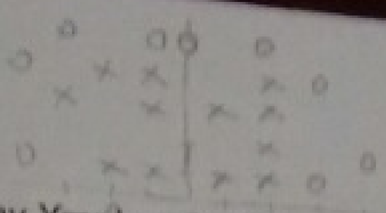
Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών & Μηχανικών Υπολογιστών
Τομέας Τεχνολογίας Πληροφορικής & Υπολογιστών



12 Φεβρουαρίου 2016

Διάρκεια: 2,5 ώρες

Εξέταση στο μάθημα ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ ΚΑΙ ΕΥΦΥΗ ΥΠΟΛΟΓΙΣΤΙΚΑ ΣΥΣΤΗΜΑΤΑ

Θέμα 1ο (30%)

α) Θέλετε να σχεδιάσετε ένα νευρωνικό δίκτυο με perceptrons το οποίο να μπορεί να διαχωρίσει τα επόμενα δεδομένα του 2-Δ χώρου στην κατηγορία στην οποία ανήκουν:

Κατηγορία 1: (1,1), (2,1), (2,2), (1,3), (2,3), (2,4), (-1,1), (-1,3), (-1,4), (-2,1), (-3,3), (-2,4)

Κατηγορία 2: (0,5), (2,5), (3,4), (4,2), (3,1), (-1,5), (-3,5), (-4,4), (-4,1), (-2,-1), (2,-1), (0,-1)

Σχεδιάστε ένα δίκτυο ορίζοντας τους νευρώνες και τα βάρη του το οποίο επιτυγχάνει την ανωτέρω κατηγοριοποίηση. Εξηγείστε.

β) Εστω ότι έχουμε ένα perceptron με δυο εισόδους και ένα νευρώνα εξόδου το οποίο έχει μάθει να υλοποιεί μια συνάρτηση f με τη βοήθεια ενός συγκεκριμένου συνόλου δεδομένων εκπαίδευσης. Είναι πάντοτε δυνατό να υλοποιήσουμε ένα άλλο perceptron, πάλι με δυο εισόδους και ένα νευρώνα εξόδου που θα μάθει να υλοποιεί τη συνάρτηση f με τη βοήθεια του ίδιου συνόλου δεδομένων εκπαίδευσης, αλλά με σταθερό κατώφλι του νευρώνα εξόδου ίσο με 1; b=1

γ) Σε ένα νευρωνικό δίκτυο με ένα κρυμμένο νευρώνα και ένα νευρώνα εξόδου με συνάρτηση ενεργοποίησης την βηματική σιγμοειδή συνάρτηση, υπολογίστε τον τύπο ανανέωσης του βάρους που συνδέει τον κρυμμένο νευρώνα με τον νευρώνα εξόδου σύμφωνα με τον αλγόριθμο backpropagation/steepest descent στις επόμενες δύο περιπτώσεις:

(i) αν η συνάρτηση κόστους είναι το τετράγωνο της διαφοράς επιθυμητής και πραγματικής εξόδου: $E = ½ (d-y)^2$

(ii) αν η συνάρτηση εξόδου είναι η επόμενη: E = -(d logy + (1-d) log (1-y)).

Μπορείτε να δικαιολογήσετε ότι η χρήση της δεύτερης συνάρτησης είναι καλύτερη από την πρώτη;

Σε ποια περίπτωση θα χρησιμοποιούσατε μια γραμμική συνάρτηση ενεργοποίησης αντί της σιγμοειδούς;

δ) Ποια είναι τα χαρακτηριστικά ενός συνελικτικού βαθιού νευρωνικού δικτύου; Αναφέρετε δυο τεχνικές κανονικοποίησης (regularization) που χρησιμοποιούνται στην υλοποίηση/μάθηση των δικτύων αυτών. Θα χρησιμοποιούσατε τον αλγόριθμο backpropagation για την εκμάθηση αυτών των δικτύων;

Θέμα 2ο (35%)

α) (15%) (i) (5%) Θεωρούμε διδιάστατο χάρτη Kohonen σε διδιάστατο χώρο εισόδου. Αν ο χάρτης είναι διατεταγμένος, θα μπορούσε η διάταξη να ανατραπεί μετά από μερικές επαναλήψεις του αλγορίθμου εκπαίδευσης;

(ii) (10%) Θεωρούμε τη διαδικασία εκπαίδευσης δικτύου SOM με 4 κόμβους στο επίπεδο εξόδου διατεταγμένους σε μονοδιάστατο δακτύλιο. Τα αντίστοιχα βάρη στον τρισδιάστατο χώρο εισόδου έχουν τρέχουσες τιμές $\mathbf{w}_1$ = [-3,1, 0,7, 5,6], $\mathbf{w}_2$ = [2,5, 4,0, 0,3], $\mathbf{w}_3$ = [1,5, -1,5, 3,4] και $\mathbf{w}_4$ = [-2,0, -1,5, 0,8]. Κατά την παρούσα φάση εκπαίδευσης του χάρτη χρησιμοποιούνται συντελεστής μάθησης η=0,3 κ

συνάρτηση γειτονιάς για κάθε κόμβο *i*: $h(i,k) = \begin{cases} 1, & i = k \\ 0{,}2, & i = k \pm 1 \\ 0, & \alpha\lambda\lambda\iota\acute{\omega}\varsigma \end{cases}$ (αρίθμηση modulo 4), όπου ο νικητής

αναδεικνύεται με βάση το κριτήριο του εσωτερικού γινομένου.

Αν παρουσιαστεί ως είσοδος το πρότυπο $\mathbf{x}$ = [-1,7, 0,3, 2,2] να γίνει η κατάλληλη ενημέρωση των βαρών σύμφωνα με τον κανόνα εκπαίδευσης του χάρτη.

β) (10%) Έστω διακριτό δίκτυο Hopfield μεγέθους N, με διπολικές τιμές στους κόμβους (συνάρτηση πρόσημο). Θεωρούμε ακολουθία M προτύπων ξ^k ($k = 1, 2, \ldots, M$), όπου το πρότυπο 1 έπεται του προτύπου M κυκλικά. Τα βάρη του δικτύου είναι μη συμμετρικά και προσδιορίζονται σύμφωνα με τη γενική σχέση:

$$w_{ij} = \frac{\lambda}{N}\sum_k \xi_i^{k+1}\xi_j^k + \frac{\mu}{N}\sum_k \xi_i^{k+2}\xi_j^k, \forall i, j.$$

Έστω ότι η τρέχουσα κατάσταση του δικτύου είναι το πρότυπο ξ^m. Υποθέτουμε ότι γίνεται ασύγχρονη ενημέρωση και ότι οι όροι διαφωνίας είναι αμελητέοι. Να δειχθεί ότι, αν $\lambda > \mu$, το δίκτυο τείνει να μεταβεί στο επόμενο πρότυπο ξ^{m+1} της ακολουθίας, ενώ, αν $\lambda < \mu$, το δίκτυο τείνει να μεταβεί στο μεθεπόμενο πρότυπο ξ^{m+2} της ακολουθίας.

γ) (10%) <u>Γενετικοί αλγόριθμοι</u>

(i) (3%) Θεωρούμε κωδικοποίηση χρωμοσωμάτων με ακολουθίες 6 δυαδικών ψηφίων. Πόσα είναι τα δυνατά σχήματα;

(ii) (7%) Ο τρέχων πληθυσμός περιλαμβάνει τα εξής χρωμοσώματα: $f = \#1$ στο χρωμόσωμα

101101, 011100, 010110, 011011, 010010 ← συνάρτηση καταλληλότητας

Αν η πιθανότητα διασταύρωσης είναι 0,7 και η πιθανότητα μετάλλαξης ανά δυαδικό ψηφίο είναι 0,001, ζητείται κάτω φράγμα για τον μέσο αριθμό στιγμιοτύπων του σχήματος s=[0****0] στην επόμενη γενιά.

<u>**Θέμα 3ο (25%)**</u>

α) (5%) Μπορείτε να δείξετε με βάση τις αντίστοιχες εξισώσεις ότι η επιλογή των διανυσμάτων υποστήριξης γενικά επηρεάζει την μορφή της καμπύλης διαχωρισμού;

β) (5%) Ποιο είναι το αποτέλεσμα της υπερπροσαρμογής (overfitting) σε ένα δίκτυο διανυσμάτων υποστήριξης (να θεωρήσετε μη-γραμμικά διαχωρίσιμο πρόβλημα).

γ) (15%) Έστω οι επόμενοι κανόνες:

(1) Αν X είναι A_1 ή X είναι A_3, τότε Y είναι B_1

(2) Αν X είναι A_2, τότε Y είναι B_2

όπου A_i και B_j ασαφή σύνολα ($i=1,2,3$, $j=1,2$) που δίνονται από τις σχέσεις:

$$A_1 = 0/x_1 + 0.2/x_2 + 1/x_3 + 0.2/x_4, \quad A_2 = 0.7/x_1 + 0.9/x_2 + 0.2/x_3 + 0/x_4$$

$$A_3 = 0/x_1 + 0/x_2 + 0.3/x_3 + 1/x_4$$

$$B_1 = 0.9/y_1 + 0.4/y_2 + 0.1/y_3, \quad B_2 = 0.3/y_1 + 0.8/y_2 + 1/y_3$$

Έστω ότι δίνεται το γεγονός A', όπου:

$$A' = 0.3/x_1 + 0.4/x_2 + 0.8/x_3 + 0.5/x_4$$

Να χρησιμοποιήσετε τη μέθοδο της παρεμβολής για να υπολογίσετε το συμπέρασμα B'.

ΚΑΛΗ ΕΠΙΤΥΧΙΑ

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών & Μηχανικών Υπολογιστών
Τομέας Τεχνολογίας Πληροφορικής & Υπολογιστών

2 Φεβρουαρίου 2010

Διάρκεια: 2,5 ώρες

Εξέταση στο μάθημα ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ ΚΑΙ ΕΥΦΥΗ ΥΠΟΛΟΓΙΣΤΙΚΑ ΣΥΣΤΗΜΑΤΑ

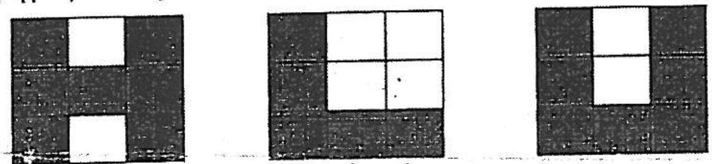
Θέμα 1ο (30%)

α) (7%) Εξ ορισμού, ένας ταξινομητής h_1 είναι περισσότερο γενικός από έναν άλλο h_2 εάν οποιοδήποτε διάνυσμα εισόδου που κατηγοριοποιείται θετικά από τον h_2 θα κατηγοροποιείται επίσης θετικά από τον h_1. Θεωρείστε δύο perceptrons που το καθένα έχει δύο εισόδους και μια έξοδο:

- perceptron A με βάρη συνδέσεων a_1, a_2 και κατώφλι a_3
- perceptron B με βάρη συνδέσεων b_1, b_2 και κατώφλι b_3.

Ποια συνθήκη πρέπει να ικανοποιούν τα a_i και b_i, i=1,2,3, ούτως ώστε το perceptron B να είναι πιο γενικό από το perceptron A;

β) (15%) Θέλουμε να χρησιμοποιήσουμε νευρωνικά δίκτυα για την ταξινόμηση των γραμμάτων H, L, U που έχουν ψηφιοποιηθεί σε ένα 3x3 πλέγμα μαύρων και άσπρων pixels, όπως φαίνεται στο Σχήμα 1. Σχεδιάστε ένα νευρωνικό δίκτυο perceptrons με τρεις εξόδους που επιλύει το πρόβλημα αυτό. Εάν παρουσιάσετε στην είσοδο του δικτύου 3 θορυβώδεις αναπαραστάσεις των γραμμάτων, στις οποίες τα pixels της πρώτης σειράς είναι όλα άσπρα, ποια θα είναι η απόδοση του κατηγοριοποιητή σας σε αυτά τα δεδομένα; Περιγράψτε πώς θα λύνατε το πρόβλημα που προκύπτει.



