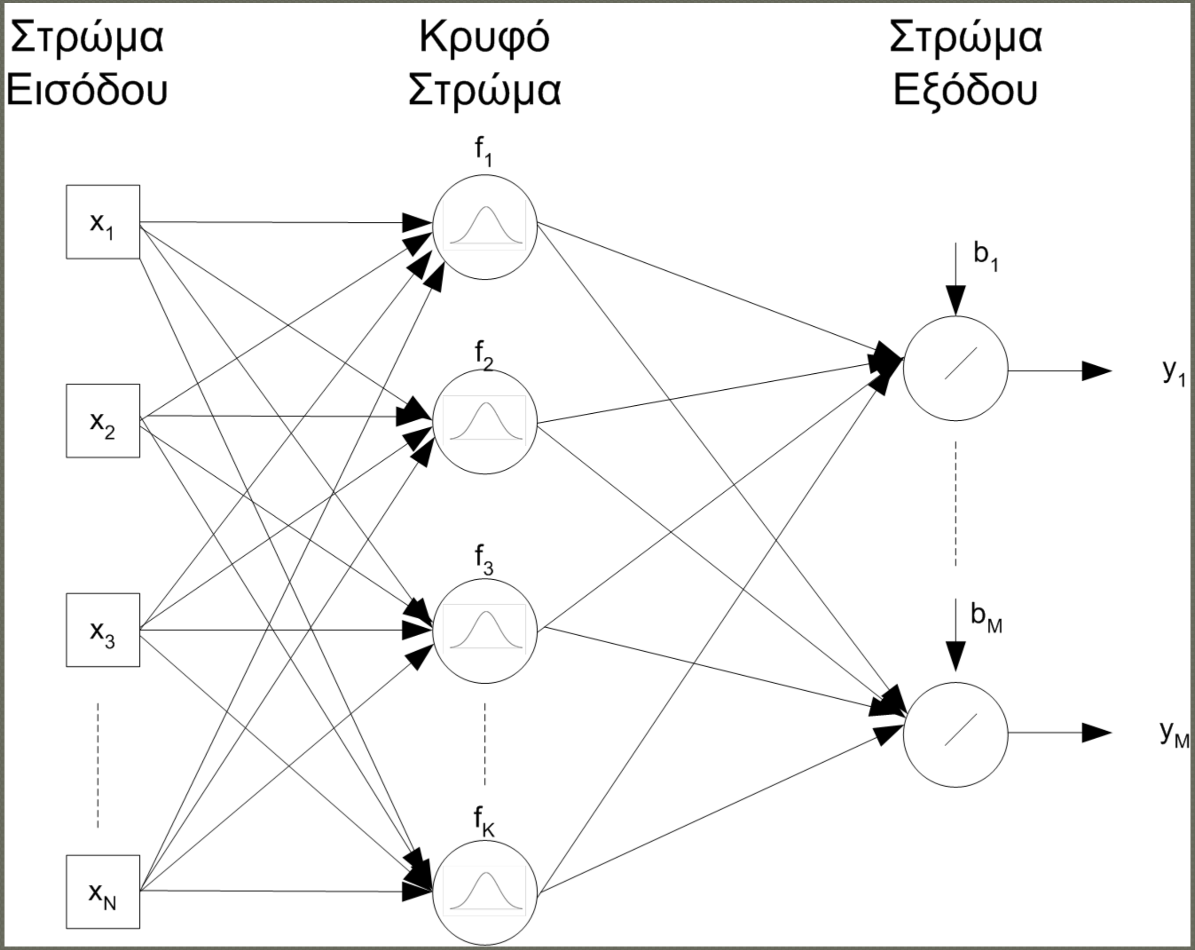
Οι μηχανισμοί ασαφοποίησης /συμπερασμού /από-ασαφοποίησης είναι δεδομένοι.

Προσαρμόζεται τόσο το σχήμα όσο και η θέση των συναρτήσεων συμμετοχής ώστε να ελαχιστοποιείται μια καλώς ορισμένη συνάρτηση σφάλματος

Δίκτυα Ακτινωτής Βάσης (ΔΑΒ)

- Τα δίκτυα αυτά είναι προσωτροφοδοτούμενα και περιλαμβάνουν ένα στρώμα νευρώνων εισόδου, ένα κρυφό στρώμα και ένα στρώμα εξόδου.
- Κύριο χαρακτηριστικό των ΔΑΒ είναι η εφαρμογή *ακτινωτών συναρτήσεων ενεργοποίησης* στους νευρώνες του κρυφού στρώματος, ενώ οι έξοδοι των κρυφών νευρώνων αθροίζονται σταθμισμένα στο στρώμα εξόδου.

