ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ	
Θεματική Ενότητα	**ΠΛΗ 31: Τεχνητή Νοημοσύνη - Εφαρμογές**
Ακαδημαϊκό Έτος	**2021 – 2022**
Γραπτή Εργασία	**#2**
Ημερομηνία Παράδοσης	**Παρασκευή, 18.02.2022, ώρα 23:59:59**

Παρατηρήσεις

Η καταληκτική ημερομηνία υποβολής της εργασίας είναι η Τετάρτη, *23 Φεβρουαρίου 2022*, ώρα 23:59:59, μετά από την οποία οποιαδήποτε υποβολή εργασίας θεωρείται ως εκπρόθεσμη.

Η παραπάνω ημερομηνία συνυπολογίζει την καθολική προκαταβολική έγκριση τυχόν αιτήματος παράτασης, κατά τον κανονισμό του ΕΑΠ (https://www.eap.gr/wp-content/uploads/2020/11/eap_kanonismos_spoydwn.pdf), με την υπόθεση πως η αρχικά οριζόμενη ημερομηνία παράδοσης θα ήταν η ακριβώς προηγούμενη Παρασκευή.

Την *Παρασκευή, 25 Φεβρουαρίου 2022, ώρα 11:59:59* (το αργότερο), θα δημοσιευθεί ενδεικτική επίλυση της εργασίας στο διαδίκτυο.

Περιμένουμε όλες οι εργασίες να υποβληθούν μέσω του συστήματος study. Πρέπει να υποβάλετε ΕΝΑ μόνο αρχείο (συμπιεσμένο), στο οποίο θα περιλαμβάνονται το πρωτότυπο κείμενο, σε οποιοδήποτε επεξεργαστή κειμένου έχετε χρησιμοποιήσει, και το παραγόμενο PDF, καθώς και όποια άλλα συνοδευτικά αρχεία. Αν υποβάλετε τμήματα κώδικα, θα πρέπει να βρίσκονται σε ξεχωριστά αρχεία, μέσα στο συμπιεσμένο αρχείο, και θα πρέπει ν’ αναφέρονται στο κείμενο της εργασίας.

Αν υπάρχουν διευκρινιστικά ερωτήματα που θεωρείτε ότι είναι αναγκαία για την επίλυση της εργασίας και τα οποία δεν έχουν απαντηθεί, μπορείτε στην επίλυσή σας να τα αναφέρετε μαζί με τις παραδοχές που θα κάνετε και τις οποίες θα πρέπει να ακολουθήσετε στην επίλυσή σας.

ΘΕΜΑ	ΜΟΝΑΔΕΣ
ΠΕ1.Α	10
ΠΕ1.Β	10
1ο	15 (4 + 7 + 2+2)
2ο	20 (4 + 4 + 4 + 4 + 4)
3ο	20 (4 + 2 + 9 + 3 + 2)
4ο	15 (3+3 + 6 + 3)
5ο	15 (3 + 2 + 2 + 4 + 4)
ΣΥΝΟΛΟ	**105**

Θέμα 1: Δίκτυα Hopfield [*15 μονάδες*]

Δύο παραδείγματα x=(x_1, x_2, …, x_n) και y=(y_1, y_2, …, y_n) ονομάζονται *ορθογώνια* εάν το εσωτερικό τους γινόμενο είναι μηδέν, δηλ. $\sum_{j=1}^{n} x_j y_j$=0.

Έστω ότι σε ένα δίκτυο Hopfield με n νευρώνες (n>1) *αποθηκεύουμε δύο ορθογώνια παραδείγματα* x=(x_1, x_2, …, x_n) και y=(y_1, y_2, …, y_n) (όπου τα x_i και τα y_i είναι 1 ή -1) χρησιμοποιώντας τον κανόνα Hebb.

Α. [*4 μονάδες*] Χρησιμοποιώντας την ιδιότητα της ορθογωνιότητας, να αποδείξετε ότι τα παραδείγματα x και y θα γίνουν καταστάσεις ισορροπίας του δικτύου. (Σημείωση: Για να είναι μια κατάσταση x=(x_1,x_2, …, x_n) κατάσταση ισορροπίας θα πρέπει:I_i= $x_i u_i \geq 0$, για κάθε νευρώνα i, όπου $u_i = \sum_{j=1}^{n} w_{ij} x_j + w_{i0}$). Να παραθέσετε την απόδειξη μόνο για το παράδειγμα x.

Β. [*7 μονάδες*] Να δείξετε ότι η ενέργεια του παραπάνω δικτύου Hopfield είναι ίδια και στις δύο καταστάσεις ισορροπίας x και y και ίση με $-n^2/2 + n$. (Να παραθέσετε την απόδειξη μόνο για την κατάσταση ισορροπίας y).

