ΤΕΧΝΗΤΗ *ΝΟΗΜΟΣΥΝΗ* – *ΥΠΟΛΟΓΙΣΤΙΚΗ ΝΟΗΜΟΣΥΝΗ*



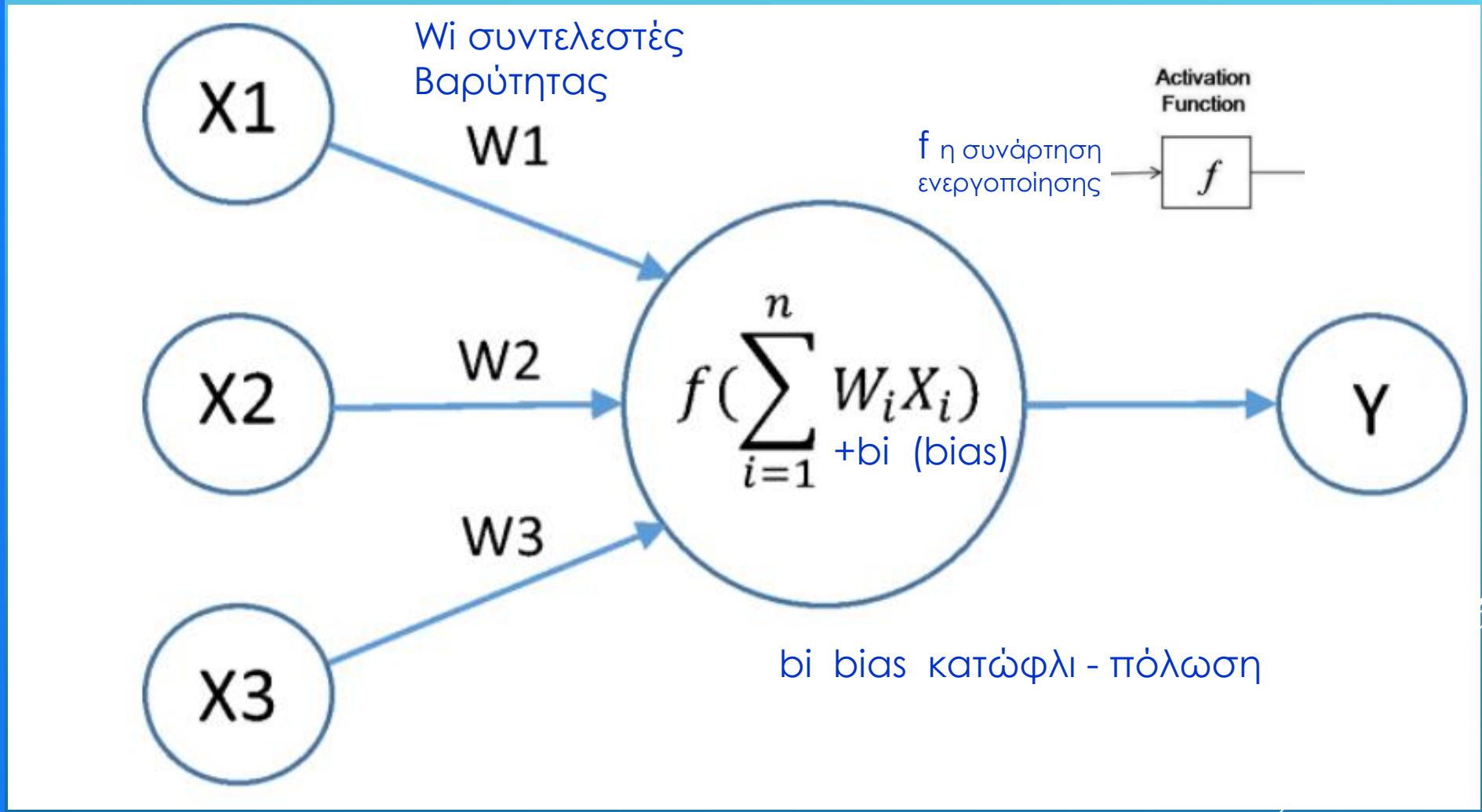
Λάζαρος Ηλιάδης Καθηγητής ΔΠΘ
Email: *liliadis@civil.duth.gr*

SHALLOW NEURAL NETWORKS
MULTI LAYER FEED FWD- BACK PROPAGATION!!

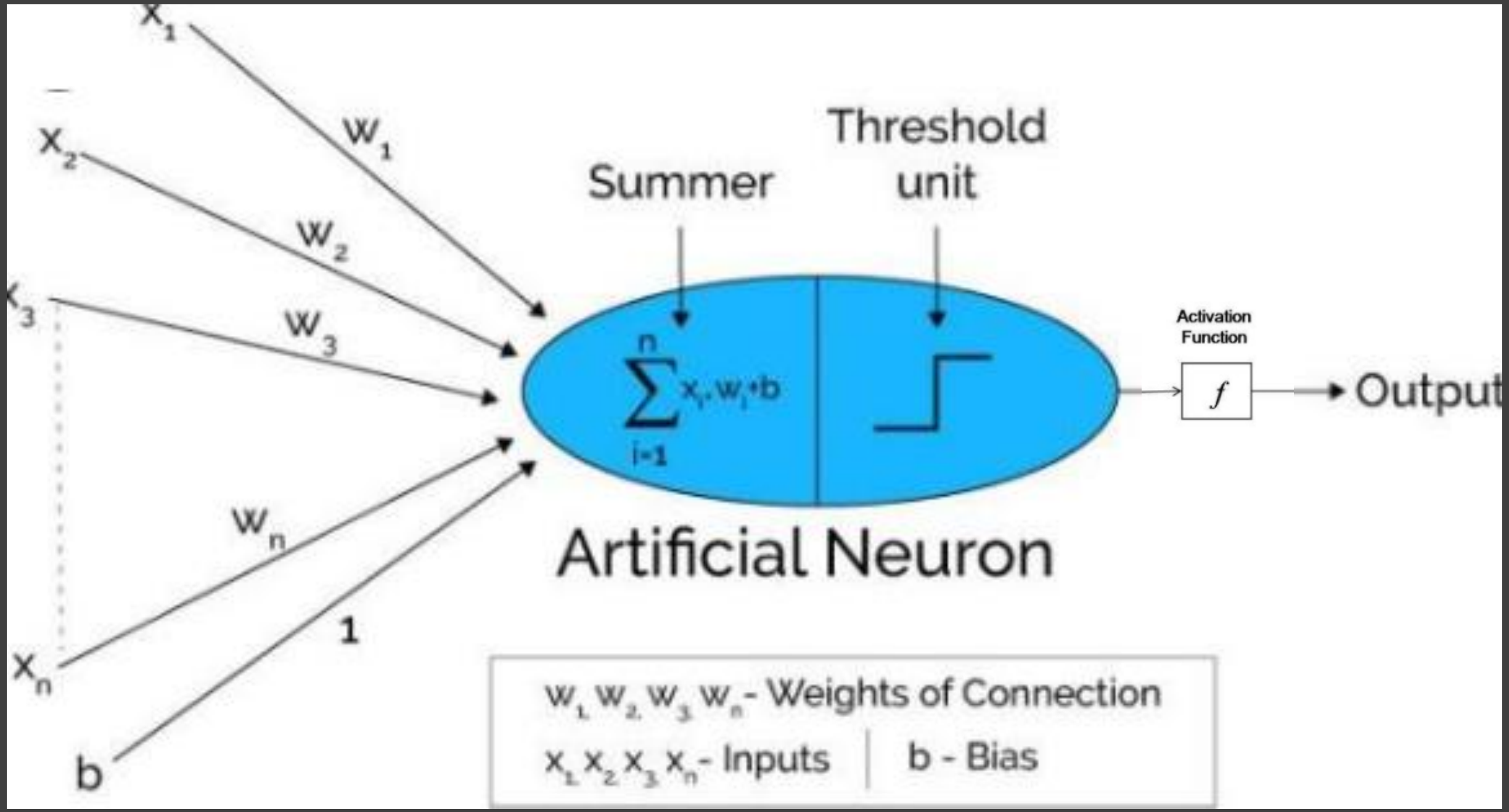
ΤΕΧΝΗΤΟΣ ΝΕΥΡΩΝΑΣ

Πόλωση ή κατώφλι (bias, threshold).

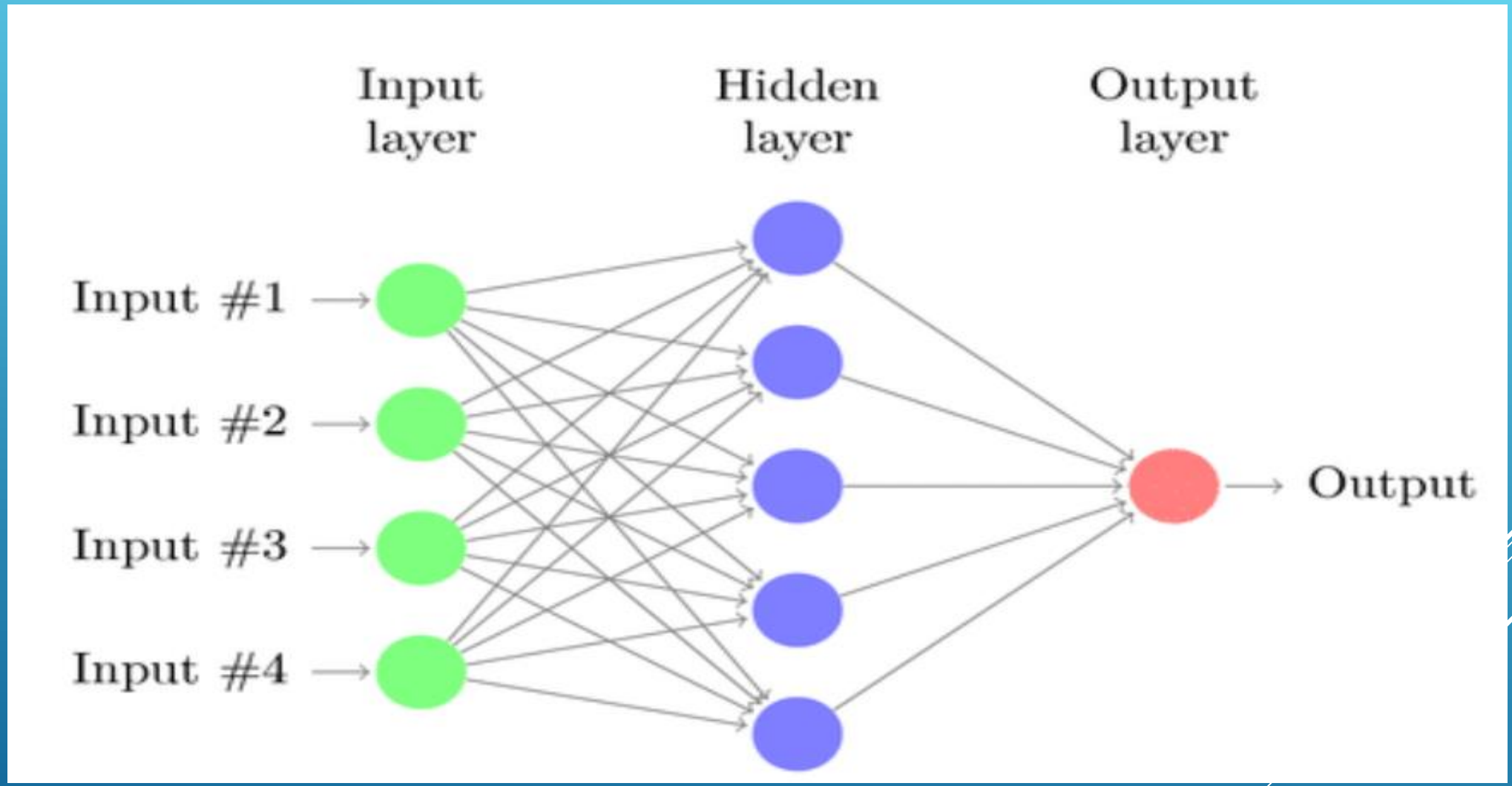
Η τιμή της εισόδου του είναι πάντα η μονάδα, Χκ=1. Εάν το συνολικό άθροισμα από τις υπόλοιπες εισόδους του νευρώνα είναι μεγαλύτερο από την τιμή αυτή, τότε ο νευρώνας ενεργοποιείται. Εάν είναι μικρότερο, τότε ο νευρώνας παραμένει ανενεργός. Η ιδέα προέκυψε από τα βιολογικά νευρικά κύτταρα.