Σχήμα 1

γ) (8%) Ποια είναι η μαθηματική σημασία του ρυθμού μάθησης στον αλγόριθμο backpropagation και πώς θα προσαρμόζατε την τιμή του σε πραγματικά προβλήματα αναγνώρισης προτύπων;

Θέμα 2ο (13%)

α) (5%) Να δοθούν συνοπτικά οι βασικές διαφορές ανάμεσα σε ένα νευρωνικό δίκτυο πρόσθιας τροφοδότησης (όπως το πολυστρωματικό perceptron με τον αλγόριθμο backpropagation) και ένα αναδρομικό δίκτυο (όπως το δίκτυο Hopfield με τον Χεμπιανό κανόνα):

- Πώς διαφοροποιούνται κατά τη φάση της εκπαίδευσης;
- Πώς διαφοροποιούνται κατά τη φάση της λειτουργίας;

Αυτό ένα σε Ασυγχρ Hopfield

β) (8%) Θεωρούμε διακριτό δίκτυο Hopfield N κόμβων με συνάρτηση ενεργοποίησης: $y_k = f(u_k) = 1$ αν $u_k \geq 0$ και -1 αλλιώς, όπου $u_k=\Sigma_j w_{kj} y_j$. Στο δίκτυο αποθηκεύουμε M πρότυπα ξ^m σύμφωνα με τον κανόνα του Hebb:

$$w_{kj} = 1/N \, \Sigma_m \xi_k^m \xi_j^m, \quad j \neq k, \quad w_{kk} = 0.$$

Να δειχθεί ότι ένα αποθηκευμένο πρότυπο ξ^p αποτελεί ευσταθή κατάσταση του δικτύου αν ισχύει

$$|\Sigma_{j \neq k} \Sigma_{m \neq p} \xi_k^m \xi_j^m \xi_j^p| < N-1, \ \forall k.$$

Θέμα 3ο (10%)

α) (5%) Γιατί χρειάζεται η συνάρτηση γειτονιάς στον αλγόριθμο εκπαίδευσης του χάρτη Kohonen; (Π.χ. θα μπορούσε να ενημερώνεται μόνο ο νικητής κόμβος;) Γιατί το εύρος της γειτονιάς πρέπει να μειώνεται με την πάροδο του χρόνου εκπαίδευσης;

Λύση σε ασκήσεις SOM

β) **(5%)** Θεωρούμε δίκτυο ανταγωνιστικής μάθησης με 2 κόμβους εισόδου και 3 κόμβους εξόδου. Τα διανύσματα βαρών για τους κόμβους εξόδου είναι $\boldsymbol{w_1}$ = [7, 1], $\boldsymbol{w_2}$ = [1, 2] και $\boldsymbol{w_3}$ = [-3, 0]. Να εφαρμοστεί ο αλγόριθμος μάθησης για το πρότυπο εισόδου $\boldsymbol{x}$ = [4, 4] χρησιμοποιώντας (i) το κριτήριο του εσωτερικού γινομένου και (ii) το κριτήριο της ευκλείδιας απόστασης για επιλογή του νικητή, με συντελεστή μάθησης α=0,5. Να σχολιαστούν τα αποτελέσματα.

Θέμα 4ο (20%)

α) **(10%)** Να εξηγήσετε διαισθητικά το λόγο για τον οποίο η αύξηση της διάστασης του χώρου χαρακτηριστικών σε ένα δίκτυο διανυσμάτων υποστήριξης αυξάνει την πιθανότητα να μετατρέψουμε ένα μη γραμμικά διαχωρίσιμο πρόβλημα σε γραμμικά διαχωρίσιμο.

β) **(10%)** Έστω τα υπερσύνολα αναφοράς $X=\{x_1, x_2, x_3\}$, $Y=\{y_1, y_2\}$ και τα ασαφή σύνολα

$A_1=\{1/x_1, 0.9/x_2, 0.1/x_3\}$, $A_2=\{0.9/x_1, 1/x_2, 0.2/x_3\}$, $B_1=\{1/y_1, 0.2/y_2\}$ και $B_2=\{0.2/y_1, 0.9/y_2\}$.

Έστω επίσης οι κανόνες:

IF x IS A_1 THEN y IS B_1

IF x IS A_2 THEN y IS B_2

που συνδέουν τα παραπάνω ασαφή σύνολα. Δεδομένου του γεγονότος $A'=\{0.8/x_1, 0.9/x_2, 0.1/x_3\}$, να χρησιμοποιήσετε τη μέθοδο παρεμβολής για να υπολογίσετε το συμπέρασμα B'.

Θέμα 5ο (7%)

Πώς μπορεί να χρησιμοποιηθεί ένας γενετικός αλγόριθμος για την ομαδοποίηση δεδομένων; Ειδικότερα, διαθέτουμε ένα σύνολο N σημείων (δεδομένων) x_i τα οποία θέλουμε να εντάξουμε σε M ομάδες που θα περιγράφονται από τα κέντρα τους c_j(κάθε σημείο x_i ανατίθεται στην ομάδα της οποίας το κέντρο είναι πλησιέστερο προς το x_i). Να περιγράψετε συνοπτικά τα βασικά χαρακτηριστικά του γενετικού αλγορίθμου (χρωμοσώματα, συνάρτηση προσαρμογής, κλπ) για τον προσδιορισμό των κέντρων.

ΚΑΛΗ ΕΠΙΤΥΧΙΑ

ΓΡΑΠΤΟ ΜΑΡΤΙΟΥ 2014

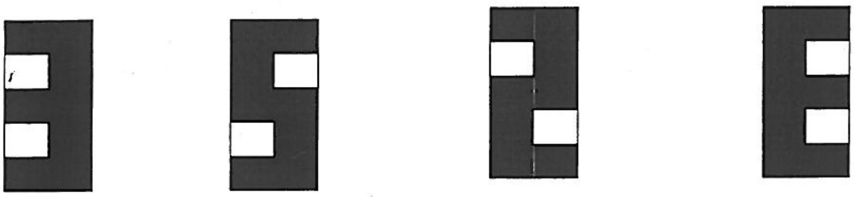
ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ ΚΑΙ ΕΥΦΥΗ ΥΠΟΛΟΓΙΣΤΙΚΑ ΣΥΣΤΗΜΑΤΑ

(Διάρκεια: 2.30 ώρες)

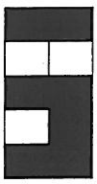
(Ακολουθούν τα θέματα όπως τα θυμάμαι, εκ μνήμης. Αν κάτι δεν βγάζει νόημα ίσως να μην το θυμάμαι καλά και να έχω γράψει μπούρδες.)

ΘΕΜΑ 1 (30%) (Κόλλιας)

1) Σχεδιάστε (συναρτήσεις ενεργοποίησης, βάρη) απλούς perceptrons που να αναγνωρίζουν τους ακόλουθους χαρακτήρες. Κάθε χαρακτήρας αποτελείται από ένα πλέγμα 5x2, και αυτά τα εικονοστοιχεία είναι οι 10 είσοδοι του δικτύου.



Τέλος σχολιάστε τι θα γίνει αν πάρει είσοδο το πλέγμα και υπολογίστε ποια θα είναι η έξοδος του δικτύου που σχεδιάσατε.



2)

(α) Τι συμβαίνει στο BackPropagation αν τα βάρη αρχικοποιηθούν σε πολύ μεγάλες τιμές οι πολύ μικρές τιμές.

(β) Απαντήστε αν τα παρακάτω είναι ορθά και πότε: Αυξάνεται η γενίκευση αν (i)μειωθούν οι εποχές (ii)αυξηθούν οι εποχές (iii)μειωθούν οι νευρώνες (iv)αυξηθούν οι νευρώνες

(γ) Έχουμε ένα νευρωνικό δίκτυο με ένα κρυμμένο επίπεδο, και με εξόδους ίσες με τις εισόδους. Που νομίζετε μπορεί να χρησιμοποιηθεί και γιατί.

<u>ΘΕΜΑ 2 (30%) (Σταφυλοπάτης)</u>

1) Κάτι με χάρτες Κοχόνεν και SOM, θεωρητικό όμως και όχι υπολογιστικό, κάτι σχετικό με τις τοπολογικές γειτονιές. (πάντως πέσανε αρκετές απορίες στο συγκεκριμένο ερώτημα, ήταν αρκετά περίεργο, ασαφές)

2) Hopfield: Διπολικό, συνάρτηση προσήμου, ασύγχρονη ενημέρωση, έχουμε αποθηκευμένα ξ^k πρότυπα (k=1,2,..,M), συνάρτηση βαρών $w_{ij} = \frac{1}{N}\sum_k \xi_i^{k-1}\xi_j^k + \frac{\mu}{N}\sum_k \xi_i^{k+1}\xi_j^k$, όροι διαφωνίας (crosstalk) μηδενικοί. Αποδείξτε ότι ξεκινώντας από το πρότυπο ξ^m, τότε αν $\mu>0$ η επόμενη κατάσταση είναι ξ^{m+1}, ενώ αν $\mu<0$ τότε ξ^{m-1}.

3) Ομοιότητες και διαφορές γενετικών αλγορίθμων και προσομοιωμένης ανόπτησης.

<u>ΘΕΜΑ 3 (20%) (Στάμου)</u>

1) Τι αντιπροσωπεύει κάθε νευρώνας του κρυμμένου στρώματος ενός δικτύου Μηχανών Διανυσμάτων Υποστήριξης (SVM);

2) Έχουμε το εξής σύστημα κλιματισμού: 3 εισόδους: θερμοκρασία T με Πεδίο Ορισμού(ΠΟ) [10,40], επιθυμητή θερμοκρασία Td με ΠΟ [18,28] και θερμοκρασία του αέρα που διοχετεύει το κλιματιστικό Tα με ΠΟ [10,60]. 1 έξοδος: ταχύτητα του ανεμιστήρα P με ΠΟ [0,1].

(α) Φτιάξε ασαφής διαμέριση

(β) Όρισε κανόνες

(γ) Υπολόγισε έξοδο του ασαφές δικτύου με είσοδο (16,55,22) με διαδικασία παρεμβολής.

ΚΑΛΗ ΕΠΙΤΥΧΙΑ

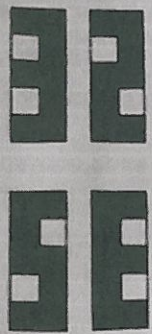
Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών & Μηχανικών Υπολογιστών
Τομέας Τεχνολογίας Πληροφορικής & Υπολογιστών

31 Αυγούστου 2015
Διάρκεια: 2,5 ώρες

Επαναληπτική Εξέταση στο μάθημα
ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ ΚΑΙ ΕΥΦΥΗ ΥΠΟΛΟΓΙΣΤΙΚΑ ΣΥΣΤΗΜΑΤΑ

Θέμα 1ο (30%)

α) (15%)Περιγράψτε καθορίζοντας και τα βάρη ένα δίκτυο από απλά perceptrons, το οποίο να μπορεί να αναγνωρίζει τους επόμενους 4 χαρακτήρες (4 έξοδοι). Οι χαρακτήρες σχεδιάζονται σε ένα πλέγμα 5 x 2 εικονοστοιχείων (10 είσοδοι) μόνο με οριζόντιες και κατακόρυφες γραμμές.



Αφού σχεδιάσετε το δίκτυο, υπολογίστε την έξοδό του, αν ως είσοδος δοθεί ο επόμενος χαρακτήρας. Σχολιάστε.

β) (15%)

1) Ποια είναι τα χαρακτηριστικά των βαθιών αρχιτεκτονικών νευρωνικών δικτύων και τι προσπαθούν να επιτύχουν αυτά;

2) Ποιά από τα επόμενα είναι σωστά και πότε;
Η ικανότητα γενίκευσης ενός πολυστρωματικού perceptron αυξάνεται
α) αν αυξήσουμε τον αριθμό των νευρώνων,
β) αν μειώσουμε τον αριθμό των νευρώνων,
γ) αν αυξήσουμε τον αριθμό των εποχών στη μάθηση,
δ) αν μειώσουμε τον αριθμό των εποχών.

3) Εχετε ένα δίκτυο με ένα μόνο κρυμμένο επίπεδο, στο οποίο ο αριθμός των νευρώνων εισόδου είναι ίσος με τον αριθμό των νευρώνων εξόδου και οι επιθυμητές έξοδοι ίδιες με τις αντίστοιχες εισόδους. Πότε και γιατί θα ήταν χρήσιμο ένα τέτοιο δίκτυο;

Θέμα 2ο (30%)

α) (7%) Έστω μονοδιάστατος χάρτης Kohonen σε μονοδιάστατο χώρο εισόδου. Υποθέτουμε ότι η ενημέρωση γίνεται με ενιαίο τρόπο για όλους τους κόμβους που περιέχονται στη γειτονιά του νικητή (συντελεστής μάθησης η). Θεωρούμε πρότυπο εισόδου x και δύο κόμβους i, j που ανήκουν στη γειτονιά του νικητή. Αν η απόσταση του βάρους w_i από το x είναι μικρότερη από την απόσταση του βάρους w_j από το x, να δειχθεί ότι αυτό ισχύει και μετά την ενημέρωση των βαρών.