Γ. [*4 μονάδες*] Έστω ότι χρησιμοποιώντας τον κανόνα Hebb αποθηκεύουμε σε ένα δίκτυο Hopfield με 4 νευρώνες τα *ορθογώνια* παραδείγματα x=(1,-1,1,-1) και y=(-1,1,1,-1).

Γ1. [*2 μονάδες*] Να επιβεβαιώσετε ότι το x έγινε κατάσταση ισορροπίας του δικτύου συμπληρώνοντας τον παρακάτω πίνακα.

x_i	$u_i = \sum_{\substack{j=1 \\ j \neq i}}^{4} w_{ij} x_j + w_{i0}$	$x_i u_i \geq 0$
x_1=1	2	Ισχύει
x_2=-1		
x_3=1		
x_4=-1		

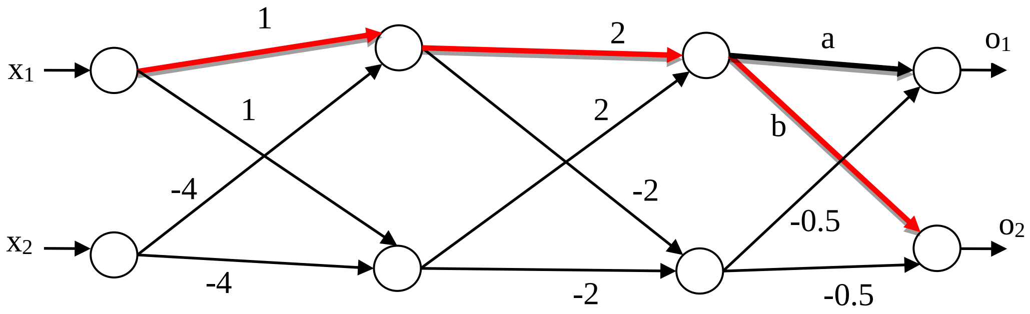
Γ2. [*2 μονάδες*] Χρησιμοποιώντας τον τύπο της ενέργειας ενός δικτύου Hopfield, να υπολογίσετε την ενέργεια της κατάστασης y και να επαληθεύσετε τον τύπο που αποδείξατε στο θέμα Β.

<u>Απάντηση:</u>

Δεν απαντήθηκε αυτό το ερώτημα

Θέμα 2: Εκπαίδευση σε MLP [*20 μονάδες*]

Θεωρούμε το MLP του παρακάτω σχήματος με 2 εισόδους, 2 εξόδους με γραμμική συνάρτηση ενεργοποίησης, πρώτο κρυμμένο επίπεδο με 2 νευρώνες με λογιστική συνάρτηση ενεργοποίησης και δεύτερο κρυμμένο επίπεδο με 2 νευρώνες με relu συνάρτηση ενεργοποίησης. Η συνάρτηση relu (rectified linear unit) ορίζεται ως g(u)=max(0,u) και χρησιμοποιείται συχνά σε δίκτυα βαθιάς μάθησης (deep neural networks) τα οποία έχουν πολλά κρυμμένα επίπεδα.



Για τα βάρη και τις πολώσεις χρησιμοποιούμε το συμβολισμό του Τόμου Β, ενότητα 4.2.1, όπως συνοψίζονται παρακάτω:

θα χρησιμοποιούμε το συμβολισμό i^ℓ για να αναφερόμαστε στο νευρώνα i του ℓ επιπέδου. Κατά συνέπεια, ορίζουμε:

- $u_i^{(\ell)}$ τη συνολική είσοδο στο νευρώνα i^ℓ
- $y_i^{(\ell)}$ την έξοδο του νευρώνα i^ℓ
- $\delta_i^{(\ell)}$ το σφάλμα του νευρώνα i^ℓ
- $w_{i0}^{(\ell)}$ την πόλωση του νευρώνα i^ℓ
- g_ℓ τη συνάρτηση ενεργοποίησης των νευρώνων στο επίπεδο ℓ
- d_ℓ τον αριθμό των νευρώνων στο επίπεδο ℓ
- $w_{ij}^{(\ell)}$ το βάρος της σύνδεσης από το νευρώνα $j^{\ell-1}$ στο νευρώνα i^ℓ.

Τα επίπεδα αριθμούνται από την είσοδο προς την έξοδο και για $\ell = 0$ έχουμε το επίπεδο εισόδου. Αν $\mathbf{x} = (x_1, \ldots, x_d)^T$ είναι το διάνυσμα εισόδου, για τους νευρώνες στο επίπεδο εισόδου έχουμε $y_i^{(0)} = x_i,\ x_0 = 1$. Αν q είναι το επίπεδο εξόδου, τότε η έξοδος του MLP είναι $o_i(\mathbf{x}; \mathbf{w}) = y_i^{(q)}$.