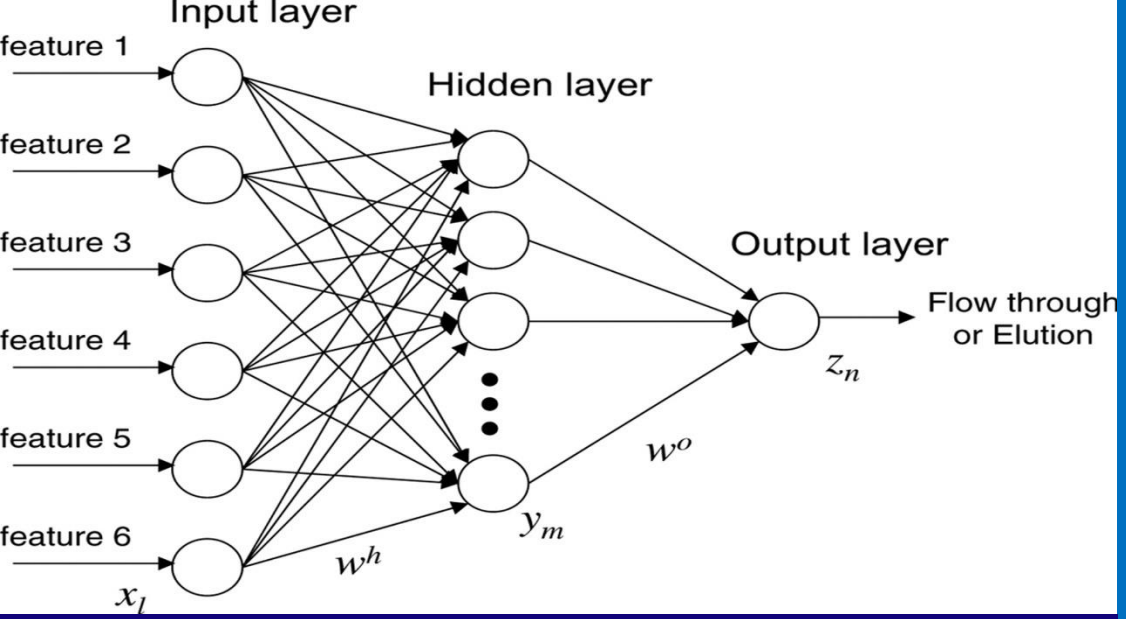


f (χ) ΕΙΝΑΙ Η ΣΥΝΑΡΤΗΣΗ ενεργοποίησης - ΜΕΤΑΦΟΡΑΣ


x1
x2
x3
xn
b
w1
w2
w3
wn
1
Summer
Threshold unit
Activation Function
f
Output
Artificial Neuron
w1 w2 w3 wn - Weights of Connection
x1 x2 x3 xn - Inputs
b - Bias

ΠΟΛΥΕΠΙΠΕΔΟ ΤΝΔ ΕΜΠΡΟΣΘΙΑΣ ΤΡΟΦΟΔΟΣΙΑΣ





- Εκπαίδευση

Οι μέθοδοι εκπαίδευσης μπορούν να διαχωριστούν σε:

- ✓ **Επιβλεπόμενη (supervised learning)**

  Το ΤΝΔ εκπαιδεύεται με συγκεκριμένες εισόδους οι οποίες ταιριάζουν με τις εξόδους.

- ✓ **Μη επιβλεπόμενη (unsupervised learning)**

  Μια μονάδα εξόδου εκπαιδεύεται να ανταποκρίνεται σε ομάδες προτύπων που υπάρχουν στην είσοδο..

- ✓ Η εκπαίδευση δύο σταδίων **(semi-supervised training)**

  Πρόκειται για τον συνδυασμό των παραπάνω μεθόδων εκπαίδευσης και εκτελείται σε δύο στάδια.

  Στο 1ο στάδιο, γίνεται εκπαίδευση χωρίς επίβλεψη, η οποία ομαδοποιεί τα δεδομένα εισόδου.

  Στο 2ο στάδιο γίνεται εκπαίδευση με επίβλεψη και δημιουργείται η συνάρτηση απεικόνισης εισόδου – εξόδου.

  Οι ομάδες του 1ου σταδίου χρησιμεύουν για να αρχικοποιηθεί το ΤΝΔ στο 2ο στάδιο.

- ❖ Ενισχυόμενη μάθηση (reinforcement learning)
  - ➢ Βασίζεται σε διαδικασίες ανταμοιβής σε σχέση με το αποτέλεσμα.

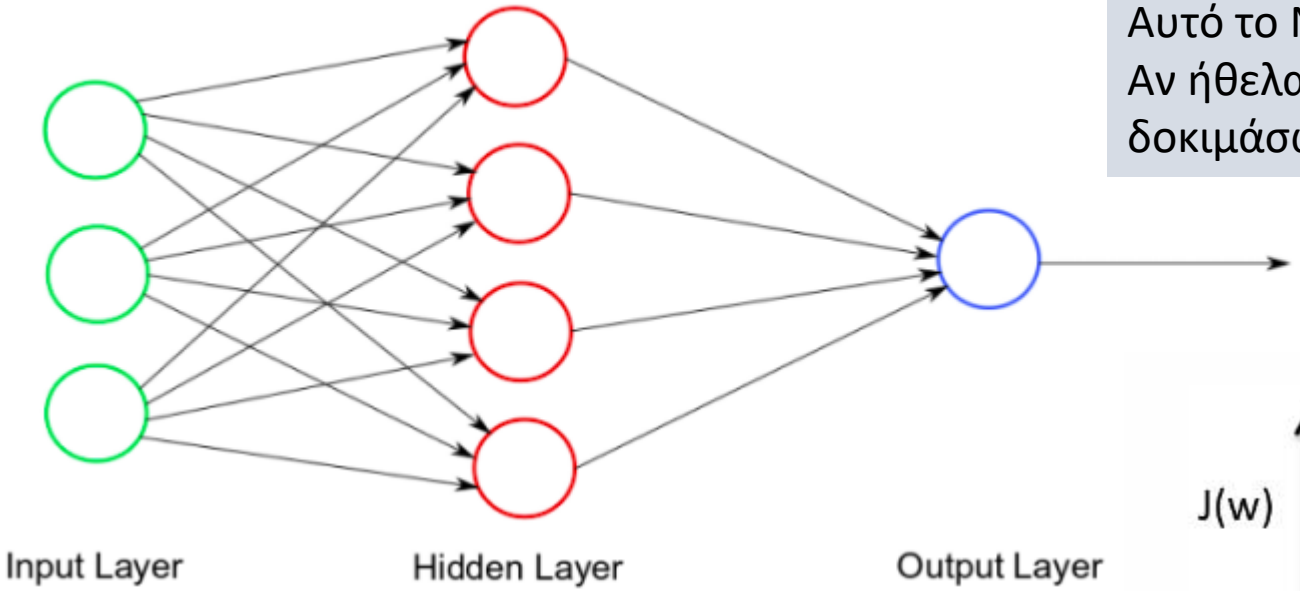
- ✓ Προσαρμογή βαρών
  - ❖ Η εκπαίδευση αποσκοπεί στην αναπροσαρμογή των βαρών των συνδέσεων μεταξύ των νευρώνων του ΤΝΔ, σύμφωνα με κάποιο κανόνα μετατροπής (Hebbian Learning Rule).