β) (8%) Θεωρούμε δίκτυο ανταγωνιστικής μάθησης με 2 κόμβους εισόδου και 3 κόμβους εξόδου. Τα διανύσματα βαρών για τους κόμβους εξόδου είναι $w_1 = [7, 1]$, $w_2 = [1, 2]$ και $w_3 = [-3, 0]$. Να εφαρμοστεί ο αλγόριθμος μάθησης για το πρότυπο εισόδου $x = [4, 4]$ χρησιμοποιώντας (i) το κριτήριο του εσωτερικού γινομένου και (ii) το κριτήριο της ευκλείδιας απόστασης για επιλογή του νικητή, με συντελεστή μάθησης α=0,5. Να σχολιαστούν τα αποτελέσματα.

γ) (10%) Θεωρούμε διακριτό δίκτυο Hopfield N κόμβων με συνάρτηση ενεργοποίησης: $y_k = f(u_k) = 1$ αν $u_k \geq 0$ και -1 αλλιώς, όπου $u_k=\Sigma_j w_{kj} y_j$. Στο δίκτυο αποθηκεύουμε M πρότυπα ξ^m σύμφωνα με τον κανόνα του Hebb. Υποθέτουμε ότι τα βάρη προσδιορίζονται από τη γενική σχέση:

$w_{kj} = 1/N\ \Sigma_m \xi_k^m \xi_j^m,\ \ \forall k, j$, δηλαδή υπάρχουν και αυτοβρόχοι.

Να δειχθεί ότι ένα αποθηκευμένο πρότυπο ξ^p αποτελεί ευσταθή κατάσταση του δικτύου αν ισχύει

$|\Sigma_{j\neq k} \Sigma_{m\neq p} \xi_k^m \xi_j^m \xi_j^p| < M+N-1,\ \forall k.$

δ) (5%) Γενετικοί αλγόριθμοι: Γιατί το θεώρημα των σχημάτων παρέχει ένα κάτω φράγμα για τον αναμενόμενο αριθμό στιγμιοτύπων ενός σχήματος στην επόμενη γενιά;

Θέμα 3° (20%)

α) (10%) Έστω μία Μηχανή Διανυσμάτων Υποστήριξης η οποία διαχωρίζει σε δύο κλάσεις στο διδιάστατο επίπεδο και έστω ότι το διάνυσμα x1=[0 1] είναι ένα διάνυσμα υποστήριξης της μίας κλάσης και το x2=[1 0] είναι ένα διάνυσμα υποστήριξης της άλλης κλάσης. Να υπολογίσετε τα βάρη w και την τιμή του κατωφλίου b.

β) (10%) Έστω $U=[-10,45]$ το σύνολο των θερμοκρασιών της ατμόσφαιρας που παρατηρούνται στην Αθήνα (σε βαθμούς Κελσίου). Να οριστεί και να σχεδιαστεί μία ασαφής διαμέριση, τάξης 8, του συνόλου αυτού.

ΚΑΛΗ ΕΠΙΤΥΧΙΑ

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών & Μηχανικών Υπολογιστών
Τομέας Τεχνολογίας Πληροφορικής & Υπολογιστών

25 Φεβρουαρίου 2013

Διάρκεια: 2,5 ώρες

Εξέταση στο μάθημα ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ ΚΑΙ ΕΥΦΥΗ ΥΠΟΛΟΓΙΣΤΙΚΑ ΣΥΣΤΗΜΑΤΑ

Θέμα 1ο (30%)

α) (10%) Ένα 1-1-1 δίκτυο εκπαιδεύεται με τον αλγόριθμο backpropagation (BP). Ο νευρώνας εξόδου έχει γραμμική συνάρτηση ενεργοποίησης $f(n) = n$, όπου n είναι η είσοδός του, ενώ ο κρυμμένος νευρώνας έχει αντίστοιχα τη συνάρτηση $f(n) = 1/(1+e^{-\beta n})$. Η παράμετρος β ανήκει στις παραμέτρους που πρέπει να προσδιοριστούν με τη μάθηση. Ονοματίστε τις άγνωστες παραμέτρους του δικτύου και υπολογίστε την εξίσωση ανανέωσης με τον αλγόριθμο BP για κάθε μια από αυτές χωριστά.

β) (10%) Η έξοδος ενός δικτύου αποτελούμενου από ένα νευρώνα με δυο εισόδους p1, p2, είναι:

y = w1 * p1 + w12 * p1 * p2 + w2 * p2 + b.

i) Βρείτε ένα κανόνα μάθησης για τα w1, w12, w2, b χρησιμοποιώντας την μέθοδο της βαθύτατης καθόδου (ελαχιστοποίηση του τετραγωνικού σφάλματος μεταξύ επιθυμητών και πραγματικών εξόδων στην έξοδο του νευρώνα), όπως στον αλγόριθμο BP;

ii) Μπορεί το δίκτυο αυτό να λύσει το πρόβλημα X-OR; Δικαιολογείστε την απάντησή σας.

γ) (10%)

i) Όταν σχεδιάζουμε ένα αλγόριθμο μάθησης επιλέγουμε τον χώρο υποθέσεων. Δώστε ένα πλεονέκτημα και ένα μειονέκτημα του να επιλέξετε ένα μεγάλο ή ένα μικρό χώρο υποθέσεων.

ii) Σε ένα πολυστρωματικό perceptron η ικανότητα γενίκευσης βελτιώνεται από:
1) αύξηση αριθμού κόμβων
2) μείωση αριθμού βαρών
3) αύξηση αριθμού εποχών του αλγορίθμου μάθησης
4) μείωση του αριθμού των δεδομένων μάθησης.

Αιτιολογείστε την ορθότητα ή μη των ανωτέρω κανόνων.

Θέμα 2ο (10%)

Έστω διακριτό δίκτυο Hopfield N κόμβων, με συνάρτηση ενεργοποίησης $y_i = f(u_i) = 1$, αν $u_i \geq 0$, και -1 αλλιώς, όπου. Θεωρούμε ακολουθία M προτύπων ξ^k ($k = 1, 2, \dots, M$), όπου το πρότυπο 1 έπεται του προτύπου M κυκλικά. Τα βάρη του δικτύου προσδιορίζονται σύμφωνα με τη γενική σχέση:

$$w_{ij} = \frac{1}{N}\sum_k \xi_i^k \xi_j^k + \frac{\lambda}{N}\sum_k \xi_i^{k+1}\xi_j^k, \forall i, j,$$

δηλαδή οι συνδέσεις είναι μη συμμετρικές. Έστω ότι η τρέχουσα κατάσταση του δικτύου είναι το πρότυπο ξ^m. Να δειχθεί ότι, αν $\lambda > 1$, το δίκτυο τείνει να μεταβεί στο επόμενο πρότυπο ξ^{m+1} της ακολουθίας (ασύγχρονη ενημέρωση). Υποθέτουμε ότι οι όροι διαφωνίας είναι αμελητέοι.

Θέμα 3ο (12%)

α) (8%) Θεωρούμε δίκτυο ανταγωνιστικής μάθησης, στο οποίο η ενημέρωση των βαρών του νικητή i για το πρότυπο εισόδου $\bar{x}$ γίνεται σύμφωνα με τον κανόνα:

$$w_{ij} = w_{ij} + \eta\left(\frac{x_j}{\sum_j x_j} - w_{ij}\right), \forall j.$$

Να δειχθεί ότι ο κανόνας αυτός εξασφαλίζει την κανονικοποίηση $\sum_j w_{ij} = 1$, σε κάθε βήμα.

β) (4%) Γιατί χρειάζεται η συνάρτηση γειτονιάς στον αλγόριθμο εκπαίδευσης του χάρτη Kohonen; (Π.χ. θα μπορούσε να ενημερώνεται μόνο ο νικητής κόμβος;) Γιατί το εύρος της γειτονιάς πρέπει να μειώνεται με την πάροδο του χρόνου εκπαίδευσης;

Θέμα 4ο (20%)

α) (5%) Σε ποιες περιπτώσεις και γιατί χρησιμοποιούμε μεταβλητές χαλαρότητας στις Μηχανές Διανυσμάτων Υποστήριξης; Περιγράψτε, χρησιμοποιώντας σχηματικά παραδείγματα.

β) (15%) Να σχεδιάσετε έναν ασαφές σύσημα τριών εισόδων και μίας εξόδου, το οποίο ελέγχει την ομαλή κίνηση ενός οχήματος προς ένα συγκεκριμένο προορισμό. Με δεδομένη την ταχύτητα $V=[0,100]$, την επιτάχυνση $G=[-1,1]$ και την απόσταση από τον προορισμό $D=[0,1000]$, το σύστημα θα ελέγχει την ισχύ της μηχανής $P=[0,1]$ (πιθανά μέσω της ροής καυσίμου). Για το σκοπό αυτό:

1. Να ορίσετε τις ασαφείς διαμερίσεις στα πεδία ορισμού των εισόδων και της εξόδου (με τρεις γλωσσικούς όρους για κάθε μία από τις εισόδους).
2. Να ορίσετε τους ασαφείς κανόνες που συνδέουν την είσοδο με την έξοδο.
3. Να υπολογίσετε την έξοδο του συστήματος για είσοδο (50, 0.6, 100), με όποια μέθοδο επιθυμείτε.

Θέμα 5ο (8%)

Γενετικοί αλγόριθμοι: Θεωρούμε κωδικοποίηση χρωμοσωμάτων με δυαδικές ακολουθίες μήκους L.
i) Πόσα είναι τα δυνατά σχήματα;
ii) Μπορεί κάθε σύνολο χρωμοσωμάτων να περιγραφεί ως σχήμα;
Να αιτιολογηθούν οι απαντήσεις σας.

ΚΑΛΗ ΕΠΙΤΥΧΙΑ

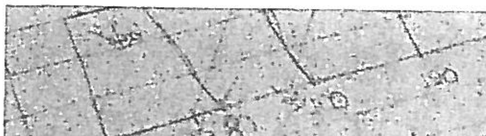
Θέμα 1ο

α)Θεωρείστε τα σύνολα δεδομένων του σχήματος 1 με συντεταγμένες (x1, x2) τα οποία ανήκουν σε 2 κατηγορίες που
συμβολίζονται με 'x' και 'o' αντίστοιχα. Υπολογίστε τα βάρη του νευρωνικού δικτύου του σχήματος το οποίο είναι σε θέση
να πετύχει την κατηγοριοποίηση των δεδομένων και στο οποίο οι συναρτήσεις ενεργοποίησης των νευρώνων είναι βηματικές
perceptron. Εξηγείστε με συντομία τις επιλογές σας.

β)Σχεδιάστε τώρα το νευρωνικό δίκτυο απλούστερα (με λιγότερα βάρη/νευρώνες) από αυτό του Σχήματος το οποίο επίσης
επιτυγχάνει επιτυχή κατηγοριοποίηση του συνόλου των δεδομένων. Εξηγείστε.

γ)Έστω ότι αντικαθιστούμε τις συναρτήσεις ενεργοποίησης με σιγμοειδείς δηλ. f(x) = 1/(1+e^x) και υπολογίζουμε τα βάρη
του δικτύου με τον αλγόριθμο backpropagation. Μια μέθοδος για αποφυγή overfitting είναι η χρήση μιας τεχνικής
κανονικοποίησης όπως η weight decay. Γράψτε την εξίσωση ανανέωσης της τιμής του w κατά τη χρήση του αλγορίθμου
backpropagation με weight decay συναρτήσει των μεταβλητών που φαίνονται στο σχήμα 1 της επιθυμητής εξόδου d της
παραμέτρου μάθησης μ και της τιμής της παραμέτρου κανονικοποίησης λ. Ποια η επίδραση της τιμής των παραμέτρων μ και λ
στην εκπαίδευση του δικτύου;

δ)Περιγράψτε ένα time-delay νευρωνικό δίκτυο και εξηγείστε πώς μαθαίνει και αναγνωρίζει ίδια patterns σε χρονικά
μεταβαλλόμενα δεδομένα/σήματα εισόδου.