Αρχιτεκτονική ΔΑΒ



Βασικές Εξισώσεις ΔΑΒ

- Έξοδος $y_{\mathrm{i}} = \sum_{j=1}^{K} w_{\mathrm{ji}} f_{\mathrm{j}}(\|\mathbf{x} - \mathbf{c}_{\mathrm{j}}\|) + b_{\mathrm{i}}$

- Συναρτήσεις ενεργοποίησης (Gaussian):

$$f_{\mathrm{j}}(\|\mathbf{x} - \mathbf{c}\|) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{c}\|^2}{2\sigma^2} \right)$$

Χρήση των ΔΑΒ

- Τα ΔΑΒ αποτελούν έναν **καθολικό προσεγγιστή**.

- Η λειτουργία των ΔΑΒ βασίζεται στην αρχή του μετασχηματισμού του προβλήματος από τον χώρο των εισόδων στον χώρο των χαρακτηριστικών πολύ μεγαλύτερης διάστασης:
Ένα μη-γραμμικώς διαχωρίσιμο πρόβλημα κατηγοριοποίησης μπορεί να μετατραπεί σε γραμμικώς διαχωρίσιμο αν μετασχηματιστεί σε ένα χώρο αρκετά μεγάλης διάστασης.

Εκπαίδευση των ΔΑΒ

- Ένα ΔΑΒ αποτελείται από δύο τμήματα που λειτουργούν εντελώς διαφορετικά μεταξύ τους. Επομένως, η εκπαίδευση των κρυφών νευρώνων και των νευρώνων εξόδου, μπορεί να πραγματοποιηθεί με διαφορετικό τρόπο και σε διαφορετικό χρόνο.

- Βελτιστοποίηση (α) των κέντρων $\mathbf{c}_1,\mathbf{c}_2,\ldots,\mathbf{c}_K$ των κρυφών νευρώνων, (β) των βαρών $\{\mathbf{w}_{1j},\mathbf{w}_{2j},\ldots,\mathbf{w}_{Kj}\}$, $j=1,\ldots,M$ των συνδέσεων με το στρώμα εξόδου.

Τα κέντρα $\mathbf{c}_1,\mathbf{c}_2,\ldots,\mathbf{c}_K$ των κρυφών νευρώνων υπάρχουν μπορούν να επιλεγούν ως ακολούθως :

- να επιλεγούν τυχαία από το σύνολο των δεδομένων εκπαίδευσης,
- να υπολογιστούν με μια μέθοδο ομαδοποίησης (π.χ. με τον **αλγόριθμο κ-μέσων**),
- να υπολογιστούν με εκπαίδευση *εποπτείας*, π.χ. με την μέθοδο **κατάβασης βαθμίδας**.

- Μια συνηθισμένη επιλογή για την εύρεση των βαρών $\{\mathbf{w}_{1j},\mathbf{w}_{2j},\ldots,\mathbf{w}_{Kj}\}$, j=1,…,M είναι η **μέθοδος του ψευδο-αντίστροφου**:

$$\mathbf{w} = \mathbf{F}^{+}\mathbf{t},$$

όπου **t** είναι το διάνυσμα των επιθυμητών εξόδων του συνόλου εκπαίδευσης και $\mathbf{F}^{+}$ είναι ο **ψευδοαντίστροφος** (μπορεί να υπολογιστεί με την μέθοδο της **ανάλυσης ίδιων τιμών**) του πίνακα που περιγράφει τις χρησιμοποιούμενες συναρτήσεις ενεργοποίησης.

Μοντέλο κ Πλησιέστερων Γειτόνων (κΠΓ)

- Ένα απλό και συνήθως υψηλής απόδοσης μοντέλο ΥΝ που εφαρμόζεται τόσο σε προβλήματα **κατηγοριοποίησης** και **παλινδρόμησης**.