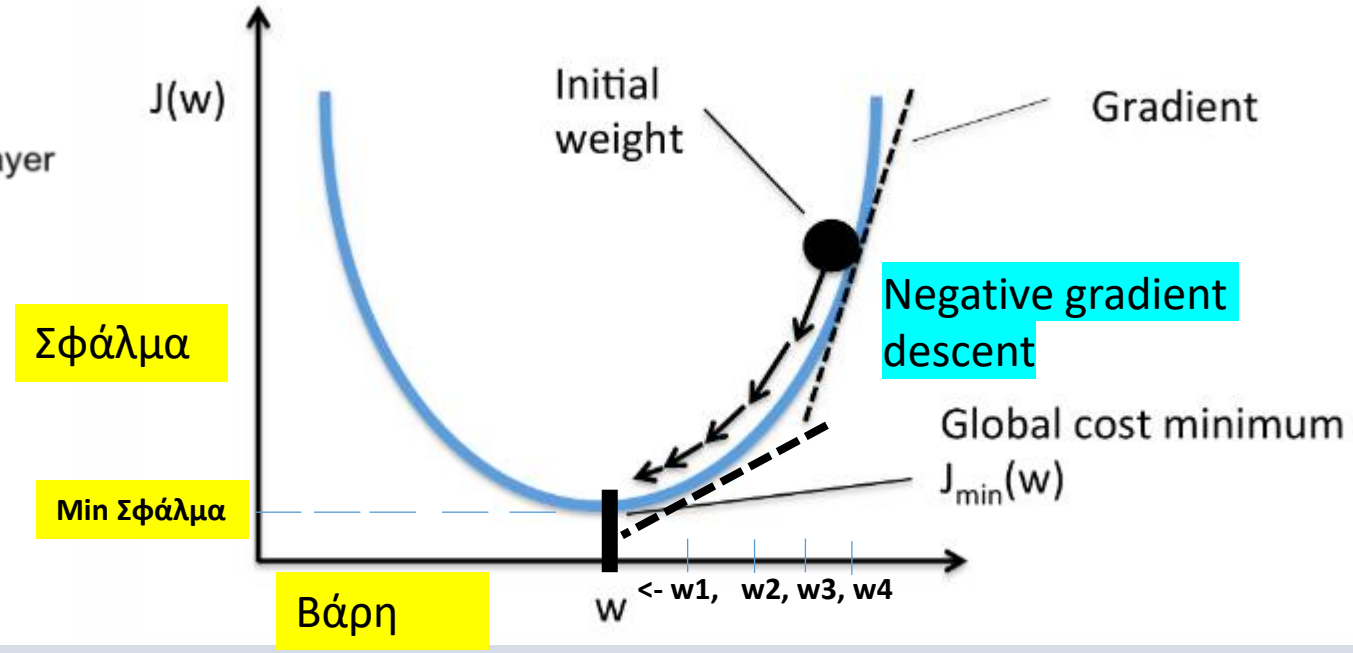
Πως θα βρω τα βέλτιστα ΒΑΡΗ;

Αυτό το Νευρωνικό δίκτυο έχει 16 βάρη!!

Αν ήθελα ακρίβεια 3 δεκαδικών ψηφίων και αν θα έπρεπε να δοκιμάσω όλους τους συνδυασμούς βαρών τότε θα προέκυπταν:

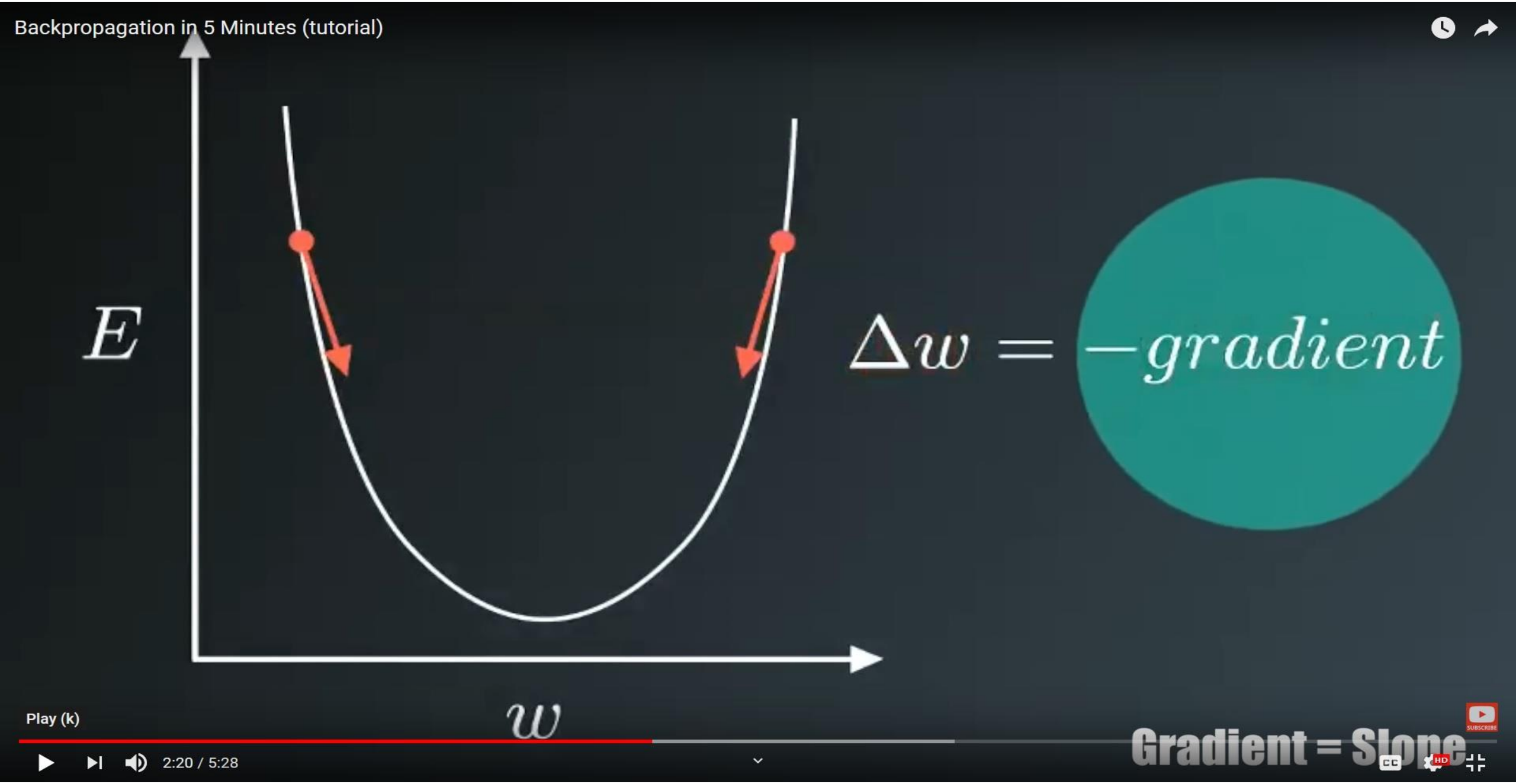
$1000^{16} = 10^{48}$ **Διαφορετικοί συνδυασμοί βαρών !!!**





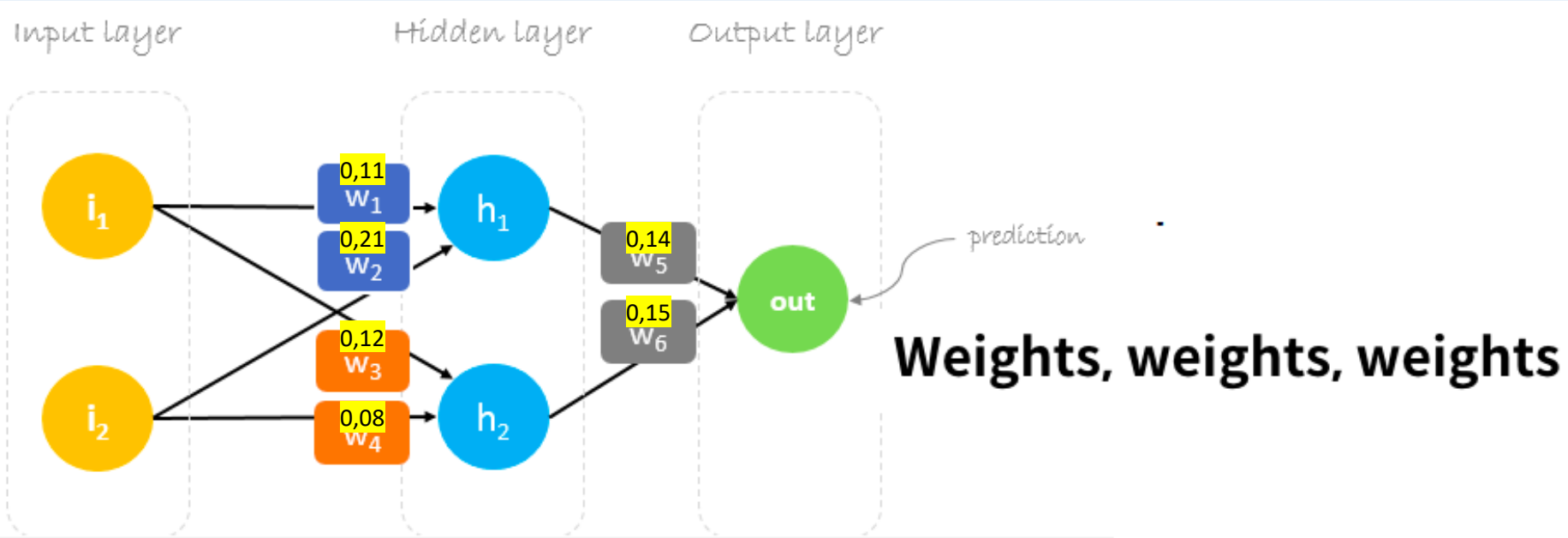
Έστω ότι αυτή είναι η συνάρτηση σφάλματος! Θα έπρεπε να μειώνω συνέχεια τα βάρη ώστε να ελαχιστοποιηθεί το σφάλμα. (ΟΛΙΚΟ ΜΕΓΙΣΤΟ). Αν συνεχίσω να μειώνω τα βάρη, μετά από κάποιο σημείο θα αρχίσει το σφάλμα να αυξάνεται. Πρέπει να σταματήσω στο κατάλληλο σημείο όσο έχω αρνητικό Gradient Descent (δηλαδή το σφάλμα μειώνεται). Η κλίση της γωνίας της εφαπτόμενης (παράγωγος) σημαίνει ότι το σφάλμα συνεχίζει να μειώνεται!! Αλλιώς αν γίνει θετική το σφάλμα θα αρχίσει να μεγαλώνει.

GRADIENT DESCENT



BACK PROPAGATION

Υπεραπλουστευμένο παράδειγμα με συνάρτηση activation F(x)=X για κάθε X>0 και στο κρυφό επίπεδο και στο εξόδου ώστε η παράγωγος να ισούται με 1.



Weights, weights, weights

Η εκπαίδευση του ΤΝΔ στοχεύει στο να βρει τα βάρη που θα ελαχιστοποιήσουν το σφάλμα πρόγνωσης. Μετά ο αλγόριθμος BP χρησιμοποιείται για να αλλάξει τα βάρη ώστε να απεικονίζονται οι εισαγόμενες τιμές ΣΩΣΤΑ στα αποτελέσματα.