πολυ μικρή γειτονιά → αποθηκευει μονο τα τοπικά χαρακτηριστικά της εισοδου και οχι τη γενική μορφή του πληγματος

(2015)

Θέμα 2ο (30%)

α) (13%) i) (5%) Ποιές συνέπειες θα μπορούσε να έχει κατά την εκπαίδευση ενός χάρτη Kohonen η χρήση πολύ μικρής γειτονιάς καθ' όλη τη διάρκεια της εκπαίδευσης. Να αιτιολογηθεί η απάντηση σας.

ii) (8%) Θεωρούμε τη διαδικασία εκπαίδευσης δικτύου SOM με 4 κόμβους στο επίπεδο εξόδου διατεταγμένους σε μονοδιάστατο δακτύλιο. Τα αντίστοιχα βάρη στον τρισδιάστατο χώρο εισόδου έχουν τρέχουσες τιμές w_1 = [-1,0, -1,5, 0,5], w_2 = [2,0, -2,0, 5,2], w_3 = [1,5, 6,0, 4,3] και w_4 = [-4,0, 7,0, 0,6]. Κατά την παρούσα φάση εκπαίδευσης του χάρτη χρησιμοποιούνται συντελεστης μάθησης η=0,5 και συνάρτηση γειτονιάς για κάθε κόμβο k: $h(i,k) = \begin{cases} 1, & i = k \\ 0{,}4, & i = k \pm 1 \\ 0, & \text{αλλιώς} \end{cases}$ (αρίθμηση modulo 4). Ο νικητής αναδεικνύεται με βάση την Ευκλείδεια απόσταση.

Αν παρουσιαστεί ως είσοδος το πρότυπο x = [-1,4, 2,3, 0,2] να γίνει η κατάλληλη ενημέρωση των βαρών σύμφωνα με τον κανόνα εκπαίδευσης του χάρτη.

β) (10%) Ένα διακριτό δίκτυο Hopfield πρόκειται να χρησιμοποιηθεί ως αυτοσυσχετιστική μνήμη. Δίνονται τα πρότυπα , τα οποία αντιστοιχίζονται με διάταξη κατά γραμμές

x_1 x_2 x_3
x_4 x_5 x_6

σε μονοδιάστατα διπολικά διανύσματα εισόδου του δικτύου (λευκό=-1, μαύρο=1)

Με χρήση των δύο πρώτων προτύπων για την εκπαίδευση του δικτύου να κατασκευαστεί ο πίνακας βαρών σύμφωνα με τον Χεμπιανό κανόνα και να ελεγχθεί αν τα δύο αυτά πρότυπα αποτελούν ευσταθείς καταστάσεις.

Λαμβάνοντας ως αρχική κατάσταση το τρίτο πρότυπο να υλοποιηθεί ακολουθία σύγχρονων ενημερώσεων. Ποιά θα είναι η τελική κατάσταση του δικτύου;

γ) (7%) Γενετικοί αλγόριθμοι: Θεωρούμε κωδικοποίηση χρωμοσωμάτων με δυαδικές ακολουθίες μήκους L.

i) (3%) Πόσα είναι τα δυνατά σχήματα;

ii) (4%) Η *τάξη* ενός σχήματος είναι το πλήθος των *ορισμένων* (0 ή 1) ψηφίων του. Πόσα είναι τ δυνατά σχήματα τάξης k,

Να αιτιολογηθούν οι απαντήσεις σας.

σελ. 278

Θέμα 3ο (20%)

α) (5%) Να εξηγήσετε με συντομία με ποιο τρόπο οι μεταβλητές χαλάρωσης επηρεάζουν την ευρωσ του περιθωρίου ταξινόμησης σε θορυβώδη δεδομένα.

β) (15%) Να σχεδιάσετε ένα ασαφές σύστημα αναπαράστασης της αρτηριακής πίεσης των ανθρώπ το οποίο πρόκειται να χρησιμοποιηθεί στο σχεδιασμό ενός συστήματος ιατρικής διάγνω (υπενθυμίζουμε ότι η μέτρηση της αρτηριακής πίεσης γίνεται με δύο αριθμούς που εκφράζου συστολική και διαστολική πίεση, πχ. 120 με 80 mmHg). Να εξηγήσετε τη λειτουργία του κα υπολογίσετε την έξοδό του σε μετρήσεις 120 με 80 mmHg, 170 με 80 mmHg και 80 με 50 mmHg. θα ήταν σημαντικό να μπορεί να προσαρμόζεται το σύστημα και με ποιο τρόπο θα μπορούσε να γί αυτό;

ΚΑΛΗ ΕΠΙΤΥΧΙΑ

Θέμα 1-2014

2) Αν αρχικοποιηθούν σε μικρό μικρές → αργή σύγκλιση
" σε μεγάλο μεγάλες → κορεσμό στις συναρτήσεις εξόδου άρα δεν διαδίδεται προς τα πίσω

3) i) ναι
ii) όχι
iii) ναι
iv) όχι

γ) κωδικοποίηση εισόδου (σελ 181-183)

Θέμα 2-2014

2) Ίδιο με 2013

3) Γενετικοί αλγόριθμοι vs simulated annealing (google search)

Θέμα 3-2014

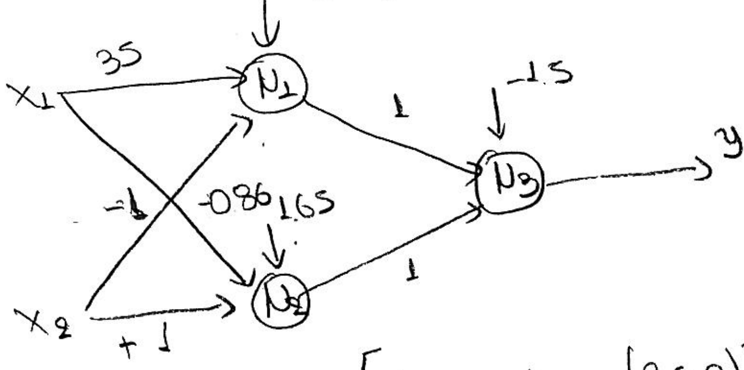
1) Κάθε κρυφός νευρώνας πέρα από αναπαράσταση σε μεγαλύτερο χώρο διαστατικότητας προσπαθεί να μάθει ένα κρυφό feature της εισόδου.

Θέμα 1-2015

1 2 ευθείες $(2.4, 0.5) \to (2.5, 4) \rightsquigarrow y = 35x + 83.5$ άρα $b = +83.5$, $w_1 = -35$, $w_2 = 1$
$(2.5, 0.5) \to (6, 3.5) \rightsquigarrow y = 0.86x + 1.65$ άρα $b = +1.65$, $w_1 = 0.86$, $w_2 = 1$

Μετά βρίσκω το θετικό και το αρνητικό για να αλλάξω τα πρόσημα

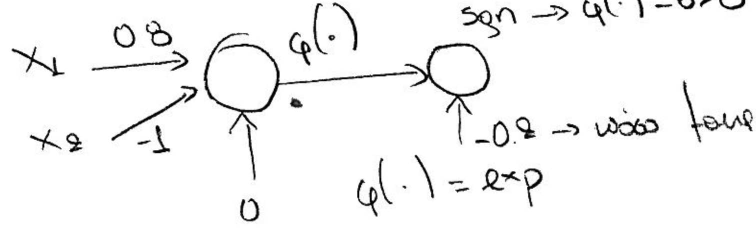
Άρα w_1 αντίστροφα, w_2 όπως είναι



3) Παίρνω κέντρο γκαουσιανής [κύκλος κέντρου (2.5, 2)] $\exp^{-\|0.8x_1 - x_2\|^2}$ (RBF)

$a \cdot 2.5 - b \cdot 2 = 0 \rightsquigarrow a = \frac{2b}{2.5} = 0.8b$

Άρα



-0.2 → πόσο μακριά από το κέντρο είναι

$\varphi(\cdot) = \exp$

$v_1 = w_{11}x_1 + w_{12}x_2 + w_{10}$

$y = \frac{1}{1+e^{-x}}$

$v_2 = w_{21}x_1 + w_{22}x_2 + w_{20}$

$v_3 = v_1 w_1 + v_2 w_2 + w_3$

δ) Time delay → Διαφάνειες κεφ 3

Θέμα 3 – 2015

α) Μεταβλητές ρυθμισης (σελ 278)

Θέμα 2 – 2015

α) i) Πολύ μικρή γειτονιά → αισθητήρες μόνο τα τοπικά χαρακτηριστικά της εισόδου και όχι τη γενική μορφή του σήματος

ii) $n = 0{,}5$

$x = [-1.4,\ 2.3,\ 0.2]$, $w_1 = [-1.0,\ -1.5,\ 0.5]$, $w_2 = [2.0,\ -2.0,\ 5.2]$

$w_3 = [1.5,\ 6.0,\ 4.3]$, $w_4 = [-4.0,\ 7.0,\ 0.6]$

$\|x - w_1\| = \left((-1.4+1.0)^2 + (2.3+1.5)^2 + (0.2-0.5)^2\right) = 0.4^2 + 3.8^2 + 0.3^2 = 14.69$ ↑ νικητής

$\|x - w_2\| = 3.4^2 + 4.3^2 + 5^2 = 55.05$

$\|x - w_3\| = 2.9^2 + 3.7^2 + 4.1^2 = 38.91$

$\|x - w_4\| = 2.6^2 + 4.7^2 + 0.4^2 = 29.01$

γειτονιά w_1, w_2, w_4

$w_1' = w_1 + n\,h\,(x - w_1) = [-1.0,\ -1.5,\ 0.5] + 0.5 \cdot 1 [-0.4,\ 3.8,\ -0.3] = [-1.2,\ 0.4,\ 0.35]$

$w_2' = [2.0,\ -2.0,\ 5.2] + 0.5 \cdot 0.4 [-3.4,\ 4.3,\ -5] = [1.32,\ -1.14,\ 4.2]$

$w_3' = [1.5,\ 6.0,\ 4.3] + 0.5 \cdot 0.4 [2.6,\ -4.7,\ -0.4] = [-3.48,\ 6.06,\ 0.52]$

β) 1ο πρότυπο: $[-1\ 1\ -1\ 1\ -1\ -1]$

2ο πρότυπο $[-1\ 1\ -1\ -1\ 1\ -1]$

3ο πρότυπο $[-1\ -1\ 1\ -1\ 1\ -1]$

$$W = \frac{1}{6}\begin{bmatrix}-1\\1\\-1\\1\\-1\\-1\end{bmatrix}\begin{bmatrix}-1 & 1 & -1 & 1 & -1 & -1\end{bmatrix} + \frac{1}{6}\begin{bmatrix}-1\\1\\-1\\-1\\1\\-1\end{bmatrix}\begin{bmatrix}-1 & 1 & -1 & -1 & 1 & -1\end{bmatrix}$$

$$\frac{2}{6}\begin{bmatrix}1&0&0&0&0&0\\0&1&0&0&0&0\\0&0&1&0&0&0\\0&0&0&1&0&0\\0&0&0&0&1&0\\0&0&0&0&0&1\end{bmatrix}=\begin{bmatrix}0&-1/3&1/3&0&0&1/3\\-1/3&0&-1/3&0&0&-1/3\\1/3&-1/3&0&0&0&1/3\\0&0&0&0&-1/3&0\\0&0&0&-1/3&0&0\\1/3&-1/3&1/3&0&0&0\end{bmatrix}$$

ια την ακολουθία σύγχρονων εμφανίσεων αυτό κάτω το μέγιστο

i) Αν έχω $0,1,* \to 3^L$

$0,1 \to 2^L$

ii) Αν έχω $0,1,* \to 3^{L-k}$

$0,1 \to 2^{L-k} \to 2^k\binom{L}{k}$

Genetic algorithms (GA) vs simulated annealing (SA)

SA → μιλάει για λύσεις, κόστος, γείτονες και κινήσεις ενώ σε GA → μιλάει για χρωμοσώματα, καταλληλότητα, επιλογή, μίξη και μετάλλαξη

- SA ίδιο με GA με πληθυσμό ίσο με 1, άρα δεν υπάρχει μίξη αλλά μόνο μετάλλαξη
- SA δημιουργεί νέα λύση αλλάζοντας ήδη υπάρχουσα, ενώ GA συνδυάζει 2 υπάρχουσες
- SA και GA υποθέτουν ότι η λύση βρίσκεται κοντά σε γνωστές
- GA θεωρεί ότι τα παιδιά μοιράζονται ιδιότητες των γονιών, άρα 2 καλοί γονείς θα έχουν καλά παιδιά