Αρχικά θεωρούμε ότι:

- $w_{11}^{(1)}= w_{21}^{(1)}=1$ $w_{12}^{(1)}= w_{22}^{(1)}=-4$ $w_{10}^{(1)}= w_{20}^{(1)}=0$
- $w_{11}^{(2)}= w_{12}^{(2)}=2$ $w_{21}^{(2)}=w_{22}^{(2)}= -2$ $w_{10}^{(2)}= w_{20}^{(2)}=0$
- $w_{11}^{(3)}=a$ $w_{21}^{(3)}=b$ $w_{12}^{(3)}= w_{22}^{(3)}=-0.5$ $w_{10}^{(3)}= w_{20}^{(3)}=0$

Οι τιμές αυτές αναγράφονται πάνω στις συνδέσεις του σχήματος. Οι πολώσεις των νευρώνων δεν απεικονίζονται στο σχήμα.

Α. [*4 μονάδες*] Να βρεθούν οι τιμές των a και b για τις οποίες για είσοδο $x_1=2$, $x_2=0.5$ οι έξοδοι του δικτύου είναι $o_1=1$ και $o_2=-1$.

Απάντηση:

Γνωρίζοντας από την εκφώνηση ότι το πρώτο κρυφό επίπεδο έχει λογιστική συνάρτηση ενεργοποίησης, το δεύτερο κρυφό επίπεδο έχει συνάρτηση ενεργοποίησης relu και η έξοδος έχει ως συνάρτηση ενεργοποίησης τη γραμμική.

Επίσης οι πολώσεις των νευρώνων είναι 0.
Θα υπολογίσω αρχικά όλες τις τιμές των συνόλων εισόδων στους νευρώνες σε όλα τα κρυμμένα επίπεδα όπως και το επίπεδο της εξόδου κάνοντας ουσιαστικά ένα ευθύ πέρασμα (forward pass) χωρίς τον υπολογισμό του σφάλματος εκαπίδευσης.

Νευρώνας πρώτου κρυμμένου επιπέδου με συνάρτηση ενεργοποίησης τη λογιστική:
$u_1^{(1)}= \sum_{i=1}^{d} wixi + w0 = 2*1 + 0.5*(-4) + 0 = 0$ $\quad y_1^{(1)} = \sigma(0) = \frac{1}{1+e^{-a0}} = \frac{1}{1+e^{-0}} = \frac{1}{2} = 0.5$

$u_2^{(1)}= \sum_{i=1}^{d} wixi + w0 = 2*1 + 0.5*(-4) + 0 = 0$ $\quad y_2^{(1)} = \sigma(0) = \frac{1}{1+e^{-a0}} = \frac{1}{1+e^{-0}} = \frac{1}{2} = 0.5$

Νευρώνας δεύτερου κρυμμένου επιπέδου με συνάρτηση ενεργοποίησης relu:
$u_1^{(2)}= \sum_{i=1}^{d} wixi + w0 = 0.5*2 + 0.5*2 + 0 = 2$ $\quad y_1^{(2)} = g(u_1) = \max(0,2) = 2$

$u_2^{(2)}= \sum_{i=1}^{n} wixi + w0 = 0.5*(-2) + 0.5*(-2) + 0 = -2$ $\quad y_2^{(2)} = g(u_2) = \max(0,-2) = 0$

Νευρώνας εξόδου με γραμμική συνάρτηση ενεργοποίησης:
$u_1^{(3)}= \sum_{i=1}^{d} wixi + w0 = 2 * a + 0 * (-0.5) + 0 = 2a$ $\quad y_1^{(3)} = g(u_2) = \text{linear}(2a)=2a$

επειδή γνωρίζω τον στόχο target ($t_1 = 1$) από την εκφώνηση της άσκησης θα διαιρέσω με τον συντελεστή του αγνώστου:

$$2a = 1 \leftrightarrow \frac{2a}{2} = \frac{1}{2} \leftrightarrow a = 0.5$$

Σύμφωνα με τη γραμμική συνάρτηση ενεργοποίησης: $y_1^{(3)} = g(u_1) = \text{linear}(0.5) = 0.5$

$u_2^{(3)}= \sum_{i=1}^{n} wixi + w0 = 2 * b + 0 * (-0.5) + 0 = 2b$ $\quad y_2^{(3)} = g(u_2) = \text{linear}(2b)=2b$