Έστω ότι τα αρχικά βάρη είναι:

`w1 = 0.11` , `w2 = 0.21` , `w3 = 0.12` , `w4 = 0.08` , `w5 = 0.14` `w6 = 0.15`

Dataset

Το σύνολο δεδομένων έχει 2 νευρώνες Εισόδου I1 και I2 και 1 νευρώνα εξόδου (τιμή Z)

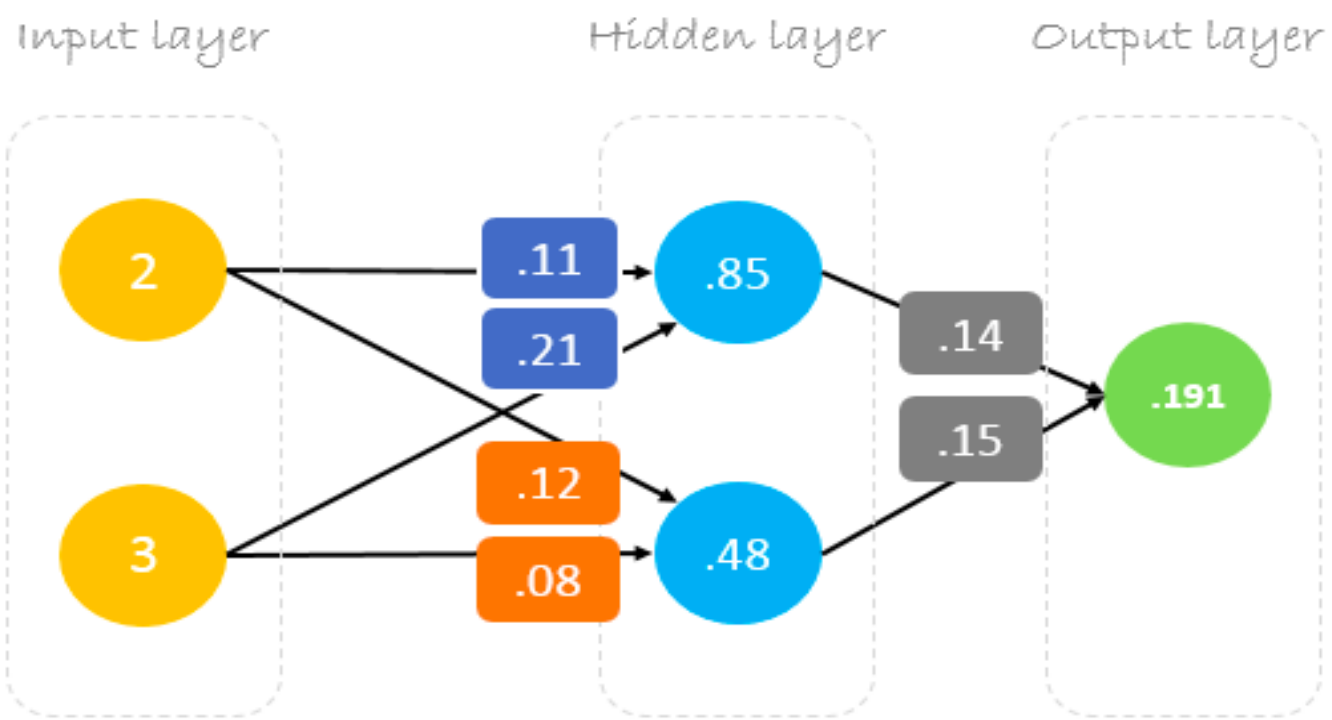


Έστω ότι στην είσοδο έχουμε τιμές Inputs=[2, 3] και στην έξοδο output=[1]

Επιθυμητή τιμή =1



FORWARD PASS ΠΡΟΩΘΗΣΗ ΠΡΟΣ ΤΟ ΕΠΙΠΕΔΟ ΕΞΟΔΟΥ



Αποτέλεσμα 0.191
ενώ ήθελα να πάρω 1

Forward Pass

ΠΟΛΛ/ΣΜΟΣ ΠΙΝΑΚΩΝ

$$[2 \quad 3] \cdot \begin{bmatrix} 0.11 & 0.12 \\ 0.21 & 0.08 \end{bmatrix} = [0.85 \quad 0.48] \cdot \begin{bmatrix} 0.14 \\ 0.15 \end{bmatrix} = [0.191]$$

ΑΝΑΛΥΣΗ ΤΟΥ ΠΟΛΛ/ΣΜΟΥ

2 x .11 + 3 x .21 = .85

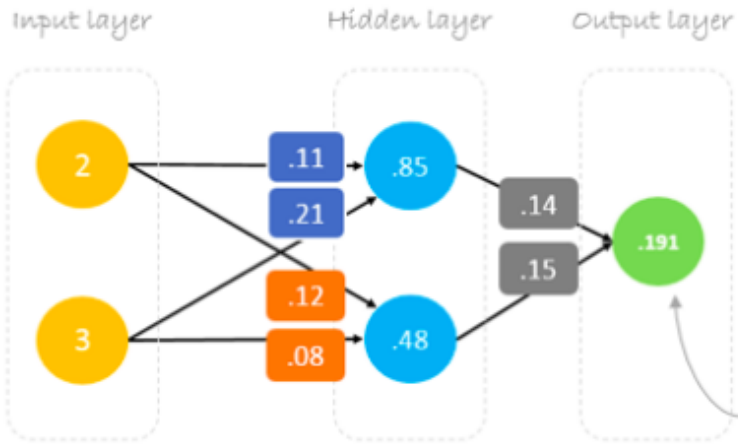
2 x .12 + 3 x .08 = .48

.85 x .14 + .48 x .15 = .191

ΕΥΡΕΣΗ ΣΦΑΛΜΑΤΟΣ!

Υπολογισμός του σφάλματος

Τώρα θα βρεθεί πόσο καλά πήγε το ΤΝΔ βρίσκοντας το σφάλμα. Το αποτέλεσμα που πήραμε δνε είναι ούτε καν κοντά σε αυτό που θέλουμε να πάρουμε.



Είσοδος

2 3

Σωστή τιμή-Actual value

1

Πρόβλεψη

actual

ΣΦΑΛΜΑ ΕΙΝΑΙ 0 ΟΤΑΝ ΕΙΝΑΙ ΣΩΣΤΗ Η ΠΡΟΒΛΕΨΗ

$$\text{Error} = \frac{1}{2}(\text{prediction} - \text{actual})^2$$

Το σφάλμα είναι πάντα θετικό λόγω ύψωσης στο τετράγωνο

το $\frac{1}{2}$ πολλαπλασιάζεται για να διευκολύνει τον υπολογισμό της παραγώγου στη συνέχεια

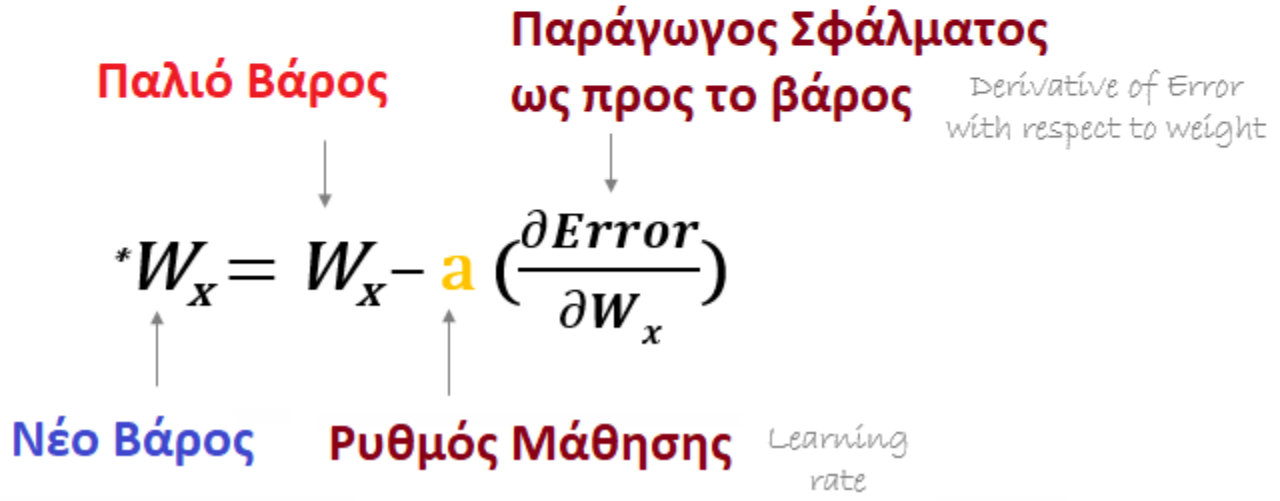
$$\text{Error} = \frac{1}{2}(0.191 - 1.0)^2 = 0.327$$

GRADIENT DESCENT ΑΛΛΑΓΗ ΒΑΡΩΝ

Πρέπει λοιπόν να αλλάξουμε τα βάρη και το ερώτημα είναι πως;
Η απάντηση λέγεται Backpropagation

Πρόκειται για αντίστροφη μετάδοση του σφάλματος. Είναι ένας μηχανισμός που χρησιμοποιεί τον αλγόριθμο Gradient Descent! Υπολογίζει την μερική παράγωγο του Σφάλματος ως προς τα βάρη του ΤΝΔ

Το Gradient Descent είναι ένας αλγόριθμος βελτιστοποίησης που χρησιμοποιεί επαναλήψεις για να βρει την ελάχιστη τιμή μιας συνάρτησης.