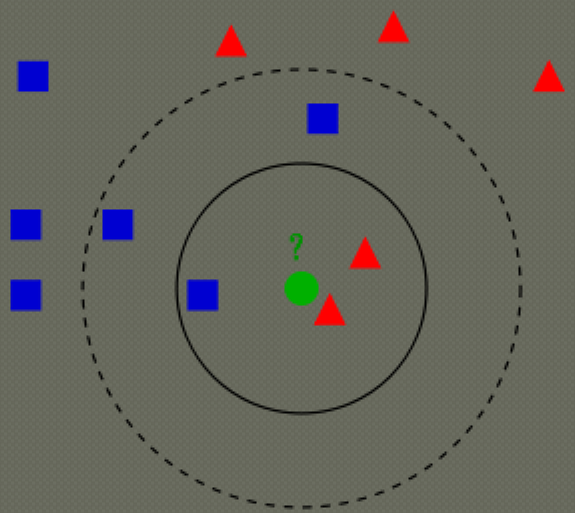
- Το κΠΓ υπολογίζει τις αποστάσεις του νέου δεδομένου από τα δεδομένα εκπαίδευσης.

- Στη συνέχεια θεωρούμε ένα πρόβλημα κατηγοριοποίησης.

- Υπολογίζονται οι αποστάσεις *dist*(.,.) ενός δεδομένου $\mathbf{x}_0$ εισόδου από κάθε ένα από τα δεδομένα εκπαίδευσης:

$$d_i = dist(\mathbf{x}_0, \mathbf{x}_i),\ i= 1,2,\ldots,N$$

- Το $\mathbf{x}_0$ κατηγοριοποιείται σε εκείνη την κατηγορία που ανήκουν τα περισσότερα από τα *κ* δεδομένα με την μικρότερη απόσταση από το $\mathbf{x}_0$.



- Το νέο δεδομένο (βλ. πράσινη τελεία) θα κατηγοριοποιηθεί στην κατηγορία των τριγώνων εάν *κ=3*, ενώ για *κ=5* θα κατηγοριοποιηθεί στην κατηγορία των τετραγώνων.

- Η απόδοση του κΠΓ μοντέλου εξαρτάται από την επιλογή της παραμέτρου *κ*. Μεγάλη τιμή *κ* έχει σαν αποτέλεσμα **ευρωστία** στον θόρυβο.

- Έχουν προταθεί μέθοδοι προ-επεξεργασίας των δεδομένων ώστε δεδομένα τις ίδιας κατηγορίας να βρεθούν σχετικά κοντά μεταξύ τους.

- Έχουν προταθεί διάφορες συναρτήσεις *dist*(.,.) για την μέτρηση της απόστασης των δεδομένων με τις ακόλουθες να είναι οι πιο δημοφιλείς:

$$dist_{Euclidean}(\mathbf{x}_0, \mathbf{x}_{\mathrm{j}}) = \sqrt{\sum_{i=1}^{n} \left(\mathbf{x}_0^i - \mathbf{x}_j^i \right)^2}$$

$$dist_{CityBlock}(\mathbf{x}_0, \mathbf{x}_{\mathrm{j}}) = \sum_{i=1}^{n} \left| \mathbf{x}_0^i - \mathbf{x}_j^i \right|$$

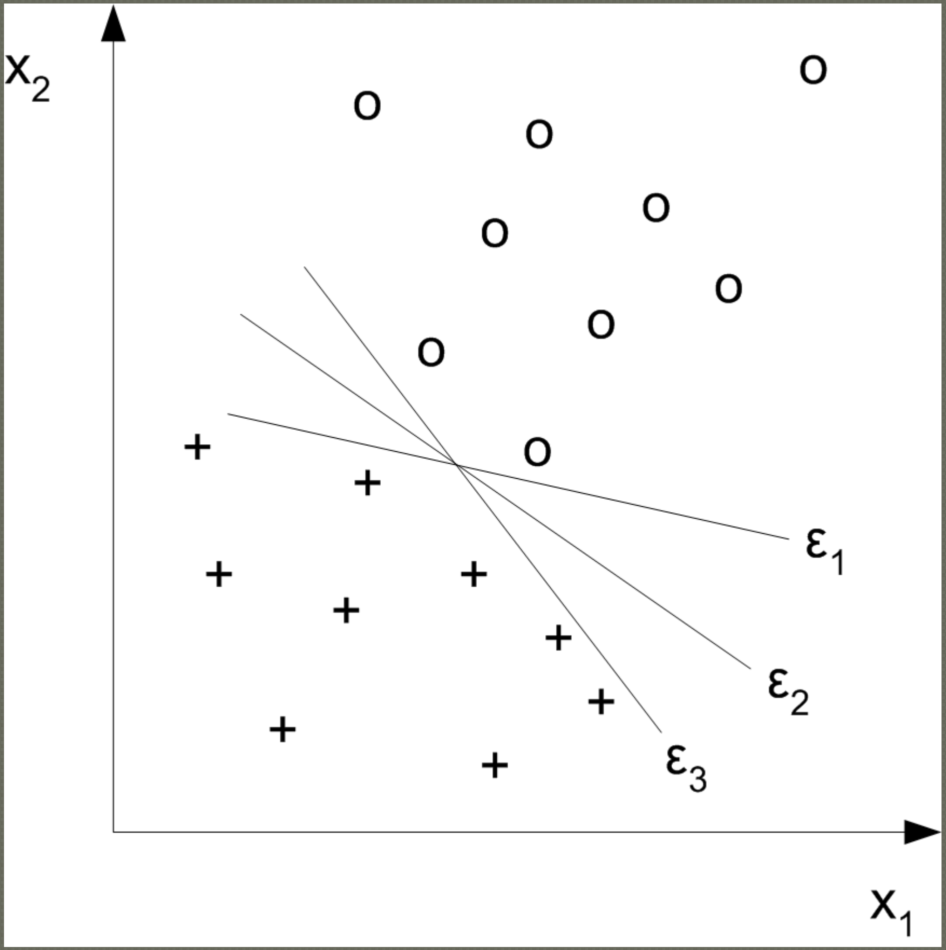
$$dist_{Chebyshev}(\mathbf{x}_0, \mathbf{x}_{\mathrm{j}}) = \max_{i} \left(\left| \mathbf{x}_0^i - \mathbf{x}_j^i \right| \right)$$