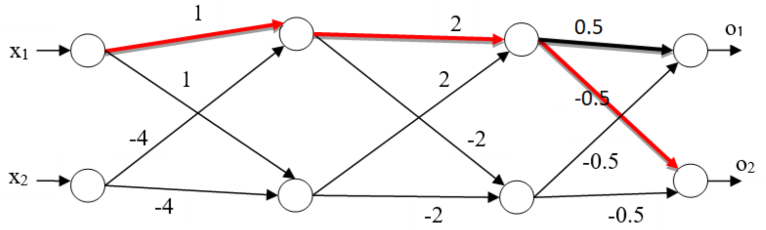
επειδή γνωρίζω τον στόχο target ($t_2 = -1$) από την εκφώνηση της άσκησης θα διαιρέσω με τον συντελεστή του αγνώστου:

$$2a = -1 \leftrightarrow \frac{2a}{2} = \frac{-1}{2} \leftrightarrow a = -0.5$$

Σύμφωνα με τη γραμμική συνάρτηση ενεργοποίησης: $y_2^{(3)} = g(u_2) = \text{linear}(-0.5) = -0.5$

B. [*4 μονάδες*] Θεωρείστε το δίκτυο με τα βάρη του ερωτήματος Α. Εστω ότι για εισόδους x_1=1, x_2=0.25 οι επιθυμητές έξοδοι είναι t_1=0.5, t_2=-0.5. Να υπολογίσετε το σφάλμα εκπαίδευσης για το συγκεκριμένο ζεύγος εισόδων-εξόδων.

Πρόσθεσα στο σχήμα της άσκησης τα υπολογισμένα βάρη του προηγούμενου ερωτήματος:



<u>Απάντηση:</u>

<u>Υπολογισμός ευθύς περάσματος (forward pass):</u>

<u>Νευρώνας πρώτου κρυμμένου επιπέδου με συνάρτηση ενεργοποίησης τη λογιστική:</u>

$u_1^{(1)} = \sum_{i=1}^{d} wixi + w0$ = 1*1 + 0.25*(-4) + 0 = 0 $y_1^{(1)} = \sigma(0) = \frac{1}{1+e^{-a0}} = \frac{1}{1+e^{-0}} = \frac{1}{2} = 0.5$

$u_2^{(1)} = \sum_{i=1}^{d} wixi + w0$ = 1*1 + 0.5*(-4) + 0 = 0 $y_2^{(1)} = \sigma(0) = \frac{1}{1+e^{-a0}} = \frac{1}{1+e^{-0}} = \frac{1}{2} = 0.5$

<u>Νευρώνας δεύτερου κρυμμένου επιπέδου με συνάρτηση ενεργοποίησης relu:</u>

$u_1^{(2)} = \sum_{i=1}^{d} wixi + w0$ = 0.5*2 + 0.5*2 + 0 = 2 $y_1^{(2)} = g(u_1) = \max(0,2) = 2$

$u_2^{(2)} = \sum_{i=1}^{n} wixi + w0$ = 0.5*(-2) + 0.5*(-2) + 0 = -2 $y_2^{(2)} = g(u_2) = \max(0,-2) = 0$

<u>Νευρώνας εξόδου με γραμμική συνάρτηση ενεργοποίησης:</u>

$u_1^{(3)} = \sum_{i=1}^{d} wixi + w0$ = 2 * 0.5 + 0 * (-0.5) + 0 = 1 $y_1^{(3)} = g(u_2) = linear(1) = 1$

άρα: o_1 = 1 είναι διάφορο του target $t_1^{(3)}$ = 0.5 υπάρχει σφάλμα και βάση του κανόνα υπολογισμού σφάλματος

$\delta_1^{(3)} = t_1^{(3)} - o_1 = 0.5 – 1 = - 0.5$

$u_2^{(3)} = \sum_{i=1}^{n} wixi + w0$ = 2 * (-0.5) + 0 * (-0.5) + 0 = -1 $y_2^{(3)} = g(u_2) = linear(-1) = -1$
άρα: o_2 = -1 είναι διάφορο του target $t_2^{(3)}$ = -0.5 υπάρχει σφάλμα και βάση του κανόνα υπολογισμού σφάλματος
$\delta_2^{(3)} = t_2^{(3)} – o_2 = (-0.5) – (-1) = 0.5$

Γ. [*4 μονάδες*] Να υπολογίσετε τις μερικές παραγώγους του τετραγωνικού σφάλματος εκπαίδευσης ως προς **τα βάρη των συνδέσεων που απεικονίζονται με κόκκινο χρώμα**, δηλαδή τα βάρη $w_{11}^{(1)}$, $w_{11}^{(2)}$ και $w_{21}^{(3)}$.

Απάντηση:

Η Μερική παράγωγος = σφάλμα προορισμού x έξοδος πηγής $\frac{\partial E^n}{\partial w_{ij}^{(h)}} = \delta_i^{(h)} y_j^{(h-1)}$.