Time delay νευρωνικό

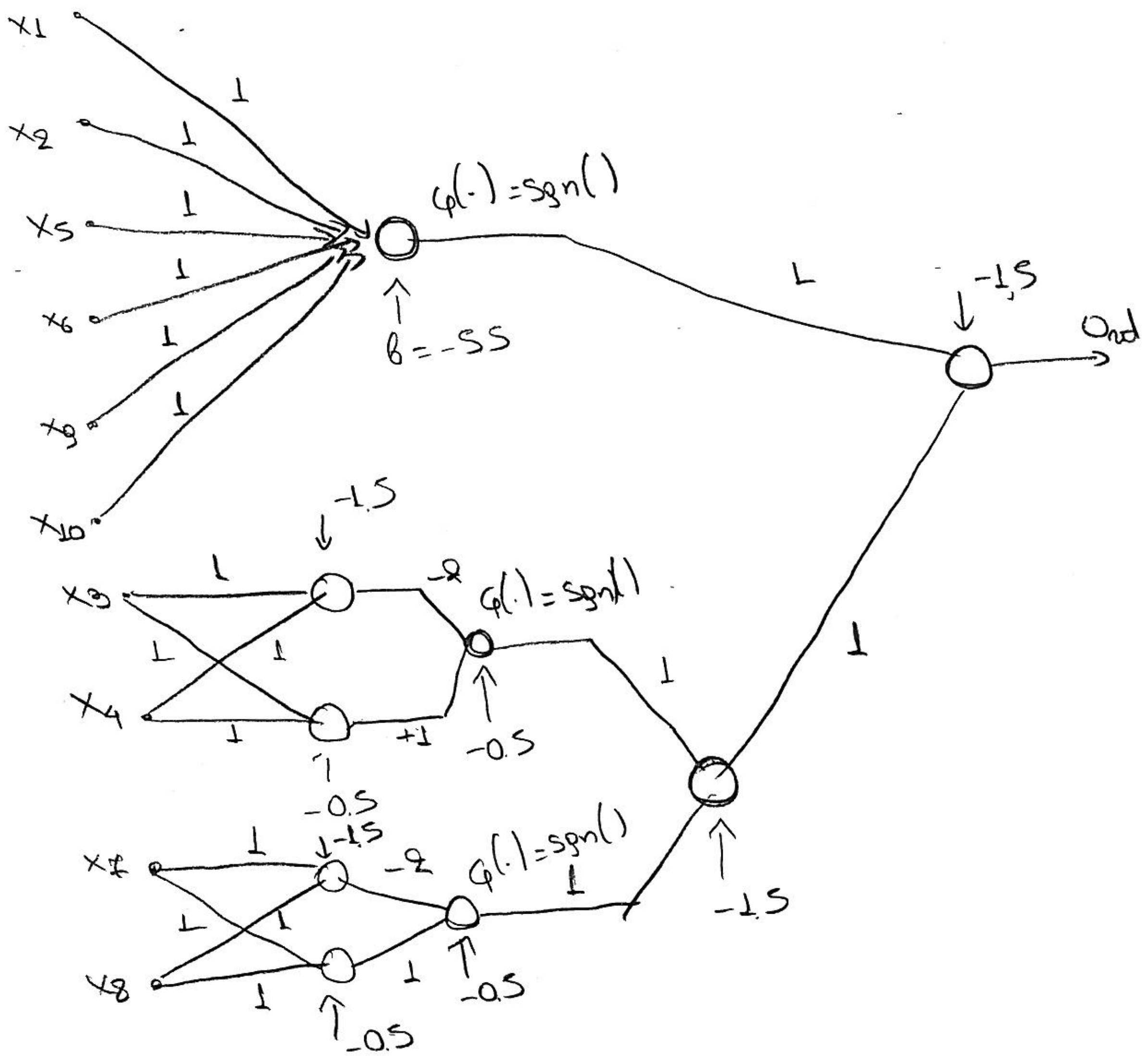
Η είσοδος x_i κατά τη χρονική στιγμή k ταυτίζεται με την είσοδο x_{i+1} κατά τη χρονική στιγμή k+1. Συνάρτηση που υλοποιείται από το δίκτυο $y = F(s(k), s(k-1), \dots, s(k-n+1))$ αντίγραφα του σήματος με γραμμικό φίλτρο, άρα τρέχω backpropagation για shifted αντίγραφα του σήματος

Αρτηριακή πίεση

Κατηγορία	Συστολική mm Hg		Διαστολική mm Hg
Φυσιολογική πίεση	κάτω από 120	και	κάτω από 80
Προ-υπέρταση	120-139	ή	80-89
Υπέρταση στάδιο 1	140-159	ή	90-99
Υπέρταση στάδιο 2	160-179	ή	100-109
Υπερτασική κρίση	180<	ή	110<

1) Περνάω την 1η, 3η, 5η γραμμή (ίδιες και έξω)

x_3 x_4
x_7 x_8 και πρέπει να κάνω AND & XOR



) $\eta \to$ συντελεστής "βάρους" η προσαρμογής της παραγώγου

. παραμετρικά για προβλήματα → όρος ορμής, προσαρμογή ανάλογα με τον αριθμό των εισόδων,
μεταβλητός ανάλογα με το επίπεδο που βρίσκεται,
προσαρμοστικό έλεγχο ρυθμού μάθησης με βέλτιστη απόσταση
και ανάλογα με το χρόνο

Θέμα 2

perceptron ≠ Hopfield

→ perceptron διαχωρίζει ενώ το Hopfield να τα αποθηκεύει

→ perceptron προσαρμόζει βάρη, Hopfield ελαχιστοποιεί ενέργεια προσεγγιστικά } φάση εκπαίδευσης

→ perceptron ρυθμό μάθησης

→ Hopfield → ταξινόμηση αν δεν το έχει αποθηκευμένο
↳ θα σου επιστρέψει ποιο μοιάζει πιο πολύ απ' αυτά } φάση λειτουργίας

→ perceptron → ταξινόμηση

β) (Λυμένο σε ασκήσεις Hopfield)

$w_{ij} = \frac{1}{N}\sum_m \xi_i^m \xi_j^m \,,\ j \neq i,\ w_{ii}=0$, αν έχω $y = \xi^p \to f(u) = f(\dots)$

$u_i = \sum_j w_{ij}\xi_j^p = \frac{1}{N}\sum_{j\neq i}\sum_m \xi_i^m \xi_j^m \cdot \xi_j^p = \frac{1}{N}\sum_{j\neq i}\left(\sum_{m\neq p}\xi_i^m \xi_j^m \cdot \xi_j^p + \xi_i^p(\xi_j^p)^2\right)$

$= \frac{1}{N}\left[\sum_{j\neq i}\xi_i^p + \sum_{j\neq i}\sum_{m\neq p}\xi_i^m \xi_j^m \xi_j^p\right] = \frac{1}{N}\left[(N-1)\xi_i^p + \underbrace{\sum_{j\neq i}\sum_{m\neq p}\xi_i^m \xi_j^m \cdot \xi_j^p}_{q_i}\right]$

αν $\xi_i^p = 1 \to \frac{1}{N}\left[(N-1)+q_i\right]$ είναι ≥ 0 αν $q_i \leq 1-N$ (1)

$\xi_i^p = -1 \to \frac{1}{N}\left(q_i-(N-1)\right)$ είναι ≥ 0 αν $q_i-(N-1) \geq 0 \to q_i \geq N-1$ (2)

Άρα από (1), (2) $\to |q_i| < N-1$

Θέμα 3ο

α) Συνάρτηση γειτονιάς → για να προσαρμόζεται το SOM σε αυτά που του παρουσιάζεις
πολύ μικρή → σύγκλιση σε τοπικό επίπεδο προσαρμόζεται μόνο κάποιο κομμάτι σε
τοπικό επίπεδο και όχι όλο, άρα τη μειώνεις για να συγκλίνει
όλο το πλέγμα και στη συνέχεια τα τοπικά χαρακτηριστικά

) $w_1 = [7\ \ 1]$, $w_2 = [1\ \ 2]$, $w_3 = [-3\ \ 0]$ $x = [4, 4]$

i) $w_1 x = 32 \leftarrow$ νικητής

$w_2 x = 12$

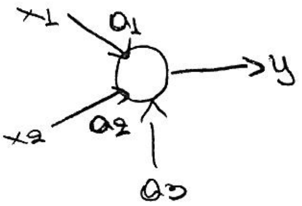
$w_3 x = -12$

$w_1' = w_1 + a(x - w_1) = [7\ \ 1] + \frac{1}{2}[-3\ \ 3]$

$= [5.5\ \ 2.5]$

α)

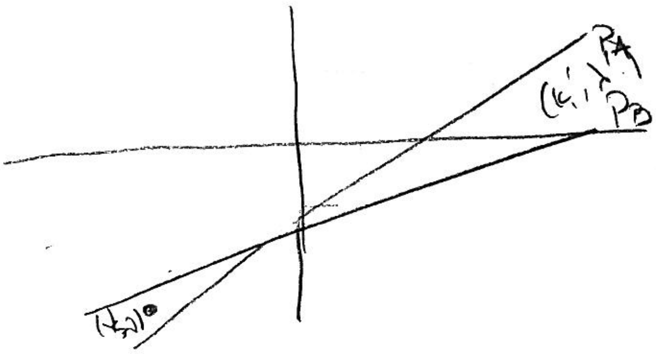
Perceptron A



$$x \cdot w > 0 \xrightarrow{P_A} [x_1\ x_2\ 1]\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} > 0 \rightsquigarrow x_1 a_1 + x_2 a_2 + a_3 > 0$$

$$\xrightarrow{P_B} [x_1\ x_2\ 1]\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} > 0 \rightsquigarrow b_1 x_1 + b_2 x_2 + b_3 > 0$$

$$\left.\begin{aligned} b_1 &= a_1 \cdot \lambda \\ b_2 &= a_2 \cdot \lambda \\ b_3 &< a_3 \end{aligned}\right\} \text{Πρέπει οι ευθείες να είναι παράλληλες}$$

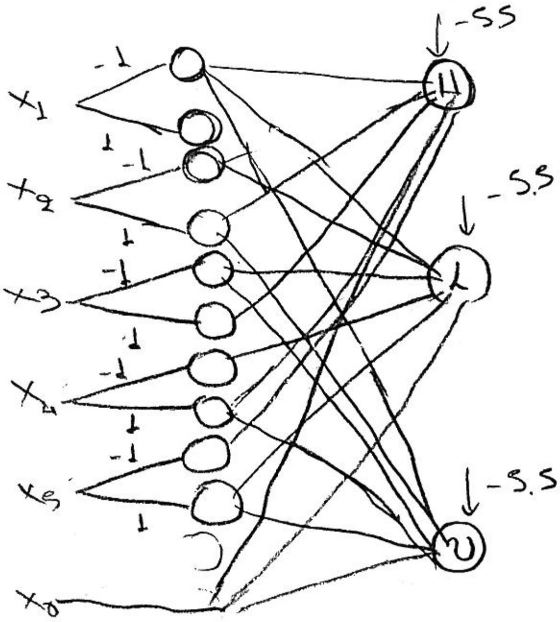


β) $\begin{pmatrix} x_1 & x_2 \\ x_3 & x_4 \\ x_5 & x_6 \end{pmatrix}$

$H \rightsquigarrow x_1'$ and x_2 and x_3 and x_4 and x_5' and x_6

$L \rightsquigarrow x_1'$ and x_2' and x_3' and x_4' and x_5 and x_6

$U \rightsquigarrow x_1'$ and x_2 and x_3' and x_4 and x_5 and x_6



Αν η τελευταία σειρά είναι όλη άσπρη θα αναγνωρίζει μόνο το L. Για να αντιμετωπιστεί το πρόβλημα είτε θα εφαρμόζεις εκπαίδευση με πρότυπα με θόρυβο είτε θα έδινες μικρότερο βάρος στην τελευταία σειρά

$\|w_1 - x\| = 3^2 + 3^2 = 18$

$\|w_2 - x\| = 3^2 + 2^2 = 13 \leftarrow$ νικητής

$\|w_3 - x\| = 7^2 + 4^2 = 65$

$w_2' = w_2 + a(x - w_2) =$

$= [1 \quad 2] + \frac{1}{2}[3 \quad 2]$

$= [2.5 \quad 3]$

Θέμα 4ο - 2010

αύξηση χώρου διαστατικότητας → αυτόματη [illegible] (σελ 233, θεώρημα Cover)