Μηχανές Διανυσμάτων Στήριξης (ΜΔΣ)

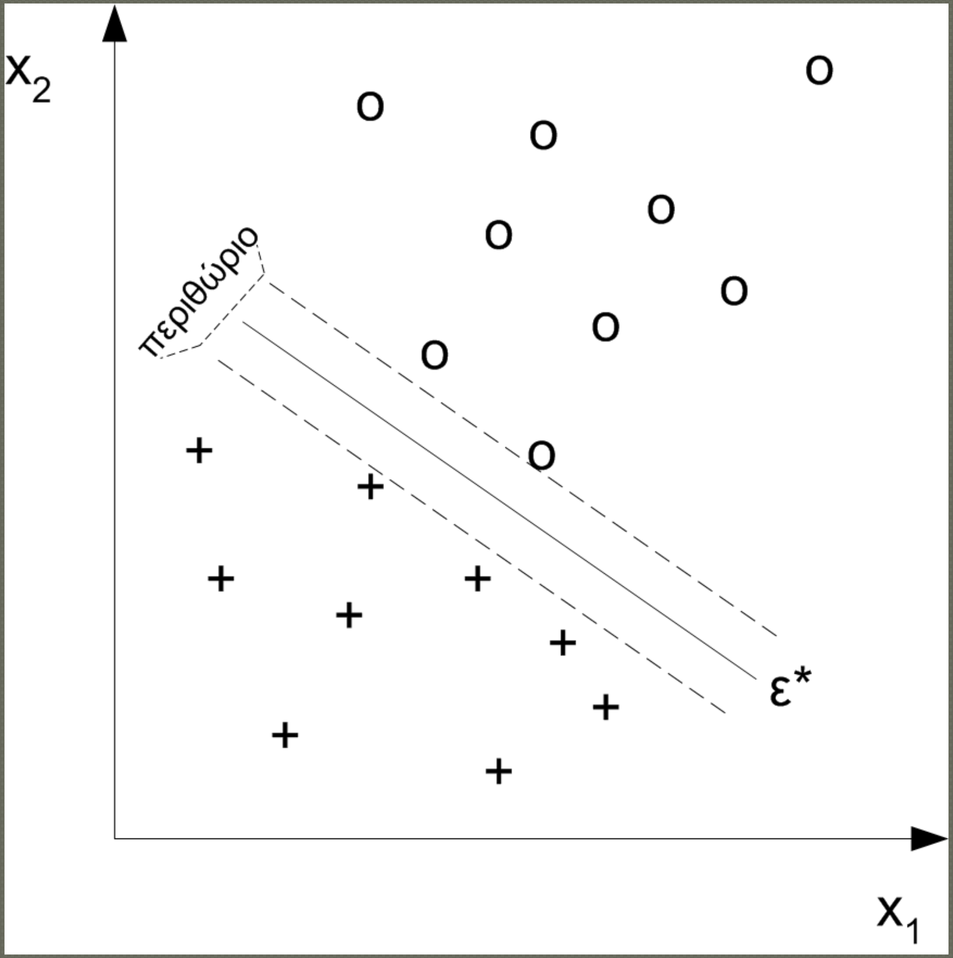
- Αρχικά προτάθηκαν για δύο κατηγορίες.