$\delta_i^{(H+1)} = (o_i - t_{ni})$, i=1,...,p (γραμμική συν. ενεργοποίησης)

$\delta_i^{(H+1)} = o_i(1-o_i)(o_i - t_{ni})$, i=1,...,p (λογιστική συν. ενεργοποίησης)

Υπολογισμός σφάλματος:

$\delta_1^{(3)} = (o_1-t_1) = (-1) - (-0.5) = 0.5$

$\delta_2^{(3)} = (o_2-t_2) = (-1) - (-0.5) = -0.5$

$\delta_1^{(2)} = g_1^{(2)}{}' * (\delta_1^{(3)}*w_{11}^{(3)} + \delta_2^{(3)}*w_{21}^{(3)} = 1 * (0.5*0.5+(-0.5)*(-0.5)) = 0.5$

$\delta_2^{(2)} = g_2^{(2)}{}' * (\delta_1^{(3)}*w_{21}^{(3)} + \delta_2^{(3)}*w_{22}^{(3)} = 0 * (0.5*(-0.5)+(-0.5)*(-0.5)) = 0$

$\delta_1^{(1)} = y_1^{(1)}*(1- y_1^{(1)})*((w_{11}^{(2)}*\delta 1^{(2)})+(w_{12}^{(2)}*\delta 2^{(2)})) = 0.5*(1-0.5)*((2*0.5)+((-2)*0)) = 0.25$

$\delta_2^{(1)} = y_2^{(1)}*(1- y_2^{(1)})* ((w_{21}^{(2)}*\delta 1^{(2)})+(w_{22}^{(2)}*\delta 2^{(2)})) = 0.5*(1-0.5)* ((2*0.5)+((-2)*0)) = 0.25$

άρα υπολογισμός των μερικών παραγώγων για τα μαρκαρισμένα βάρη είναι:

$\mathbf{w_{11}^{(1)} = \delta_1^{(1)} * y_1^{(0)} = \delta_1^{(1)} * x_1 = 0.25 * 1 = 0.25}$

$\mathbf{w_{11}^{(2)} = \delta_1^{(2)} * y_1^{(1)} = 0.5 * 0.5 = 0.25}$

$\mathbf{w_{21}^{(3)} = \delta_2^{(3)} * y_1^{(2)} = 2 * (-0.5) = -1}$

Δ. [*4 μονάδες*] Εφαρμόζοντας gradient descent (εξίσωση (4.18) Τόμος Β) με ρυθμό μάθησης η=0.2, **να υπολογίσετε τις νέες τιμές των βαρών μόνο για τις συνδέσεις που απεικονίζονται με κόκκινο χρώμα.** Τα βάρη των υπόλοιπων συνδέσεων και οι πολώσεις παραμένουν αμετάβλητες.

Απάντηση:

Για ρυθμό εκπαίδευσης η=0.2 οι τιμές των βαρών $w_{11}^{(1)}$, $w_{11}^{(2)}$, $w_{21}^{(3)}$που προκύπτουν από την εξίσωση ενημέρωσης gradient descent (4.18) είναι: $w_i(\tau+1)=w_i(\tau)\text{-}n\frac{\partial E^n}{\partial w_i}$, i=1,..,L βάρος μείον ρυθμό εκπαίδευσης επί μερική παράγωγος.

Βάρη:

$w_{11}^{(1)} = w_{11}^{(1)}\text{-}\eta * \frac{\partial E^n}{\partial w_i} = 1 - 0.2 * 0.25 = 0.95$

$w_{11}^{(2)} = w_{11}^{(2)}\text{-}\eta * \frac{\partial E^n}{\partial w_i} = 2- 0.2 * 0.25 = 1.95$

$w_{21}^{(3)} = w_{21}^{(3)}\text{-}\eta * \frac{\partial E^n}{\partial w_i} = (-0.5) - 0.2 * (-1) = - 0.3$

Ε. [*4 μονάδες*] Για τις νέες τιμές των βαρών να υπολογίσετε τη νέα τιμή του σφάλματος εκπαίδευσης. Έχει προκύψει μείωση του σφάλματος;

Απάντηση:

Υπολογισμός ευθύς περάσματος (forward pass):

Νευρώνας πρώτου κρυμμένου επιπέδου με συνάρτηση ενεργοποίησης τη λογιστική:

$u_1^{(1)}= \sum_{i=1}^{d} wixi + w0$ = 1*0.95 + 0.25*(-4) + 0 = -0.05 $y_1^{(1)} = \sigma(-0.05) = \frac{1}{1+e^{-a0}} = \frac{1}{1+e^{-0.05}} = 0.4875$

$u_2^{(1)}= \sum_{i=1}^{d} wixi + w0$ = 1*1 + 0.25*(-4) + 0 = 0 $y_2^{(1)} = \sigma(0) = \frac{1}{1+e^{-a0}} = \frac{1}{1+e^{-0}} = \frac{1}{2} = 0.5$

Νευρώνας δεύτερου κρυμμένου επιπέδου με συνάρτηση ενεργοποίησης relu:

$u_1^{(2)}= \sum_{i=1}^{d} wixi + w0$ = 0.4875 * 1.95 + 0.5*2 + 0 = 1.95 $y_1^{(2)}$ = g(u_1) = max(0,1.95) = 1.95

$u_2^{(2)}= \sum_{i=1}^{n} wixi + w0$ = 0.4875*(-2) + 0.5*(-2) + 0 = -1.975 $y_2^{(2)}$ = g(u_2) = max(0,-1.975) = 0

Νευρώνας εξόδου με γραμμική συνάρτηση ενεργοποίησης:

$u_1^{(3)}= \sum_{i=1}^{d} wixi + w0$ = 1.95 * 0.5 + 0 * (-0.5) + 0 = 0.975 $y_1^{(3)}$ = g(u_2) = linear(0.975) = 0.975

άρα: o_1 = 0.975 είναι διάφορο του target $t_1^{(3)}$ = 0.5 υπάρχει σφάλμα και βάση του κανόνα υπολογισμού σφάλματος

$\delta_1^{(3)} = t_1^{(3)} - o_1$ = 0.5 – 0.975 = - 0.475

$u_2^{(3)}= \sum_{i=1}^{n} wixi + w0$ = 1.95 * (-0.3) + 0 * (-0.5) + 0 = -0.585 $y_2^{(3)}$ = g(u_2) = linear(-0.585) = -0.585

άρα: o_2 = -0.585 είναι διάφορο του target $t_2^{(3)}$ = -0.5 υπάρχει σφάλμα και βάση του κανόνα υπολογισμού σφάλματος: $\delta_2^{(3)} = t_2^{(3)} – o_2$ = (-0.5) – (-0.585) = 0.085

Στο νέο δίκτυο βρίσκουμε έξοδο o=(-0.475,0.085) ενώ στο παλιό δίκτυο πριν τις μεταβολές βρίσκουμε o=(-0.5,0.5)

Βλέπουμε ότι υπάρχει μείωση του σφάλματος αλλά για Ως άθροισμα τετραγώνων έχουμε κάτω φράγμα την τιμή μηδέν η οποία προκύπτει όταν έχουμε τέλεια εκπαίδευση

$$E(w)=\frac{1}{2}\sum_{n=1}^{N}||t^n-o(x^n;w)||^2=\frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{p}(t_{nm}-o_m(x^n;w))^2$$

υπολογίζοντας το μέσο τετραγωνικό σφάλμα του παλιού δικτύου υπό ερωτήματος β:

$$E(w) = \frac{1}{2}((-0.5)^2+(0.5)^2)=0.25$$

υπολογίζοντας το μέσο τετραγωνικό σφάλμα του νέου δικτύου υπό ερωτήματος Ε:

$$E(w) = \frac{1}{2}((-0.475)^2+(0.085)^2)=0.116$$

Παρατηρώ ότι έχει προκύψει σημαντική μείωση του σφάλματος στο νέο δίκτυο.

Θέμα 3: ΤΝΔ με βηματικούς νευρώνες [*20 μονάδες*]

Στα επόμενα ερωτήματα θα χρησιμοποιήσουμε βηματικούς νευρώνες με d εισόδους $x_1,\ldots,x_d$, βάρη $w_1,\ldots,w_d$ και πόλωση w_0. Η ενεργοποίηση u του νευρώνα υπολογίζεται ως: $u=w_1x_1+\ldots+w_dx_d+w_0$. Η έξοδος o του νευρώνα παίρνει τιμές 0 ή 1 και η βηματική συνάρτηση ενεργοποίησης του νευρώνα ακολουθεί την παρακάτω σχέση:

$$o(u)=\begin{cases}1, & u\geq 0\\ 0, & u<0\end{cases}$$

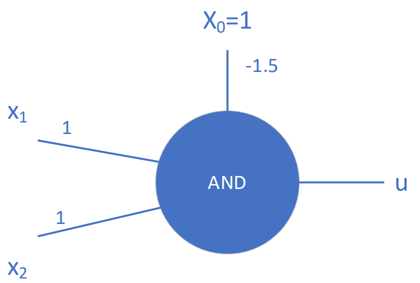
A. [*4 μονάδες*] Για καθεμιά από τις παρακάτω τέσσερις περιπτώσεις, να ορίσετε (εάν είναι εφικτό) τα βάρη w_1, w_2 και την πόλωση w_0 ενός βηματικού νευρώνα με δύο εισόδους x_1 και x_2 (που παίρνουν τιμές 0 ή 1) και έξοδο o οποίος υλοποιεί την αντίστοιχη λογική συνάρτηση:

A1. o=1 εάν x_1=1 AND x_2=1, αλλιώς o=0.