$A_1 = \{1/x_1, 0.9/x_2, 0.1/x_3\}$ $B_1 = \{1/y_1, 0.2/y_2\}$

$A_2 = \{0.9/x_1, 1/x_2, 0.2/x_3\}$ $B_2 = \{0.2/y_1, 0.9/y_2\}$

$A' = \{0.8/x_1, 0.9/x_2, 0.1/x_3\}$

$$r_i(A') = \sup_{j \in N_m} t[A'(x_j), A_i(x_j)]$$

$A_1 : \sup\{0.8/x_1, 0.9/x_2, 0.1/x_3\} = 0.9/x_2 \rightarrow y_1$

$A_2 : \sup\{0.8/x_1, 0.9/x_2, 0.1/x_3\} = 0.9/x_2 \rightarrow y_2$

$B'(y_1) = \sup t[r_i(A'), B_j(x_j)] = \sup[t(0.9, 1, (0.9, 0.2)] = \{0.9/y_1, 0.2/y_2\} = \{0.9/y_1\}$ $B' = \{0.9/y_1, 0.9/y_2\}$

$B'(y_2) = (\{0.9, 0.2\}, \{0.9, 0.2\} = \{0.2/y_1, 0.9/y_2\} = \{0.9/y_2\}$

Θέμα 5 - 2010

γενικευμένος αλγόριθμος για clustering

→ αντιστοιχίζεις κάθε κέντρο c_j σε 1 διάνυσμα (πρωτότυπο)

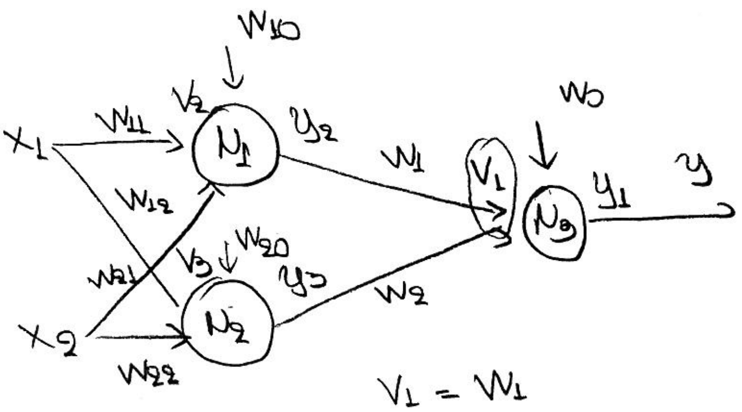
Τεχνική: χωρίζεις σε clusters με βάση τα πρωτότυπα $k\{$

και τοποθετείς κάθε πρωτότυπο να φτάσει στο κέντρο

αφού αρχικοποιήσεις τον αλγόριθμο

γ) $\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}(k-1)} - \alpha w_{ij}(k-1)$

$f(x) = \frac{1}{1+e^{-x}}$



$v_1 = w_1$

Επίπεδο εξόδου

$$\Delta w_1 = -\eta \frac{\partial E}{\partial w_1} - \alpha w_1 = -\eta \frac{\partial E(n)}{\partial e_1(n)} \cdot \frac{\partial e_1}{\partial y_1} \cdot \frac{\partial y_1}{\partial v_1} \cdot \frac{\partial v_1}{\partial w_1} - \alpha w_1$$

$$E(n) = \frac{1}{2} e_1^2(n) \rightsquigarrow \frac{\partial E(n)}{\partial e_1} = e_1$$

$$\frac{\partial e_1}{\partial y_1} = \frac{\partial}{\partial y_1}(d_1 - y_1) = -1$$

$$\frac{\partial y_1}{\partial v_1} = f(v_1)(1 - f(v_1))$$

$$\frac{\partial v_1}{\partial w_1} = y_2$$

$\Downarrow$

$$\Delta w_1 = -\eta\left(e_1(-1)(f(v_1)(1-f(v_1))y_2\right) - \alpha w_1$$

$$\Delta w_2 = -\eta\left(e_1(-1)(f(v_1)(1-f(v_1))y_3\right) - \alpha w_2$$

$$\Delta w_{11} = -\eta \frac{\partial E}{\partial w_{11}} - \alpha w_{11}$$

$$= -\eta \; f(v_2)(1-f(v_2)) \cdot \left[e_1(-1)\,(f(v_1)(1-f(v_1))w_1\right] - \alpha w_{11}$$

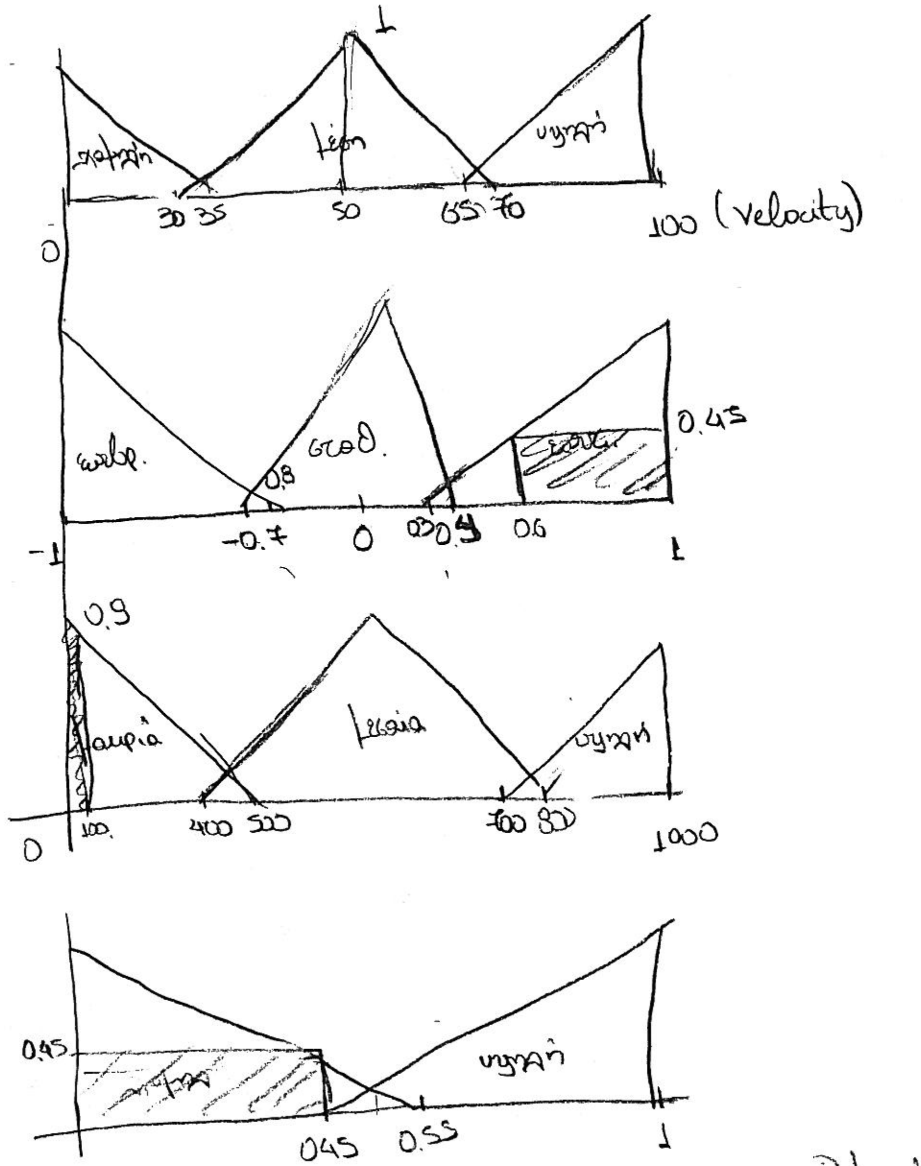
ι) Θέματα 2013

Θέμα 4 – 2013

2) Μεταβατικές συμπεριφορές → όταν θες να ταξινομήσεις μη διαχωρίσιμα πρότυπα
↳ όταν είναι γραμμικά διαχωρίσιμα και θες να βρεις διαχωριστική επιφάνεια

Σημασιολογικά παραδείγματα → σελ 277

3) $V = [0, 100]$
$G = [-1, 1]$
$D = [0, 1000]$
$P = [0, 1]$



if V = μέση AND D χαμηλή AND G υψηλή → P μικρό

AND(min) = 0.45

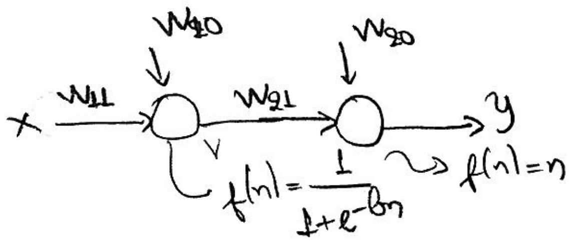
$\boxed{\text{Να ορίσουμε κανόνες}}$

Θέμα 5_2013

i) 3^L

ii) Αναγνωρίσιμο δείγμα: $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ ⇝ δεν μπορεί να το κάνει σωστά

1)



$$y_1(n) = f(w_{11} x + w_{10}) = \frac{1}{1+e^{-b v_1}}$$

$$y_2(n) = f(w_{21} y_1 + w_{20}) = v_2(n)$$

$$v_2(n) = w_{21} y_1(n) + w_{20}$$

ΘΑ 130-134

$$\mathcal{E} = \frac{1}{2} e_2^2(n) \rightsquigarrow \frac{\partial E}{\partial e_2} = e_2(n)$$

$$e_2(n) = d_2(n) - y_2(n) \rightsquigarrow \frac{\partial e_2}{\partial y_2} = -1$$

$$\frac{\partial E}{\partial w_{20}} = \underset{e_2}{\frac{\partial E}{\partial e_2}} \cdot \underset{-1}{\frac{\partial e_2}{\partial y_2}} \cdot \underset{1}{\frac{\partial y_2}{\partial v_2}} \cdot \frac{\partial v_2}{\partial w_{20}} = -e_2$$

$$\frac{\partial E}{\partial w_{21}} = \frac{\partial E}{\partial e_2} \cdot \frac{\partial e_2}{\partial y_2} \cdot \frac{\partial y_2}{\partial v_2} \cdot \frac{\partial v_2}{\partial w_{21}} = -e_2 y_1$$

$$\frac{\partial E}{\partial b} = \frac{\partial E}{\partial e_2} \cdot \frac{\partial e_2}{\partial y_2} \cdot \frac{\partial y_2}{\partial v_2} \cdot \frac{\partial v_2}{\partial b} = -e_2 \frac{\partial v_2}{\partial b}$$

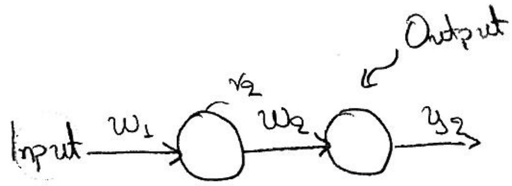
$$\frac{\partial v_2}{\partial b} = \frac{\partial}{\partial b}\left(w_{21} \frac{1}{1+e^{-b v_1(n)}} + w_{20} \right) = + w_{21} y_1(n) \cdot e^{-b v_1(n)}$$

άρα $b(n+1) = b(n) + \eta \left(-e_2 w_{21} v_1(n) y_1(n) e^{-b v_1(n)} \right)$

παραγωγικών ...