ΜΑΘΗΜΑΤΙΚΑ ΕΠΑΝΑΠΡΟΣΔΙΟΡΙΣΜΟΥ ΒΑΡΟΥΣ

Η παράγωγος της συναρτησης σφάλματος υπολογίζεται με τον πιο κάτω κανόνα της αλυσίδας

ΕΑΝ Z είναι μια σύνθετη συνάρτηση της μορφής: $z = f(x,y)$ **όπου** $x = x(t)$ **και** $y = y(t)$

τότε $\frac{dz}{dt} = \frac{\partial z}{\partial x}.\frac{dx}{dt} + \frac{\partial z}{\partial y}.\frac{dy}{dt}$ **όπου** Z=Error t=W6 X=Εκτιμώμενη τιμή Y=πραγματική

Υ είναι σταθερά πραγματική τιμή

ΝΕΟ ΒΑΡΟΣ $^{*}W_6 = W_6 - a\left(\frac{\partial Error}{\partial W_6}\right)$ *learning rate.*

Prediction=(i1 w1+i2w2)w5 +(i1w3+i2W4)w6

$$\frac{\partial Error}{\partial W_6} = \frac{\partial Error}{\partial prediction} * \frac{\partial prediction}{\partial W_6}$$ chain rule

0

Error = $\frac{1}{2}$(prediction – actual)^2

συντελεστής

$$\frac{\partial Error}{\partial W_6} = \frac{\frac{1}{2}(predictoin-actula)^2}{\partial prediciton} * \frac{\partial\left(i_1 w_1 + i_2 w_2\right) w_5 + \left(i_1 w_3 + i_2 w_4\right) w_6}{\partial W_6}$$

i1 και i2 είναι οι είσοδοι

$$\frac{\partial Error}{\partial W_6} = 2 * \frac{1}{2}(\text{predictoin} - \text{actula}) \frac{\partial(predictoin-actula)}{\partial prediciton} * \left(i_1 w_3 + i_2 w_4\right)$$ ← Το ονομάζω h2

$$\frac{\partial Error}{\partial W_6} = (\text{predictoin} - \text{actula}) * (h_2)$$

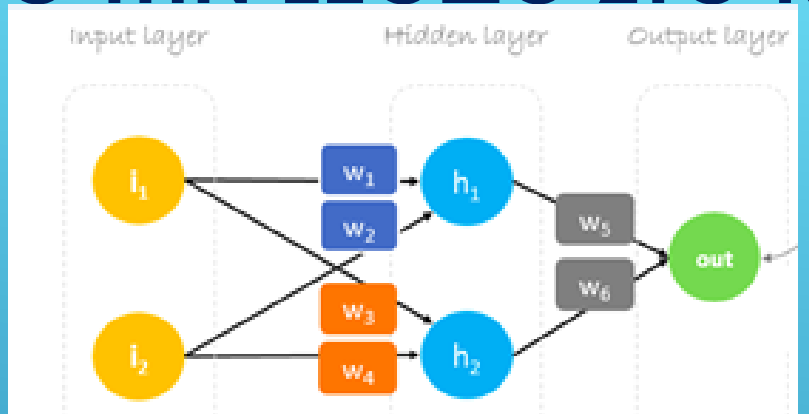
Δ = prediction – actual ← delta

Το ονομάζω Δ

$$\frac{\partial Error}{\partial W_6} = \Delta h_2$$

h2=i1w3+i2w4

ΑΛΛΑΓΗ ΒΑΡΩΝ ΑΠΌ ΤΗΝ ΕΞΟΔΟ ΣΤΟ ΚΡΥΦΟ



Για να αλλάξω το W6 εφαρμόζω τον τύπο $^*W_6 = W_6 - a\,\Delta h_2$ παρόμοια για κάθε βάρος από το επίπεδο εξόδου προς το κρυφό

$$^*W_5 = W_5 - a\,\Delta h_1$$

Έτσι όπως πηγαίνουμε προς τα πίσω για να αλλάξουμε τα `w1` , `w2` , `w3` `w4` που υπάρχουν μεταξύ του επιπέδου Εισόδου και Κρυφού ενδεικτικά η μερική παράγωγος της συνάρτησης σφάλματος όσον αφορά στο `w1` είναι η εξής!

$$\frac{\partial Error}{\partial W_1} = \frac{\partial Error}{\partial prediction} * \frac{\partial prediction}{\partial h_1} * \frac{\partial h_1}{\partial W_1}$$ chain rule

$$\frac{\partial Error}{\partial W_1} = \frac{\partial \frac{1}{2}(predictoin - actula)^2}{\partial prediciton} * \frac{\partial (h_1)\,w_5 + (h_2)\,w_6}{\partial h_1} * \frac{\partial\, i_1 w_1 + i_2 w_2}{\partial w_1}$$

w5 → 0 ; i1 → 0

$$\frac{\partial Error}{\partial W_1} = 2 * \frac{1}{2}(predictoin - actula)\frac{\partial (predictoin - actula)}{\partial prediciton} * (w_5) * (i_1)$$

Error = $\frac{1}{2}$(prediction – actual)²

prediction = $(h_1)\,w_5 + (h_2)\,w_6$

$h_1 = i_1 w_1 + i_2 w_2$

ΝΕΑ ΒΑΡΗ

Το a Learning Rate είναι πολύ σημαντικό

updated weights

ΝΕΑ ΒΑΡΗ

$$
\begin{aligned}
{}^{*}w_6 &= w_6 - a\,(h_2 \,.\, \Delta) \\
{}^{*}w_5 &= w_5 - a\,(h_1 \,.\, \Delta) \\
{}^{*}w_4 &= w_4 - a\,(i_2 \,.\, \Delta w_6) \\
{}^{*}w_3 &= w_3 - a\,(i_1 \,.\, \Delta w_6) \\
{}^{*}w_2 &= w_2 - a\,(i_2 \,.\, \Delta w_5) \\
{}^{*}w_1 &= w_1 - a\,(i_1 \,.\, \Delta w_5)
\end{aligned}
$$

Μπορούμε να παραστήσουμε τους πιο πάνω υπολογισμούς με πράξεις πινάκων

$$\begin{bmatrix} w_5 \\ w_6 \end{bmatrix} = \begin{bmatrix} w_5 \\ w_6 \end{bmatrix} - a\,\Delta \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} = \begin{bmatrix} w_5 \\ w_6 \end{bmatrix} - \begin{bmatrix} a h_1 \Delta \\ a h_2 \Delta \end{bmatrix}$$

$$\begin{bmatrix} w_1 & w_3 \\ w_2 & w_4 \end{bmatrix} = \begin{bmatrix} w_1 & w_3 \\ w_2 & w_4 \end{bmatrix} - a\,\Delta \begin{bmatrix} i_1 \\ i_2 \end{bmatrix} . \begin{bmatrix} w_5 & w_6 \end{bmatrix} = \begin{bmatrix} w_1 & w_3 \\ w_2 & w_4 \end{bmatrix} - \begin{bmatrix} a\,i_1 \Delta w_5 & a\,i_1 \Delta w_6 \\ a\,i_2 \Delta w_5 & a\,i_2 \Delta w_6 \end{bmatrix}$$

ΡΥΘΜΟΣ ΜΑΘΗΣΗΣ

Learning rate: **Είναι υπερπαράμετρος την τιμή της οποίας καλούμαστε να εκτιμήσουμε**

Αν LR Μικρό το ΤΝΔ θα κάνει άπειρα μικρά βήματα προς το minError=>άπειρος χρόνος εκτέλεσης΄ίσωςποτέ να μην φτάσει σε convergence point. Αν LR Μεγάλο θα ξεπεράσει το minError point στο gradient descend!!

$\Delta = 0.191 - 1 = -0.809$ ← Delta = prediction - actual

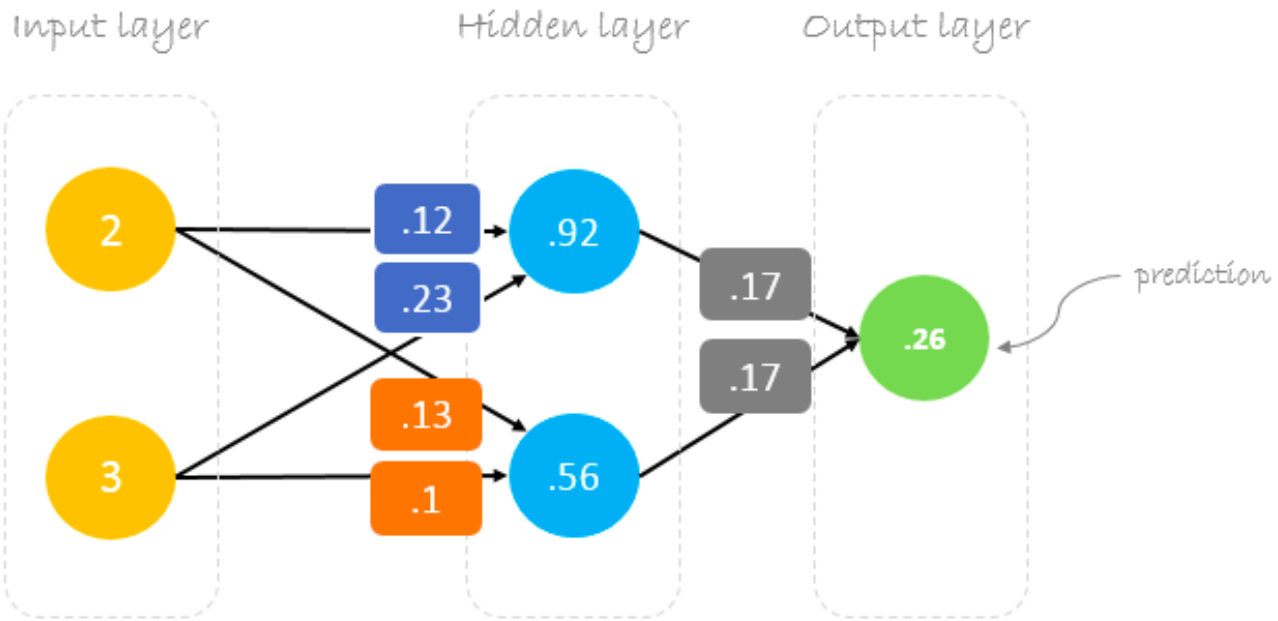
$a = 0.05$ ← Learning rate, **αρχικοποιούμε την τιμή του εμείς**

$$\begin{bmatrix} w_5 \\ w_6 \end{bmatrix} = \begin{bmatrix} 0.14 \\ 0.15 \end{bmatrix} - 0.05(-0.809)\begin{bmatrix} 0.85 \\ 0.48 \end{bmatrix} = \begin{bmatrix} 0.14 \\ 0.15 \end{bmatrix} - \begin{bmatrix} -0.034 \\ -0.019 \end{bmatrix} = \begin{bmatrix} 0.17 \\ 0.17 \end{bmatrix}$$

$$\begin{bmatrix} w_1 & w_3 \\ w_2 & w_4 \end{bmatrix} = \begin{bmatrix} .11 & .12 \\ .21 & .08 \end{bmatrix} - 0.05(-0.809)\begin{bmatrix} 2 \\ 3 \end{bmatrix}.[0.14 \quad 0.15] = \begin{bmatrix} .11 & .12 \\ .21 & .08 \end{bmatrix} - \begin{bmatrix} -0.011 & -0.012 \\ -0.017 & -0.018 \end{bmatrix} = \begin{bmatrix} .12 & .13 \\ .23 & .10 \end{bmatrix}$$

Έτσι βρίσκω τα νέα βάρη

Τώρα χρησιμοποιώντας τα νέα βάρη πάμε να κάνουμε εμπρόσθιο πέρασμα (forward pass)



Forward Pass

Πολλαπλασιασμός Πινάκων

$$[2 \quad 3] \cdot \begin{bmatrix} 0.12 & 0.13 \\ 0.23 & 0.10 \end{bmatrix} = [0.92 \quad 0.56] \cdot \begin{bmatrix} 0.17 \\ 0.17 \end{bmatrix} = [0.26]$$

Details

2 x .12 + 3 x .23 = .85

2 x .13 + 3 x .10 = .48

.92 x .17 + .56 x .17 = .26

Η πρόγνωση 0.26 **είναι πιό κοντά στο actual output**

Επαναλαμβάνεται το μπρος - πίσω ώσπου να μειωθεί πολύ το σφάλμα

ΣΥΝΑΡΤΗΣΕΙΣ ΕΝΕΡΓΟΠΟΙΗΣΗΣ/ΜΕΤΑΦΟΡΑΣ –TRANSFER FUNCTIONS

Στο κρυφό επίπεδο πάντοτε χρησιμοποιείται ΜΗ Γραμμική συνάρτηση. Στο επίπεδο εξόδου μπορεί να χρησιμοποιηθεί είτε Γραμμική πχ ReLU είτε Μη Γραμμική ίδια με αυτήν του κρυφού επιπέδου.