- Επιδιώκουν τον υπολογισμό ενός υπερ-επίπεδου, που παίζει τον ρόλο επιφάνειας λήψης απόφασης, ώστε το περιθώριο διαχωρισμού των κατηγοριών να μεγιστοποιείται όπως εξηγείται στη συνέχεια.

γραμμική διαχωρισιμότητα (πολλές λύσεις)



ΜΔΣ προτείνει μία (βέλτιστη) λύση που είναι η ευθεία ε*



Γραμμικές ΜΔΣ (υπολογισμός της ευθείας ε)*

Διαχωρίσιμα Δεδομένα

- Η ευθεία ε* υπολογίζεται έτσι ώστε το περιθώριο διαχωρισμού των κατηγοριών να μεγιστοποιείται.
- Η ευθεία ε* καθορίζεται από τα λεγόμενα **διανύσματα στήριξης** τα οποία βρίσκονται πιο κοντά στην ευθεία ε*.

Διατύπωση του προβλήματος

- Ζητούνται το διάνυσμα **w** και η σταθερά πόλωσης b που επαληθεύουν την εξίσωση $\mathbf{w}^{\mathrm{T}}\mathbf{x}+b = 0$ ώστε να ελαχιστοποιείται η ακόλουθη συνάρτηση

$$J(\mathbf{w}) = \frac{1}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w}$$

όπου $t_k(\mathbf{w}^{\mathrm{T}}\mathbf{x}_k+b) \geq +1$ με $t_k \in \{-1,+1\}$, $\mathbf{k} \in \{1,\ldots,N\}$ και $\mathbf{x}_k \in R^n$ είναι τα δεδομένα εκπαίδευσης.

- Η επίλυση βρίσκεται με την συνάρτηση **Lagrange**

$$L(\mathbf{w},b,\boldsymbol{a}) = \frac{1}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w} - \sum_{k=1}^{N} a_k \left[\mathrm{t}_{\mathrm{k}} (\mathbf{w}^{\mathrm{T}}\mathbf{x}_{\mathrm{k}} + b) - 1 \right]$$

όπου οι συντελεστές $\alpha_k \geq 0$, k=1,…,N ονομάζονται **πολλαπλασιαστές Lagrange**.

- Το πρόβλημα μετασχηματίζεται στο πρόβλημα ελαχιστοποίησης της ακόλουθης συνάρτησης

$$Q(\boldsymbol{a}) = \sum_{k=1}^{N} a_k - \frac{1}{2}\sum_{\ell=1}^{N}\sum_{m=1}^{N} a_\ell a_m \mathrm{t}_\ell \mathrm{t}_m \mathbf{x}_\ell^{\mathrm{T}} \mathbf{x}_m$$

όπου $\sum_{k=1}^{N} a_k \mathrm{t}_k = 0$ με $0 \leq \alpha_{\mathrm{k}}$, k=1,…,N .

Οι μη-μηδενικές λύσεις $\alpha_{\mathrm{k}} \geq 0$, k=1,…,N (πολλαπλασιαστές Lagrange) που προκύπτουν είναι τα ζητούμενα διανύσματα στήριξης.

Μη-διαχωρίσιμα Δεδομένα

- Το πρόβλημα μετασχηματίζεται στο πρόβλημα ελαχιστοποίησης της ακόλουθης συνάρτησης

$$Q(\boldsymbol{a}) = \sum_{k=1}^{N} a_k - \frac{1}{2}\sum_{\ell=1}^{N}\sum_{m=1}^{N} a_\ell a_m \mathrm{t}_\ell \mathrm{t}_m \mathbf{x}_\ell^{\mathrm{T}} \mathbf{x}_m$$

όπου $\sum_{k=1}^{N} a_k \mathrm{t}_k = 0$ με $0 \leq \alpha_{\mathrm{k}} \leq \mathrm{c}$, k=1,…,N .

Ίδιο πρόβλημα όπως προηγουμένως με τον επιπλέον περιορισμό "$\alpha_{\mathrm{k}} \leq \mathrm{c}$", όπου η σταθερά "c" υπολογίζεται πειραματικά.