Θέμα 1

1-1-1 δίκτυο
επίπεδο εξόδου $\to f(n)=n$
κρυφό επίπεδο $\to f(n) = 1/(1+e^{-bn})$
$\to$ παράμετρος που προσδιορίζεται από μάθηση
άγνωστες παράμετροι του δικτύου $\to$?
εξίσωση ανανέωσης με τον BP για καθεμία από αυτές



Hidden Layer

Άγνωστες παράμετροι w_1, w_2, b

Είναι $e_k = d_k - y_k$, $d_k \to$ επιθυμητή έξοδος
$y_k \to$ έξοδος perceptron
$e_k \to$ σφάλμα

$\delta_k = e_k \cdot f'(v_k)$, $f \to$ συνάρτηση ενεργοποίησης
$v_k \to$ έξοδος νευρώνα

Άρα για επίπεδο εξόδου $\delta_2 = e_k \cdot 1 = e_2$

και $\Delta w_2 = \gamma \cdot \delta_2 \cdot y$, γ ρυθμός μάθησης

και $w_2' = w_2 + \Delta w_2$

Για κρυφό επίπεδο: $\delta_1 = f'(v_1) \cdot \delta_2 \cdot w_2'$

$$= + \frac{1}{(1+e^{-bn})^2} \cdot b e^{-bn} \cdot \delta_2 \cdot w_2'$$

και $\Delta w_1 = \gamma \cdot \delta_1 \cdot x$, $x \to$ είσοδος

και $w_1' = w_1 + \Delta w_1$

εναλλακτικά με κανόνα αλυσίδας (σελ 130)

$$\frac{\partial E}{\partial b} = \frac{\partial E}{\partial e} \cdot \frac{\partial e}{\partial y_2} \cdot \frac{\partial y_2}{\partial v_2} \cdot \frac{\partial v_2}{\partial b}$$

$$\overset{\shortparallel}{e} \qquad -1 \qquad 1$$

$$E = \frac{1}{2} e^2$$

$$y_2 = f(v_2) = v_2(n)$$

$$\frac{\partial v_2}{\partial b} = \frac{\partial\left(w_2 \cdot \frac{1}{1+e^{-bv_1(n)}}\right)}{\partial b} = \underbrace{\left(-\frac{w_2}{1+e^{-bv_1(n)}}\right)}_{}^{-w_2 y_1} \cdot \frac{\partial\left(1+e^{-bv_1(n)}\right)}{\partial b} =$$

$$= -w_2 \cdot y_1(n) \cdot (-v_1) e^{-bv_1} = w_2 \cdot v_1(n)\, y_1(n)\, e^{-bv_1(n)}$$

$$b(n+1) = b(n) + n \frac{\partial v_2}{\partial b} = b(n) - n e_2 w_2 v_1(n)\, y_1(n)\, e^{-bv_1}$$

$$\frac{\partial E}{\partial w_1} = \frac{1}{2}\frac{\partial}{\partial w_1}(e_j(n))^2 = \frac{1}{2}\frac{\partial}{\partial w_1}(\ldots$$

$$= \frac{1}{2}\,2\,e(-p_1) = -e\cdot p_1$$

$$\frac{\partial E}{\partial w_2} = \frac{1}{2}\cdot 2\cdot e\cdot p_2 = -e\cdot p_2$$

$$\frac{\partial E}{\partial w_{12}} = \frac{1}{2}\cdot 2\cdot e(-p_1p_2) = -e\,p_1p_2$$

$$\frac{\partial E}{\partial b} = +e(-1) = -e$$

ii) Για $(0,0)\to 0$, $(1,0)\to 1$, $(0,1)\to 1$, $(1,1)\to 0$

$$\begin{cases} 0 = b \\ 1 = w1 \\ 1 = w2 \\ 0 = 1\cdot 1 + w12\cdot 1\cdot 1 + b\cdot 1 \rightsquigarrow w12 = -2 \end{cases}$$

γ) i) Hypothesis space → ο χώρος που ψάχνεις συναρτήσεις για fit
μεγαλύτερος → μπορείς να αναπαραστήσεις περισσότερα δεδομένα
μικρότερος → ευκολότερο να ψάξεις

ii) 1) όχι → overfitting
2) ναι → αυξάνεται δυνατότητα
3) όχι
4) ναι

Θέμα 2

Λυμένο σε ασκήσεις Hopfield (Άσκηση 9)

Θέμα 3

α) (Άσκηση 2)

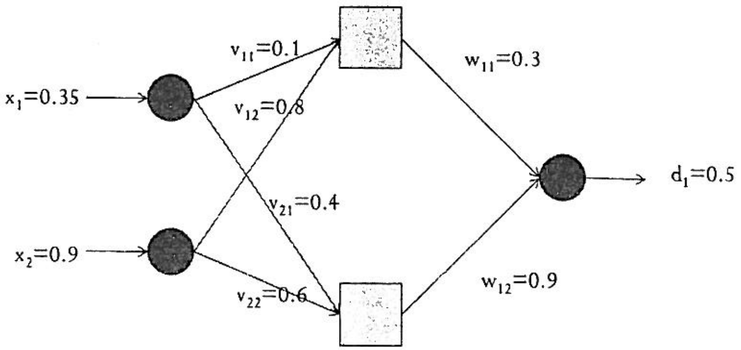
$$w'_{ij} = w_{ij} + n\left(\frac{x_j}{\sum_j x_j} - w_{ij}\right) \forall j$$

$$\sum_j w'_{ij} = \sum_j\left(w_{ij} + n\left(\frac{x}{\sum_j x_j} - w_{ij}\right)\right) = \sum_j w_{ij} + n\sum_j \frac{x_j}{\sum_j x_j} - \sum_j w_{ij}\,) =$$

$$= 1 + n(1-1) = 0 *$$

Έστω $\sum_j x_j = p$ άρα $\sum_j \frac{x_j}{p} = \frac{1}{p}\sum_j x_j = \frac{1}{p} = 1$

Παράδειγμα

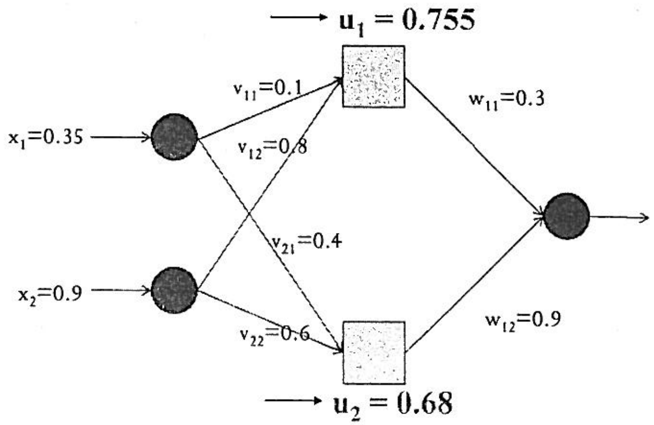


Είσοδος: $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0.35 \\ 0.9 \end{pmatrix}$ Επιθυμητή Έξοδος: $(d_1) = (0.5)$

21

Παράδειγμα

$$u_k = \sum_i w_{ki} \cdot x_i$$

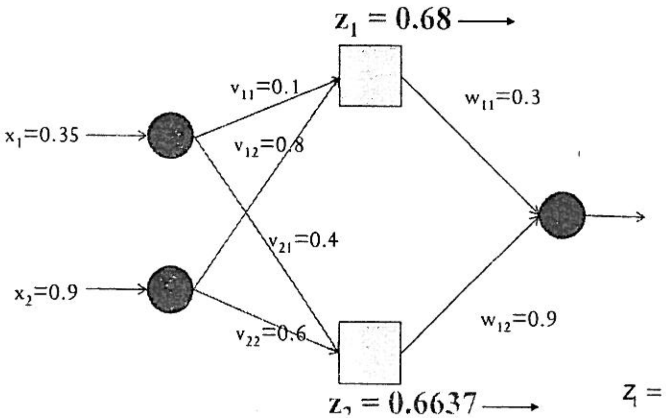


- Πέρασμα κατά την ΕΥΘΕΙΑ φορα $u_1 = v_{11} \cdot x_1 + v_{12} \cdot x_2 = 0.1 \cdot 0.35 + 0.8 \cdot 0.9 = 0.755$
- Υπολογίζουμε τις ενεργοποιήσεις 1^{ου} επιπέδου $u_2 = v_{21} \cdot x_1 + v_{22} \cdot x_2 = 0.4 \cdot 0.35 + 0.6 \cdot 0.9 = 0.68$

22

Παράδειγμα

$$y_k = f(u_k) = \frac{1}{1 + e^{-u_k}}$$



- Υπολογίζουμε τις εξόδους του πρώτου επιπέδου, περνώντας τις ενεργοποιήσεις που υπολογίσαμε προηγουμένως στη συνάρτηση ενεργοποίησης:

$$z_1 = \frac{1}{1 + e^{-0.755}} = 0.68$$

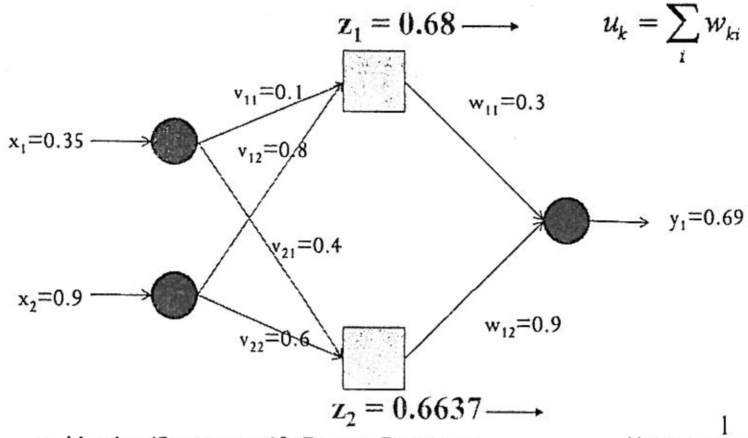
$$z_2 = \frac{1}{1 + e^{-0.68}} = 0.6637$$

23

Παράδειγμα

$$y_k = f(u_k) = \frac{1}{1 + e^{-u_k}}$$

$$u_k = \sum_i w_{ki} \cdot x_i$$



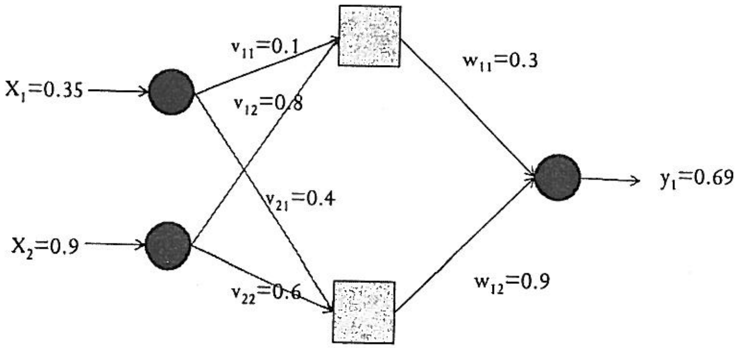
- Υπολογίζουμε την έξοδο του δευτέρου επιπέδου:

$$y_1 = \frac{1}{1 + e^{-0.8013}} = 0.69$$

24

Παράδειγμα

$$e_k = d_k \square y_k$$



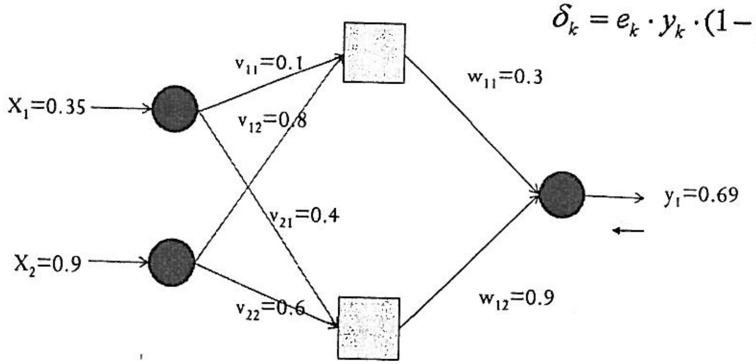
- Υπολογίζουμε το σφάλμα στην έξοδο: $e_1 = d_1 \square y_1 = 0.5 \square 0.69 = \square 0.19$