Στην αναγνώριση προτύπων – Pattern Recognition όπου οι τιμές είναι διακριτές και το αποτέλεσμα είναι Δυαδικός αριθμός μπορεί να χρησιμοποιηθεί η Sigmoid transfer function η οποία είναι αλλιώς γνωστή ως Logistic/ ΛΟΓΙΣΤΙΚΗ ΣΙΓΜΟΕΙΔΗΣ που παράγει τιμές από το 0 έως και το +1.

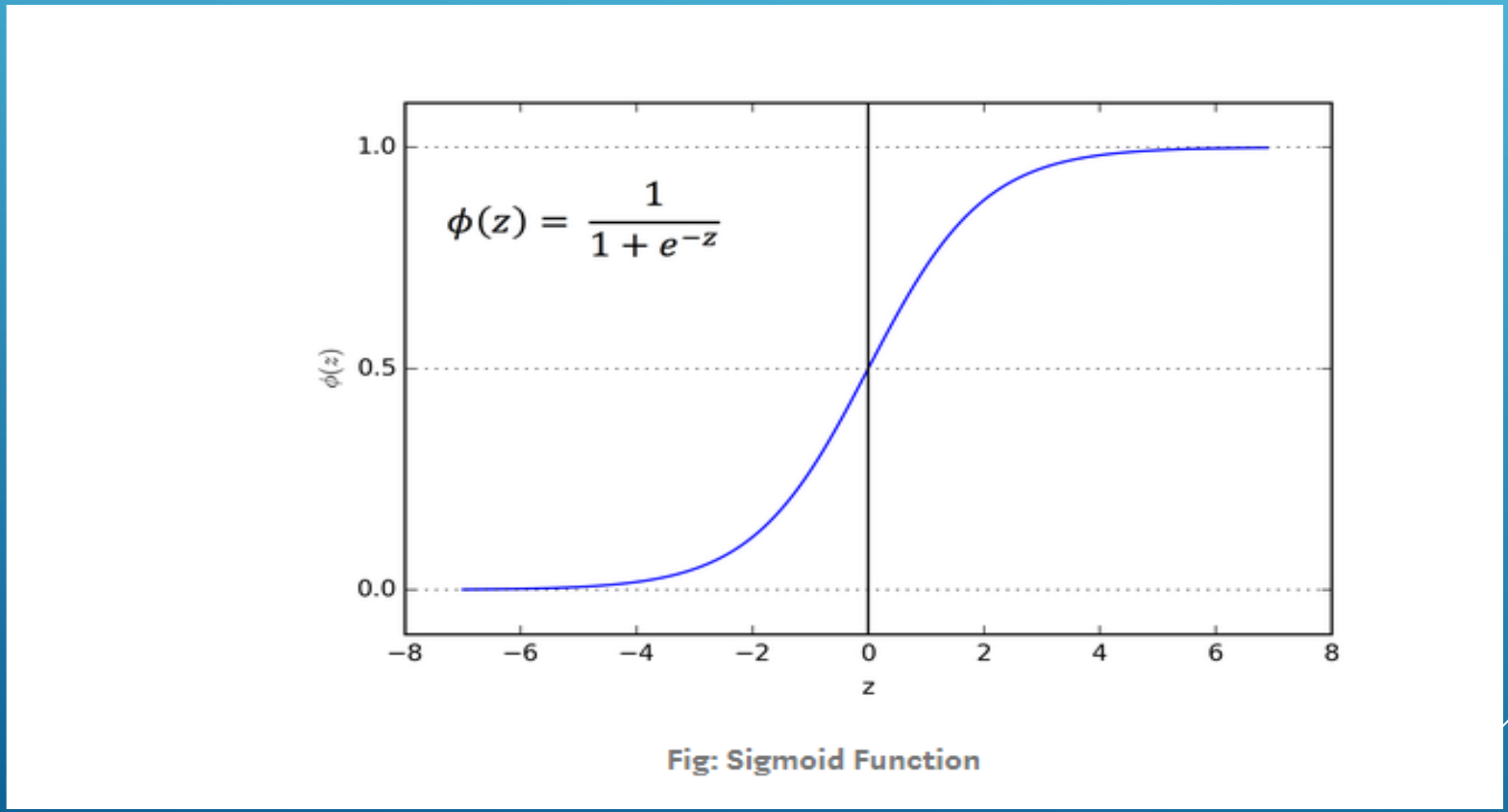
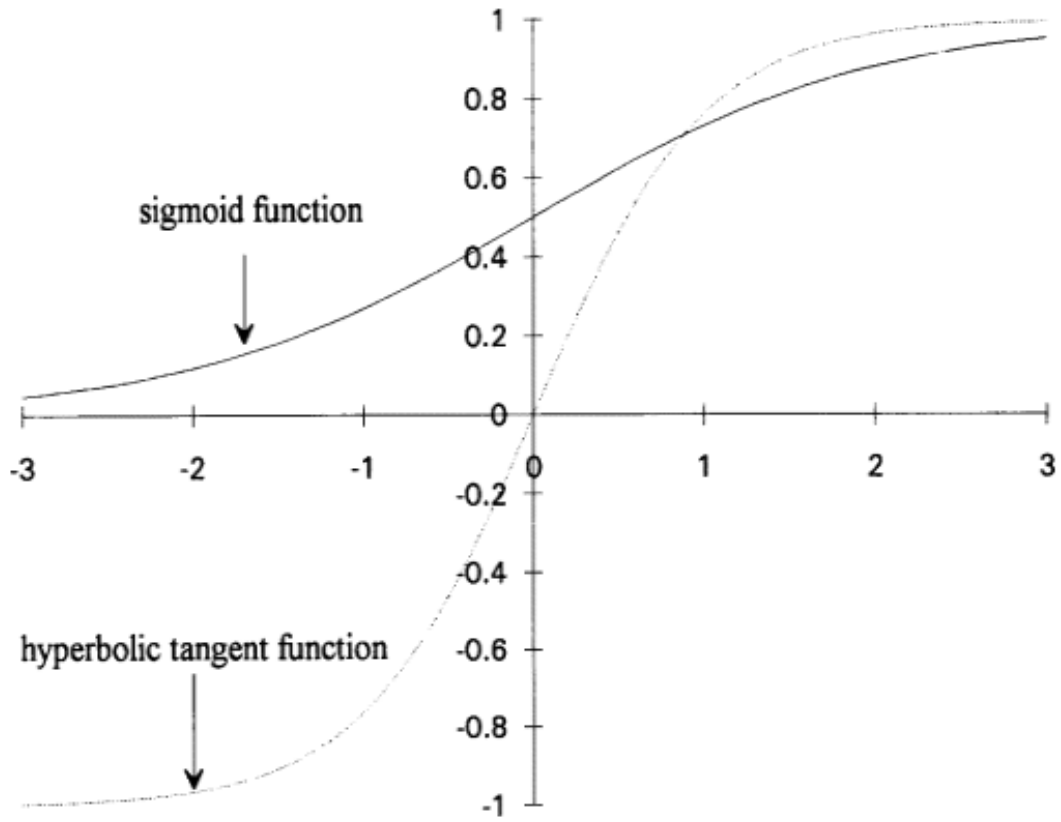


Fig: Sigmoid Function



ΥΠΕΡΒΟΛΙΚΗ ΕΦΑΠΤΟΜΕΝΗ

F(x)= TanH (X)

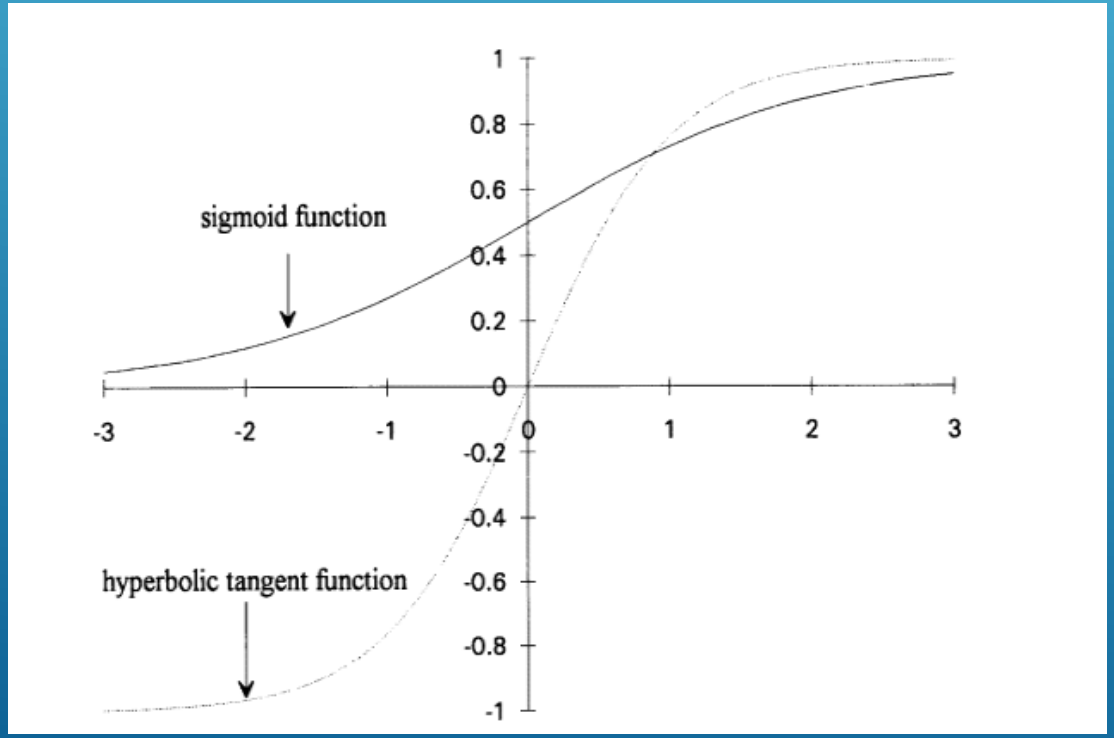
ΜΗ ΓΡΑΜΜΙΚΕΣ ΣΥΝΑΡΤΗΣΕΙΣ ΕΝΕΡΓΟΠΟΙΗΣΗΣ – ΜΕΤΑΦΟΡΑΣ ΙΙ

Στην διαδικασία regression / function approximation μπορώ να χρησιμοποιήσω είτε την Sigmoid είτε την TanH (Tangetn Hyperbolic) Εφαπτομένη Υπερβολοειδή **που παράγει τιμές από το -1 έως και το +1**