Απάντηση:

Έπειτα από την κατασκευή του ζητούμενου σε πίνακα αληθείας παρατηρώ ότι η έξοδος είναι όμοια με μια έξοδο πύλης AND η οποία είναι γραμμικά ανεξάρτητη και μπορεί να κατασκευαστεί με έναν βηματικό νευρώνα δύο εισόδων με πόλωση.

Ενδεικτική τιμή στα βάρη: w1 = 1 , w2 = 1 , w0 = -1.5

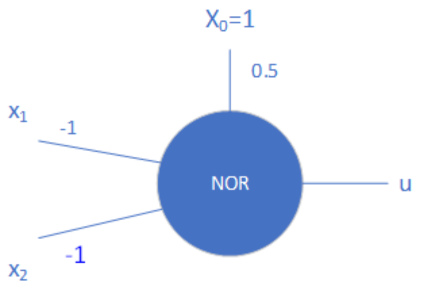


A2. o=1 εάν x_1=0 AND x_2=0, αλλιώς o=0.

Απάντηση:

Έπειτα από την κατασκευή του ζητούμενου σε πίνακα αληθείας παρατηρώ ότι η έξοδος είναι όμοια με μια έξοδο πύλης NOR η οποία είναι γραμμικά διαχωρίσιμη και μπορεί να κατασκευαστεί με έναν βηματικό νευρώνα δύο εισόδων με πόλωση.

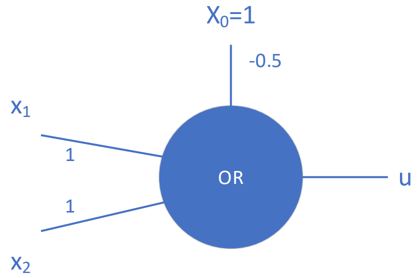
Ενδεικτική τιμή στα βάρη: w1 = -1 , w2 = -1 , w0 = +0.5



A3. o=1 εάν x_1=1 OR x_2=1, αλλιώς o=0.

Απάντηση:
Έπειτα από την κατασκευή του ζητούμενου σε πίνακα αληθείας παρατηρώ ότι η έξοδος είναι όμοια με μια έξοδο πύλης OR η οποία είναι γραμμικά διαχωρίσιμη και μπορεί να κατασκευαστεί με έναν βηματικό νευρώνα δύο εισόδων με πόλωση.

Ενδεικτική τιμή στα βάρη: w1 = 1 , w2 = 1 , w0 = -0.5



Α4. o=1 εάν $x_1 \neq x_2$, αλλιώς o=0.

Απάντηση:
Έπειτα από την κατασκευή του ζητούμενου σε πίνακα αληθείας παρατηρώ ότι η έξοδος είναι όμοια με μια έξοδο πύλης XOR η οποία δεν είναι γραμμικά διαχωρίσιμη και δεν μπορεί να κατασκευαστεί με έναν βηματικό νευρώνα αλλά με περισσότερους.

Β. [*2 μονάδες*] Να ορίσετε βάρη w_1, w_2 και την πόλωση w_0 ενός βηματικού νευρώνα με δύο εισόδους x_1 και x_2 και έξοδο o οποίος υλοποιεί την συνάρτηση:

o=1 εάν $\alpha x_1 + \beta x_2 + \gamma \geq 0$ αλλιώς o=0.

όπου α, β και γ είναι γνωστοί πραγματικοί αριθμοί.

Απάντηση:
Από την εξίσωση υπερεπιπέδου στο $\mathbb{R}^2$ που αναγράφεται στο κεφάλαιο 2.2.1 του τόμου Β στο $\sum_{i=1}^{d} w_i \cdot x_i + w_0 = 0$ για δύο εισόδους, το όριο απόφαση παίρνει τη μορφή μιας ευθείας γραμμής με εξίσωση ευθείας: $\alpha \cdot \chi_1 + \beta \cdot \chi_2 + \gamma = 0$ και με εξίσωση νευρώνα με πόλωση:
$w_1 \cdot x_1 + w_2 \cdot x_2 + w_0 = 0$ άρα τα βάρη των νευρώνων θα είναι: $w_1 = a, w_2 = \beta, w_0 = \gamma$