Μη-γραμμικές ΜΔΣ

- «*Ένα μη-γραμμικώς διαχωρίσιμο πρόβλημα αναγνώρισης προτύπων, μπορεί να μετασχηματιστεί σε γραμμικώς διαχωρίσιμο σε ένα χώρο περισσότερων διαστάσεων*»

(Cover, 1965)

- Χρησιμοποιείται μια μη-γραμμική συνάρτηση $\varphi: R^n \rightarrow R^m$ από τον χώρο εισόδων R^n στο χώρο R^m των χώρο των χαρακτηριστικών με $\mathbf{n} << \mathbf{m}$. Έτσι, αντί της εξίσωσης $\mathbf{w}^T\mathbf{x}+b = 0$ έχουμε την
- $$\sum_{i=1}^{m} w_i \varphi_i(\mathbf{x}) + b = 0$$

- Το αντίστοιχο πρόβλημα είναι η ελαχιστοποίηση της ακόλουθης συνάρτησης

$$Q(\boldsymbol{a}) = \sum_{k=1}^{N} a_k - \frac{1}{2}\sum_{\ell=1}^{N}\sum_{m=1}^{N} a_\ell a_m \mathrm{t}_\ell \mathrm{t}_m \mathrm{K}(\mathbf{x}_\ell, \mathbf{x}_m)$$

όπου $\sum_{k=1}^{N} a_k \mathrm{t}_k = 0$ με $0 \leq \alpha_{\mathrm{k}}$, k=1,…,N .

Η ποσότητα **K(x$_\ell$,x$_\mathrm{m}$)** καλείται **πυρήνας** και η επιλογή του παίζει σημαντικό ρόλο.

Διάσταση VC (Vapnik-Chervonenkis)

- Μηχανές (βλ. αλγόριθμοι) μάθησης, σε κάποιες περιπτώσεις, χαρακτηρίζονται από έναν ακέραιο αριθμό ο οποίος καλείται **διάσταση Vapnik-Chervonenkis** ή **διαVC** για συντομία.
- Η δια**VC** μιας μηχανής μπορεί να οριοθετήσει την ικανότητα μιας μηχανής για μάθηση.
- Ορίζουμε την δια**VC** στη βάση της έννοιας του *θρυμματισμού* όπως εξηγείται στη συνέχεια.

- Έστω X το σύνολο των δεδομένων ενδιαφέροντος.
- Μία **έννοια** $\boldsymbol{c}$ πάνω στο X ορίζεται ως ένα υποσύνολο του X, δηλ. $c \subseteq X$.
- Έστω $\boldsymbol{C}$ το σύνολο όλων των εννοιών τις οποίες δυνητικά μπορεί να μάθει μια συγκεκριμένη μηχανή μάθησης.

- Το ενδιαφέρον μας εστιάζεται σε διακριτά υποσύνολα S του X. Έστω D το σύνολο των διακριτών υποσυνόλων του X.

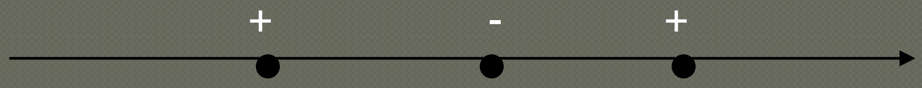
- Δοθέντος του συνόλου C εννοιών, ορίζουμε την συνάρτηση Π_C: $D \rightarrow 2^D$, με τον ακόλουθο τύπο $\Pi_C(S) = \{c \cap S \mid c \in C\}$. Κάθε στοιχείο του συνόλου $\Pi_C(S)$ καλείται **διχοτόμηση** του S.

- Εάν $|\Pi_C(S)| = 2^{|S|}$ τότε λέμε ότι το υποσύνολο S **θρυμματίζεται** από το σύνολο C.

- Η **διάσταση Vapnik-Chervonenkis** (**διαVC**) ορίζεται ως η μεγαλύτερη *πληθικότητα* συνόλου $S \subseteq X$ το οποίο μπορεί να θρυμματιστεί από το C.
- Τα ακόλουθα παραδείγματα είναι ενδεικτικά.