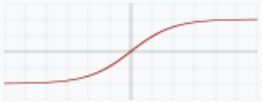
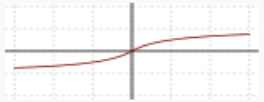
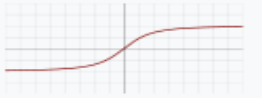
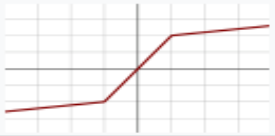
Η Sigmoid θα παράξει τιμές κοντά στο 0 εάν το όρισμα της συνάρτησης είναι ουσιαστικά αρνητικό.

Έτσι, η έξοδος αυτού του hidden neuron θα είναι κοντά στο μηδέν, μειώνοντας έτσι τον ρυθμό εκμάθησης για όλα τα επόμενα βάρη ίσως έτσι επιβραδυνθεί εξαιρετικά η μάθηση (πρακτικά θα σταματήσει)!!!

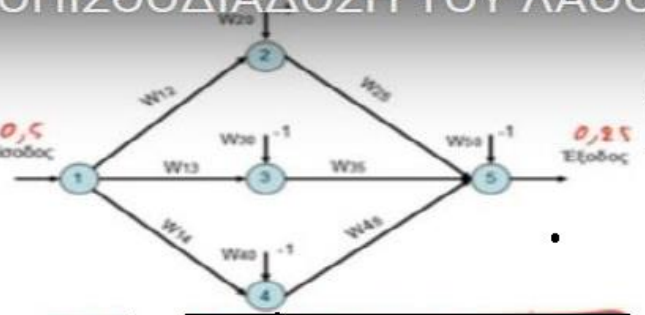
Προτιμάω την TanH στο Regression. Η TanH έχει μεγαλύτερη κλίση από την Sigmoid άρα μπορεί να αξιολογήσει καλύτερα μικρές διακυμάνσεις στην τιμή της εισόδου. Επίσης παίρνει και αρνητικές τιμές για αρνητικές εισόδους



Πχ για εισόδους 0,1 και 0,2. Η sigmoid δίνει 0,525 και 0,550 διαφορά μόνο 0,025 ενώ η TanH δίνει 0,100 και 0,197 διαφορά 0,097

Name	Plot	Equation	Derivative (with respect to x)
Identity		$f(x) = x$	$f'(x) = 1$
Binary step		$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$	$f'(x) = \begin{cases} 0 & \text{for } x \neq 0 \\ ? & \text{for } x = 0 \end{cases}$
Logistic (a.k.a. Sigmoid or Soft step)		$f(x) = \sigma(x) = \frac{1}{1+e^{-x}}$ [1]	$f'(x) = f(x)(1 - f(x))$
TanH		$f(x) = \tanh(x) = \frac{(e^x - e^{-x})}{(e^x + e^{-x})}$	$f'(x) = 1 - f(x)^2$
SQNL[10]		$f(x) = \begin{cases} 1 & : x > 2.0 \\ x - \frac{x^2}{4} & : 0 \leq x \leq 2.0 \\ x + \frac{x^2}{4} & : -2.0 \leq x < 0 \\ -1 & : x < -2.0 \end{cases}$	$f'(x) = 1 \mp \frac{x}{2}$
ArcTan		$f(x) = \tan^{-1}(x)$	$f'(x) = \frac{1}{x^2+1}$
ArSinH		$f(x) = \sinh^{-1}(x) = \ln\left(x + \sqrt{x^2+1}\right)$	$f'(x) = \frac{1}{\sqrt{x^2+1}}$
ElliotSig[11][12] Softsign[13][14]		$f(x) = \frac{x}{1+\lvert x\rvert}$	$f'(x) = \frac{1}{(1+\lvert x\rvert)^2}$
Inverse square root unit (ISRU)[15]		$f(x) = \frac{x}{\sqrt{1+\alpha x^2}}$	$f'(x) = \left(\frac{1}{\sqrt{1+\alpha x^2}}\right)^3$
Inverse square root linear unit (ISRLU)[15]		$f(x) = \begin{cases} \frac{x}{\sqrt{1+\alpha x^2}} & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$	$f'(x) = \begin{cases} \left(\frac{1}{\sqrt{1+\alpha x^2}}\right)^3 & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$
Piecewise Linear Unit (PLU)[16]		$f(x) = \max(\alpha(x+c) - c, \min(\alpha(x-c)+c, x))$	$f'(x) = \begin{cases} \alpha & \text{for } \lvert x\rvert > c \\ 1 & \text{for } \lvert x\rvert < c \end{cases}$
Rectified linear unit (ReLU)[17]		$f(x) = \begin{cases} 0 & \text{for } x \leq 0 \\ x & \text{for } x > 0 \end{cases}$	$f'(x) = \begin{cases} 0 & \text{for } x \leq 0 \\ 1 & \text{for } x > 0 \end{cases}$

ΟΠΙΣΘΟΔΙΑΔΟΣΗ ΤΟΥ ΛΑΘΟΥΣ - Εφαρμογή 1



Σχεδιάζουμε ένα απλό ΤΝΔ που παρουσιάζεται στο παρακάτω σχήμα, ώστε η έξοδός του να παράγει το τετράγωνο της εισόδου του. Οι νευρώνες του κρυφού επιπέδου έχουν συνάρτηση ενεργοποίησης τη σιγμοειδή συνάρτηση $S(x) = 1/(1+e^{-x})$, ενώ ο νευρώνας εξόδου τη γραμμική συνάρτηση αντίστοιχα. Επίσης, υπενθυμίζεται ότι τα κατώφλια έχουν είσοδο την τιμή -1.

w12=w13=0,10	W14=0,20 W25=0,30	W35=0,82	W45=0,20 W20=0,10	W30=0,10	W40=0,10 W50=0,10	

Για όλα τα παρακάτω ερωτήματα χρησιμοποιείστε ακρίβεια 2 δεκαδικών.

(α) Ποιο είναι το σφάλμα στην έξοδο του δικτύου αν εισάγουμε την τιμή 0.5 ως είσοδο

(β) Να υπολογίσετε τις τιμές των ανανεωμένων βαρών που προκύπτουν από την εκπαίδευση του δικτύου με τον αλγόριθμο οπισθοδιάδοσης του λάθους, χρησιμοποιώντας την τιμή εισόδου του ερωτήματος (α). Δίνεται ότι ο ρυθμός εκπαίδευσης είναι 0.2. Να συμπληρωθεί ο Πίνακας 1.

d_5
0,25

y_1	y_2	y_3	y_4	y_5
0,5	0,49	0,49	0,50	0,55

e_5
-0,30

δ_2	δ_3	δ_4	δ_5
			-0,30

Πίνακας 1 Ανανεωμένα βάρη του ΤΝΔ

$W_{12\text{ανεν}}$	$W_{13\text{ανεν}}$	$W_{14\text{ανεν}}$	$W_{25\text{ανεν}}$	$W_{35\text{ανεν}}$	$W_{45\text{ανεν}}$	$W_{20\text{ανεν}}$	$W_{30\text{ανεν}}$	$W_{40\text{ανεν}}$	$W_{50\text{ανεν}}$

ΠΡΟΣ ΤΑ ΕΜΠΡΟΣ ΠΕΡΑΣΜΑ

ΝΕΥΡΩΝΑΣ 1 (εισόδου)

$y_1 = x_1 = 0,5$

ΝΕΥΡΩΝΑΣ 2 (κρυφός)

Δυναμικό: $v_2 = w_{12} \cdot y_1 + w_{20}(-1) = -0,05$

Ενεργοποίηση: $y_2 = \varphi(v_2) = \frac{1}{1+e^{-(-0,05)}} = 0,49$

ΝΕΥΡΩΝΑΣ 3 (κρυφός)

Δυναμικό: $v_3 = w_{13} \cdot y_1 + w_{30} \cdot (-1) = -0,05$

Ενεργοποίηση: $y_3 = \varphi(v_3) = \frac{1}{1+e^{-(-0,05)}} = 0,49$

ΝΕΥΡΩΝΑΣ 4 (κρυφός)

Δυναμικό: $v_4 = w_{14} \cdot y_1 + w_{40}(-1) = 0$

Ενεργοποίηση: $y_4 = \varphi(v_4) = \frac{1}{1+e^{-0}} = 0,50$

ΝΕΥΡΩΝΑΣ 5 (εξόδου)

Δυναμικό: $v_5 = w_{25} \cdot y_2 + w_{35} \cdot y_3 + w_{45} \cdot y_4 + w_{50}(-1) = 0,55$

Ενεργοποίηση: $y_5 = \varphi(v_5) = v_5 = 0,55$

Σφάλμα στο νευρ. 5

$e_5 = d_5 - y_5 = -0,30$

ΠΡΟΣ ΤΑ ΠΙΣΩ ΠΕΡΑΣΜΑ

Νευρώνας 5 (εξόδου)