Παράδειγμα

$$e_k = d_k \square y_k$$

$$\delta_k = e_k \cdot f'(u_k)$$

$$\delta_k = e_k \cdot y_k \cdot (1 - y_k)$$

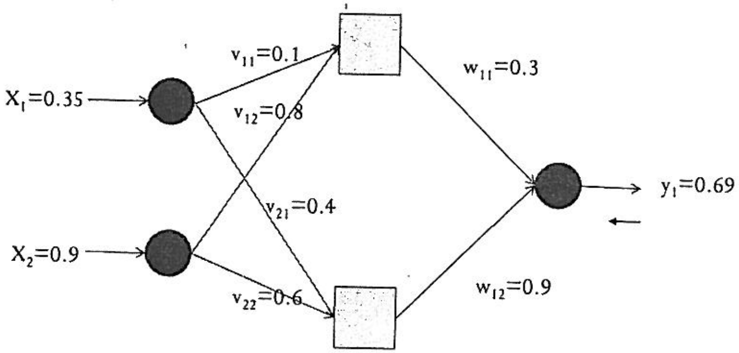


- Υπολογίζουμε το σφάλμα στην έξοδο και την τοπική κλίση: $e_1 = d_1 \square y_1 = 0.5 \square 0.69 = \square 0.19$

$$\delta_1 = -0.19 \cdot 0.69 \cdot (1 - 0.69) = -0.0406$$

Παράδειγμα

$$\Delta w_{ki} = \gamma \cdot \delta_k \cdot y_i$$

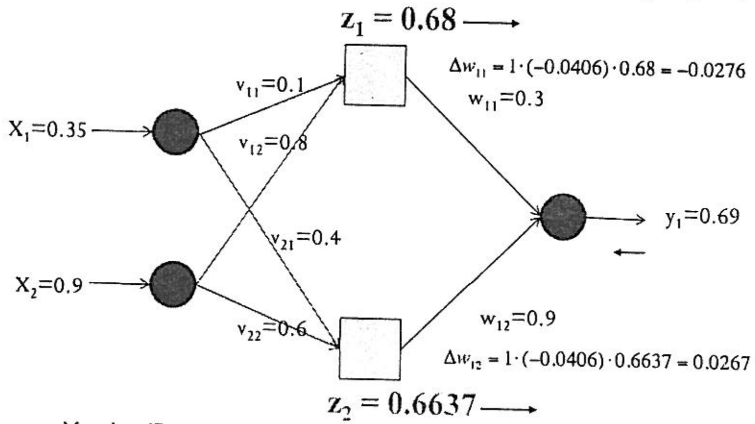


- Υπολογίζουμε το : Δw_{ki}

$$\delta_1 = \square 0.0406$$

Παράδειγμα

$$\Delta w_{ki} = \gamma \cdot \delta_k \cdot y_i$$



- Υπολογίζουμε το : Δw_{ki}

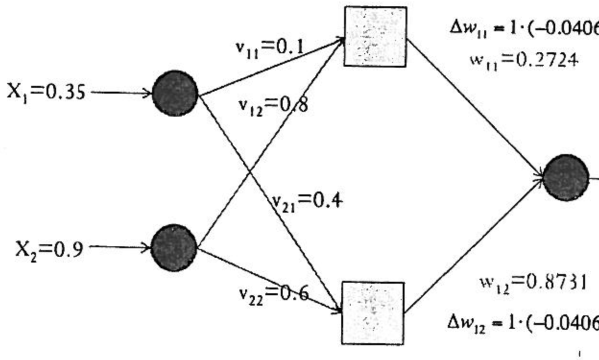
$$\delta_1 = \square 0.0406$$

Παράδειγμα

$$\Delta w_{ki} = \gamma \cdot \delta_k \cdot y_i$$

$\Delta w_{11} = 1 \cdot (-0.0406) \cdot 0.68 = -0.0276$

$w_{11}=0.2724$



$w_{12}=0.8731$

$\Delta w_{12} = 1 \cdot (-0.0406) \cdot 0.6637 = 0.0267$

- Ανανεώνουμε τα βάρη

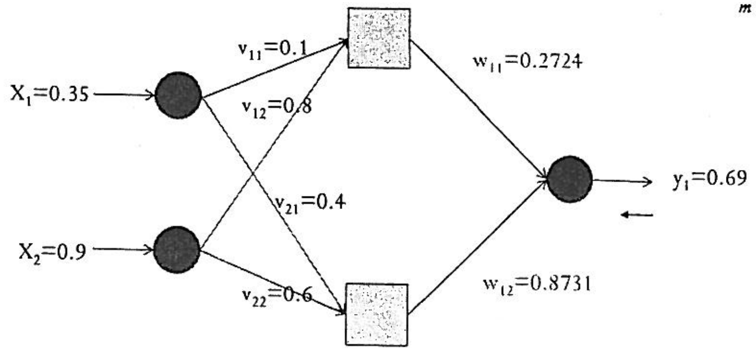
□ = □0.406

29

Παράδειγμα

$$\Delta w_{ji} = \gamma \cdot \delta_j \cdot y_i$$

$$\delta_j = y_j \cdot (1 - y_i) \cdot \sum_m \delta_m \cdot w_{mj}$$



- Υπολογίζουμε το σφάλμα προς τα πίσω

30

Παράδειγμα

$$\Delta w_{ji} = \gamma \cdot \delta_j \cdot y_i$$

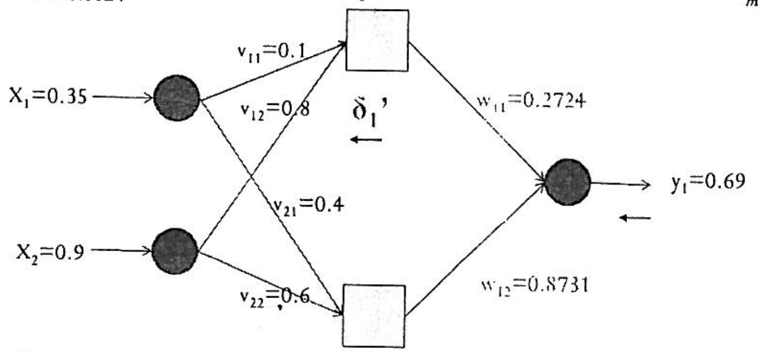
$$\delta_j = y_j \cdot (1 - y_i) \cdot \sum_m \delta_m \cdot w_{mj}$$

$\delta_1' = y_1 \cdot (1 - y_1) \cdot \delta_1 \cdot w_{11}$

$= 0.68 \cdot (1 - 0.68) \cdot (-0.0406) \cdot 0.2724$

$= -0.0024$

$z_1 = 0.68 \rightarrow$



$\delta_2' = y_2 \cdot (1 - y_2) \cdot \delta_1 \cdot w_{12}$

$= 0.6637 \cdot (1 - 0.6637) \cdot (-0.0406) \cdot 0.8731$

$= -0.0079$

31

Παράδειγμα

$$\Delta w_{ji} = \gamma \cdot \delta_j \cdot y_i$$

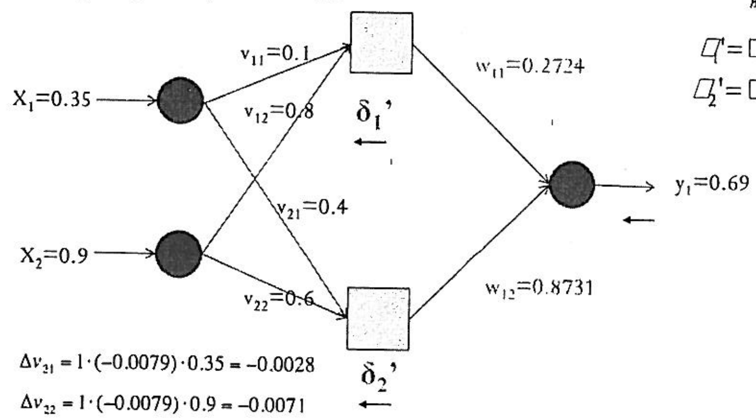
$$\delta_j = y_j \cdot (1 - y_i) \cdot \sum_m \delta_m \cdot w_{mj}$$

$\Delta v_{11} = 1 \cdot (-0.0024) \cdot 0.35 = -0.00084$

$\Delta v_{12} = 1 \cdot (-0.0024) \cdot 0.9 = -0.0022$

□' = □0.0024

$\square_2$' = □0.0079



$\Delta v_{21} = 1 \cdot (-0.0079) \cdot 0.35 = -0.0028$

$\Delta v_{22} = 1 \cdot (-0.0079) \cdot 0.9 = -0.0071$

32

Παράδειγμα

$$\Delta w_{ji} = \gamma \cdot \delta_j \cdot y_i$$

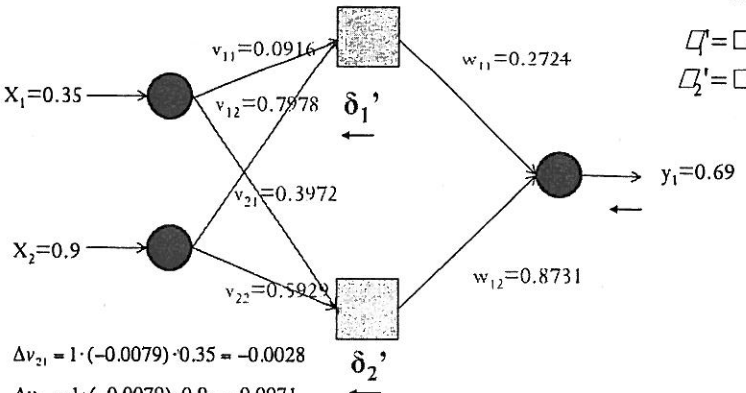
$$\delta_j = y_j \cdot (1 - y_i) \cdot \sum_m \delta_m \cdot w_{mj}$$

$\Delta v_{11} = 1 \cdot (-0.0024) \cdot 0.35 = -0.00084$

$\Delta v_{12} = 1 \cdot (-0.0024) \cdot 0.9 = -0.0022$

$\delta_1' = -0.0024$

$\delta_2' = -0.0079$



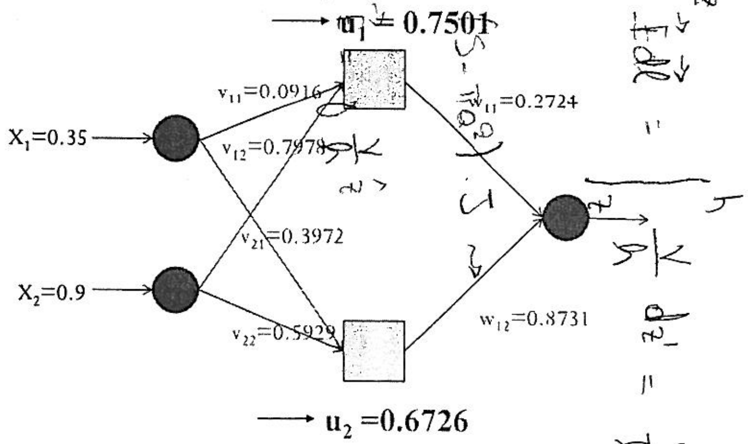
$\Delta v_{21} = 1 \cdot (-0.0079) \cdot 0.35 = -0.0028$

$\Delta v_{22} = 1 \cdot (-0.0079) \cdot 0.9 = -0.0071$

33

Παράδειγμα

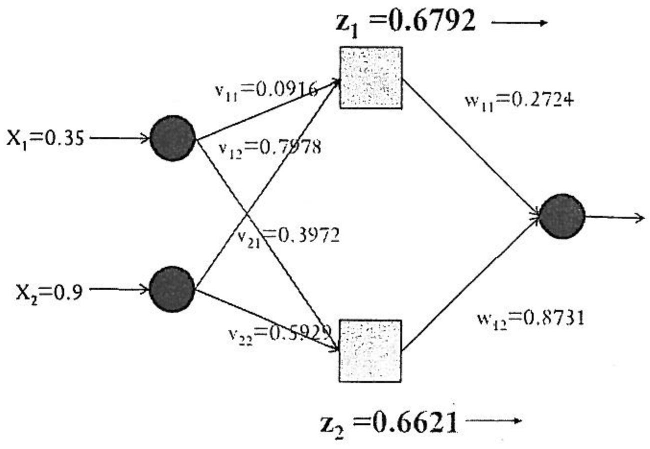
$u_k = \sum w_{ki} \cdot x_i$



34

Παράδειγμα

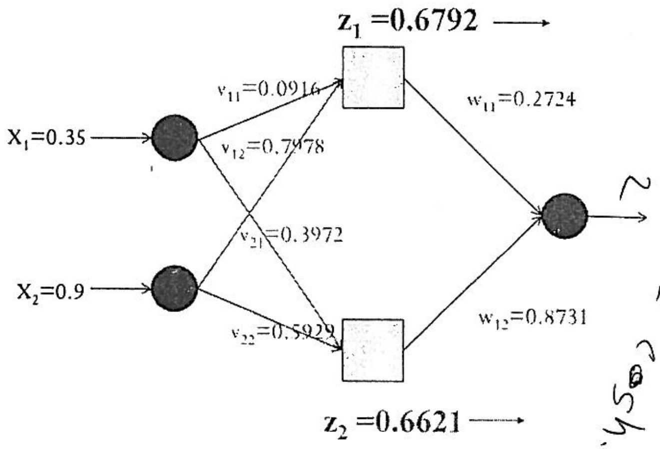
$$y_k = f(u_k) = \frac{1}{1 + e^{-u_k}}$$



35

Παράδειγμα

$$y_k = f(u_k) = \frac{1}{1 + e^{-u_k}}$$



Το σφάλμα από 0.19 έχει γίνει 0.182

36