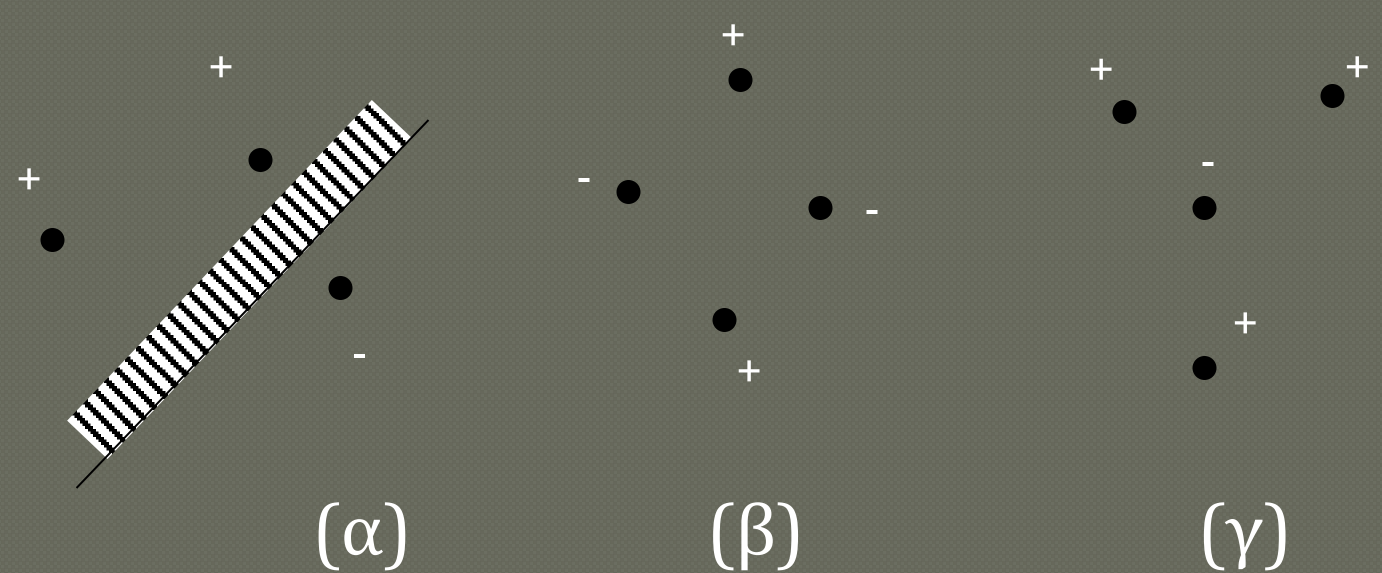
Παράδειγμα Α

- *X* είναι η ευθεία των πραγματικών αριθμών, ενώ *C* είναι το σύνολο των διαστημάτων της μορφής [*a*,*b*]. Το *C* μπορεί να θρυμματίσει οποιοδήποτε σύνολο 2 σημείων, αλλά κανένα σύνολο 3 σημείων διότι δεν μπορεί να πραγματοποιήσει τη διχοτόμηση που φαίνεται στο ακόλουθο Σχήμα. Συνεπώς, δια**VC** = 2.



Παράδειγμα Β

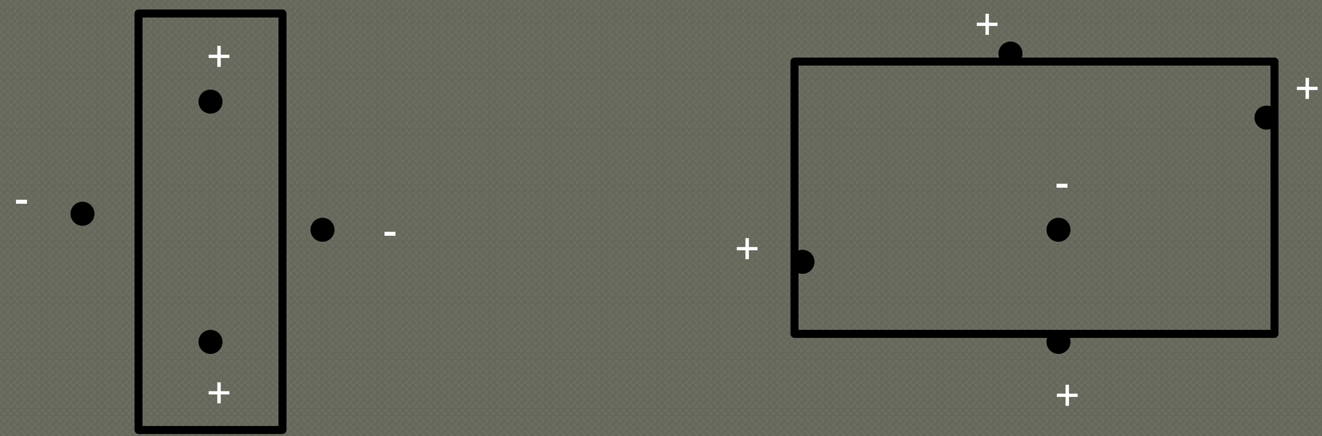
- *X* είναι το σύνολο των σημείων στο επίπεδο, ενώ *C* είναι οι γραμμικοί ημιχώροι (ημι-επίπεδα). Συνεπώς, δια**VC** = 3.