Τοπική κλίση: $\delta_5 = e_5 \cdot \varphi'(v_5) = e_5 \cdot 1 = -0,30$

Διορθ. στα βάρη: $\Delta w_{25} = \eta \cdot y_2 \cdot \delta_5 = -0,03$ | Νέα βάρη $w_{25} = w_{25} + \Delta w_{25} = 0,27$

$\Delta w_{35} = \eta \cdot y_3 \cdot \delta_5 = -0,03$ | $w_{35} = w_{35} + \Delta w_{35} = 0,79$

$\Delta w_{45} = \eta \cdot y_4 \cdot \delta_5 = -0,03$ | $w_{45} = w_{45} + \Delta w_{45} = 0,17$

$\Delta w_{50} = \eta \cdot (-1) \cdot \delta_5 = 0,06$ | $w_{50} = w_{50} + \Delta w_{30} = 0,16$

Νευρώνας 4 (κρυφός)

Τοπική κλίση: $\delta_4 = \varphi'(v_4) \cdot [w_{45} \cdot \delta_5] = y_4(1-y_4) \cdot w_{25} \cdot \delta_5 = -0,02$

Διορθ. στα βάρη: $\Delta w_{14} = \eta \cdot y_1 \cdot \delta_4 = 0$ | $w_{14} = w_{14} + \Delta w_{14} = 0,20$

$\Delta w_{40} = \eta \cdot (-1) \cdot \delta_4 = 0$ | $w_{40} = w_{40} + \Delta w_{40} = 0,10$

Νευρώνας 3 (κρυφός)

Τοπική κλίση: $\delta_3 = \varphi'(v_3)[w_{35} \cdot \delta_5] = y_3(1-y_3) \cdot w_{35} \cdot \delta_5 = -0,06$

Διορθ. στα βάρη: $\Delta w_{13} = \eta \cdot y_1 \cdot \delta_3 = -0,01$ | $w_{13} = w_{13} + \Delta w_{13} = 0,09$

$\Delta w_{30} = \eta \cdot (-1) \cdot \delta_3 = 0,01$ | $w_{30} = w_{30} + \Delta w_{30} = 0,11$

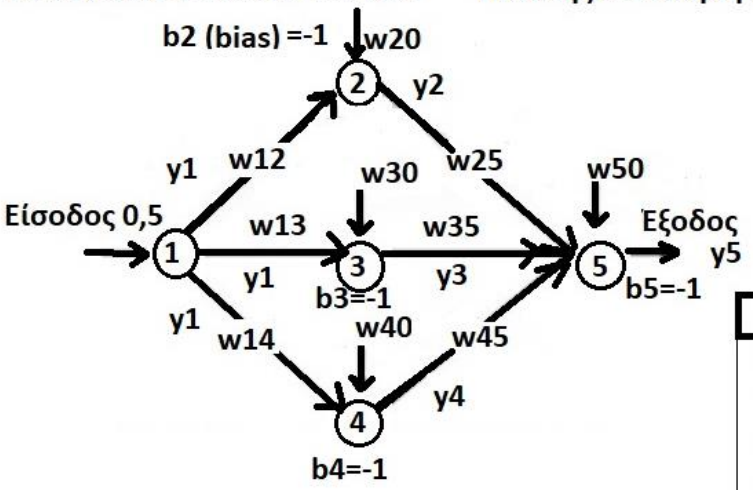
Νευρώνας 2 (κρυφός)

Τοπική κλίση: $\delta_2 = \varphi'(v_2)[w_{25} \cdot \delta_5] = y_2(1-y_2) \cdot w_{25} \cdot \delta_5 = -0,02$

Διορθ. στα βάρη: $\Delta w_{12} = \eta \cdot y_1 \cdot \delta_2 = 0$ | $w_{12} = 0,10$

$\Delta w_{20} = \eta \cdot (-1) \cdot \delta_2 = 0$ | $w_{20} = 0,10$

Το b1 (bias) ή αλλιώς κατώφλι Ηλιάδης Λ. Άσκηση 1 ΒΡ



Δίνεται το πιο κάτω νευρωνικό δίκτυο όπου η Έξοδος πρέπει να είναι το τετράγωνο της εισόδου.

Οι νευρώνες εισόδου έχουν συνάρτηση ενεργοποίησης την Σιγμοειδή και ο νευρώνας εξόδου την Φ(χ)-> Relu (γραμμική). Όπως φαίνεται τα κατώφλια bias b2 ,b3,b4 έχουν την τιμή -1.

Οι τιμές των συντελεστών βαρύτητας είναι:
w12= 0,10 w13=0,10 w14=0,20 w25=0,30 w35=0,82 w45=0,20 w20=0,10 w30=0,10 w40=0,10 w50=0,10

ΑΣΚΗΣΗ:

α) Να υπολογίσετε το σφάλμα στην έξοδο δεδομένου ότι η είσοδος είναι 0,5

β) Με δεδομένο ότι χρησιμοποιείται back Propagation να βρείτε τις ανανεωμένες τιμές των συντελεστών βαρύτητας μετά από μια ανάστροφη μετάδοση σφάλματος. Ο ρυθμός μάθησης (Learning Rate) είναι ίσος με 0,2. Όπου y1, η είσοδος που μεταβιβάζεται στους νευρώνες 2,3,4, και y2 ,y3,y4 οι τιμές που μεταφέρονται στον νευρώνα 5

y5 είναι η έξοδος

Η Σιγμοειδής είναι: $S(x) = \frac{1}{1+e^{-x}}$

ΥΠΟΔΕΙΞΗ:
Νευρώνας1: y1=0,5
Νευρώνας 2:
ΔΥΝΑΜΙΚΟ V2=W12*Y1+bias*W20
Με την εφαρμογή της συνάρτησης ενεργοποίησης στην έξοδο μεταφέρεται από τον νευρώνα 2 προς τον νευρώνα 5 το y2=S(V2) όπου S η Σιγμοειδής, V2 το δυναμικό όπου S η πιο πάνω σιγμοειδής συνάρτηση

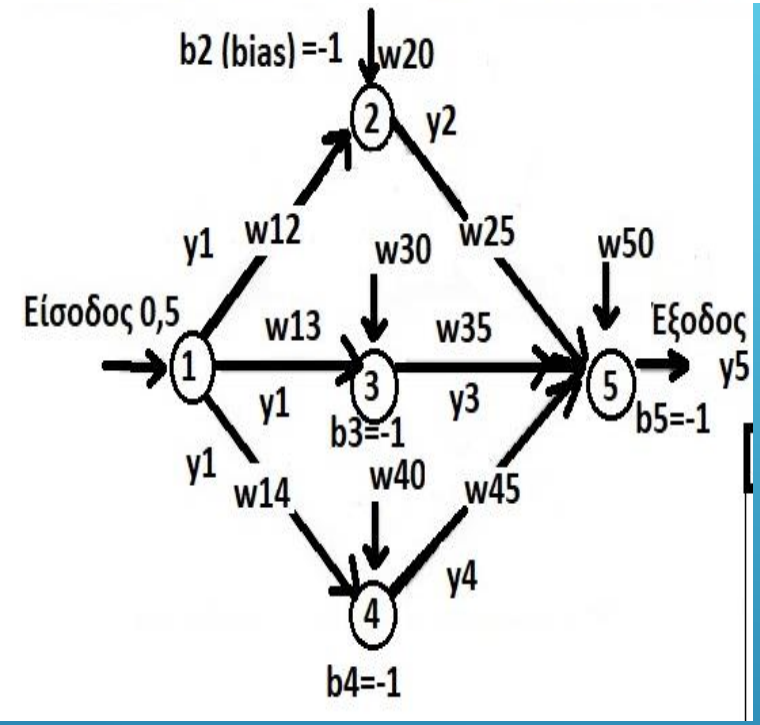
Σχετικά με το β σκέλος: Στην ανάστροφη μετάδοση σφάλματος όταν ο νευρώνας 5 θα μεταβιβάσει το σφάλμα προς τα κρυφό επίπεδο θα αλλάξουν τα βάρη w255,w35,w45,w50

Εδώ υπολογίζω την τοπική κλίση: δ5=error5 * Φ'(V5)

error5=Υπολογιζόμενη-πραγματική τιμή
Στη συνέχεια η διόρθωση στα βάρη θα είναι: ΔW25=a *Y2*δ5 όπου a το Learning Rate
Η διόρθωση στο bias w50 θα είναι: Δw50=a * (-1) * δ5
Έτσι θα βρείτε και τα υπόλοιπα νέα βάρη

Οι τιμές των συντελεστών βαρύτητας είναι:

w12= 0,10 w13=0,10 w14=0,20 w25=0,30 w35=0,82 w45=0,20 w20=0,10 w30=0,10 w40=0,10 w50=0,10



Νευρ1:Y1=X1=0,5 ΝΕΥR2:ΔΥΝΑΜΙΚΟ:V2=W12.Y1+W20.(-1) = - 0,05

ΕΝΕΡΓΟΠΟΙΗΣΗ: Y2=F(V2)=> Y2=1/[1+e^-(-0,05)] =0,49

Νευρ3: Δυναμικό V3=W13.Y1+W30.(-1)=-0,05

ΕΝΕΡΓΟΠΟΙΗΣΗ Y3=F(V3)=>Y3= Y2=1/[1+e^-(-0,05)] =0,49 αυτή η τιμή φεύγει από τον Ν3 προς τον Ν5.

Νευρ4:……..???

Νευρ5:……????……. Y5=0,55

ΝΕΥΡΩΝΑΣ5 ΔΥΝΑΜΙΚΟ V5=W25.Y2+W35.Y3+W45.Y4+w50.(-1)=0,55.

ΕΝΕΡΓΟΠΟΙΗΣΗ: Y5=F(V5)=> Y5=0,55

ΠΡΟΣ ΤΑ ΠΙΣΩ ΤΟΠΙΚΗ ΚΛΙΣΗ: δ5=e5.f'(V5)=e5.1=-0,30 LR α=0,2

Διόρθωση βαρών: ΔW25= α.Y2.δ5 = 0,2.0,49.(-0,30)=-0,03

ΔW35= α.Y3.δ5 = 0,2.0,49.(-0,30)=-0,03

ΔW45=α.Y4.δ5=0,2.Y4.(-0,3)=-0,03

ΔW50=α.(-1).δ5=0,2.(-1).(-0,3)=0,06

ΝΕΑ ΒΑΡΗ W25=W25+ΔW25=???

W35=W35+ΔW35=????

W45=W45+ΔW45=????

W50=W50+ΔW50=????