- Το Παράδειγμα Β μπορεί να επεκταθεί σε γενικό χώρο R^d, όπου προκύπτει δια**VC**= d+1. Συνεπώς, ένα γραμμικό ΤΝΔ τύπου **Perceptron** με d εισόδους και 1 έξοδο αποτελεί μια μηχανή μάθησης με **VC** διάσταση ίση με d+1.

Παράδειγμα Γ

- *X* είναι το σύνολο των σημείων στο επίπεδο, ενώ *C* είναι το σύνολο των ορθογώνιων παραλληλογράμμων. Συνεπώς, δια**VC** = 4.



(α) (β)

- Η **VC** διάσταση υπολογίζει ένα πιθανοτικό άνω όριο στο σφάλμα εξέτασης ενός μοντέλου κατηγοριοποίησης ως ακολούθως

$$P\left(\sigma\varphi E\xi\acute{\varepsilon}\tau \le \sigma\varphi E\kappa\pi + \sqrt{\frac{h(\log(2N/h)+1)-\log(\eta/4)}{N}} \right) = 1-\eta$$

όπου h είναι η **VC** διάσταση του μοντέλου κατηγοριοποίησης, $0 \le \eta \le 1$, και N το μέγεθος του υποσυνόλου εκπαίδευσης (θεωρούμε ότι $h << N$).

Σύγχρονες Τάσεις

- Όταν ένα μοντέλο της κλασικής ΥΝ υλοποιείται σε λογισμικό για εφαρμογή σε ένα συγκεκριμένο πρόβλημα τότε συχνά γίνονται επεκτάσεις με αποτέλεσμα να εμφανίζονται νέοι αλγόριθμοι όπως αλγόριθμοι instance-based, στατιστικοί, case-based reasoning, machine learning, data mining, causality.

- Ένα σημαντικό πρόβλημα των κλασικών ΤΝΔ είναι η υπολογιστική τους πολυπλοκότητα, δηλ. ο αριθμός των πράξεων για μάθηση, διότι χρησιμοποιούνται αλγόριθμοι *κατάβασης βαθμίδας* οι οποίοι υπολογιστικά είναι πολύ αργοί και, επιπλέον, συχνά παγιδεύονται σε τοπικά ελάχιστα.

- Μια **Μηχανή Ακραίας Μάθησης** (**ΜΑΜ**) είναι ένα προσωτροφοδοτούμενο ΤΝΔ τριών στρωμάτων με N νευρώνες στο κρυμμένο στρώμα, τυχαία επιλεγμένα βάρη εισόδου και τυχαίες τιμές *σταθερών πόλωσης* στους νευρώνες του κρυμμένου στρώματος, ενώ τα βάρη στην έξοδο υπολογίζονται με έναν πολλαπλασιασμό πινάκων.

- Μια ΜΑΜ μπορεί να μάθει με ακρίβεια N δείγματα, ενώ η ταχύτητα μάθησης μπορεί να είναι ακόμα και χιλιάδες φορές μεγαλύτερη από την ταχύτητα των συμβατικών ΤΝΔ τριών στρωμάτων με *οπισθόδρομη μάθηση*.

- Πέρα από τα κλασικά μοντέλα ΥΝ, τα οποία λαμβάνουν ως εισόδους διανύσματα πραγματικών αριθμών έχουν προταθεί ΤΝΔ με εισόδους διανύσματα μιγαδικών αριθμών με σκοπό να βελτιώσουν την αναπαράσταση των εισόδων τους. Παρά τα οριακά πλεονεκτήματα της αναπαράστασης με μιγαδικούς αριθμούς τα συνηθισμένα μειονεκτήματα των ΤΝΔ, όπως είναι η ερμηνεία των απαντήσεών τους, παραμένουν.

- Αναφορικά με τα ασαφή συστήματα, έχουν προταθεί πολλές επεκτάσεις της έννοιας ασαφές σύνολο όπως **τραχιά σύνολα**, **διαισθητικά ασαφή σύνολα**, κ.ά. με μαθηματικό κυρίως ενδιαφέρον.

Στοιχεία Επικοινωνίας

Δρ. Βασίλειος Γ. Καμπουρλάζος

vgkabs@teiemt. gr

Τηλ. 6945224802

Γραφείο B122 (Κτήριο βιβλιοθήκης)

ΕΥΧΑΡΙΣΤΩ