Γ. [*9 μονάδες*] Να κατασκευάσετε ένα νευρωνικό δίκτυο με βηματικούς νευρώνες το οποίο θα παίρνει ως είσοδο ένα παράδειγμα $x=(x_1,x_2)$ και θα το ταξινομεί σύμφωνα με τους παρακάτω κανόνες:

if ($x_1 \geq x_2$ AND $x_1 \leq 7$) THEN Class=C_1

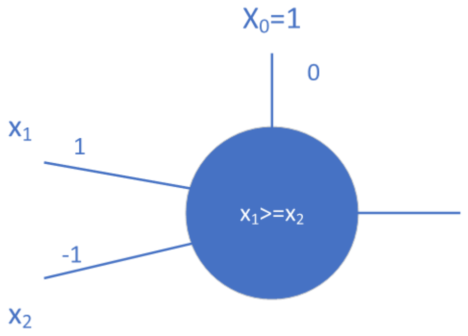
if ($x_1 \leq x_2 - 4$) THEN Class=C_2

else Class=C_3

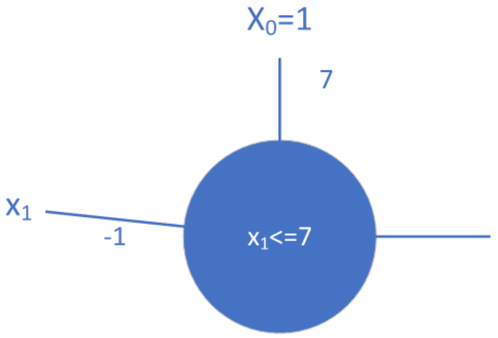
Απάντηση:

Η υλοποίηση του νευρωνικού δικτύου θα γίνει με βηματικούς νευρώνες με πολώσεις (+1)

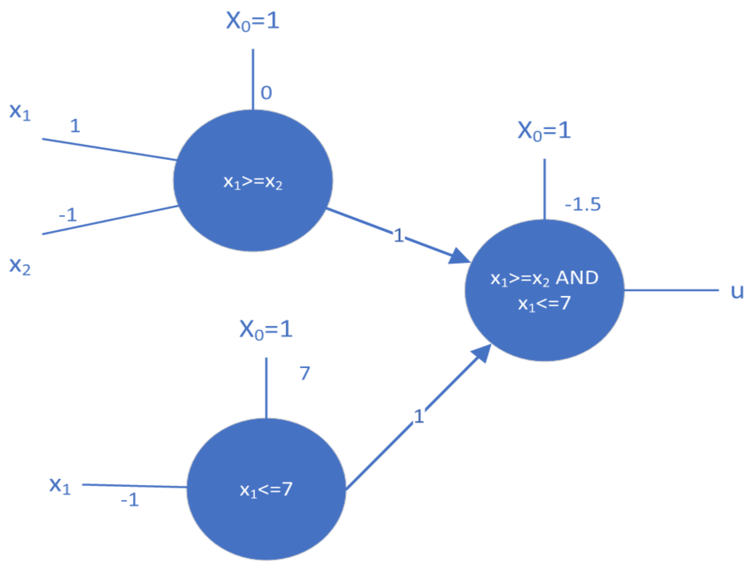
- Για πρώτο μέλος της πρώτης που ανισότητας που πρέπει να ικανοποιείται θα ελέγξω για ποιες τιμές εισόδων όταν το χ_1 είναι μεγαλύτερο του χ_2: $x_1 \geq x_2 \Leftrightarrow x1 - x2 \geq 0$ άρα η εξίσωση ευθείας του νευρώνα είναι: $w_1 \cdot x_1 + w_2 \cdot x_2 + w_0 = 0$ τότε: $1 \cdot x_1 + (-1) \cdot x_2 + 0 \cdot 1 = 0$



Για δεύτερο μέλος της πρώτης που ανισότητας που πρέπει να ικανοποιείται θα ελέγξω για ποιες τιμές εισόδων όταν το χ_1 είναι μικρότερο ή ίσο το 7. $\chi_1 \leq 7 \Leftrightarrow 0 \leq -\chi_1 + 7$. Υλοποιείται με έναν νευρώνα με μια είσοδο χ και πόλωση.

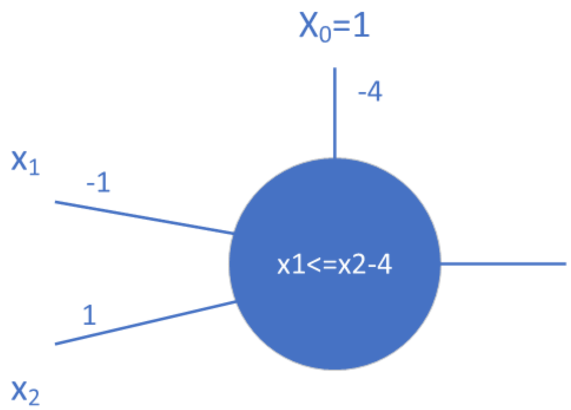


Με βάση τη σχέση AND που δίνεται στην εκφώνηση ώστε να είναι θετική η έξοδος θα πρέπει να είναι και οι δύο νευρώνες θετικοί, επομένως θα τις συνδέσω με έναν νευρώνα όπου θα υλοποιεί το λογικό AND και θα ενώσει τη σύνδεση μεταξύ των δύο αυτών ανισοτήτων.



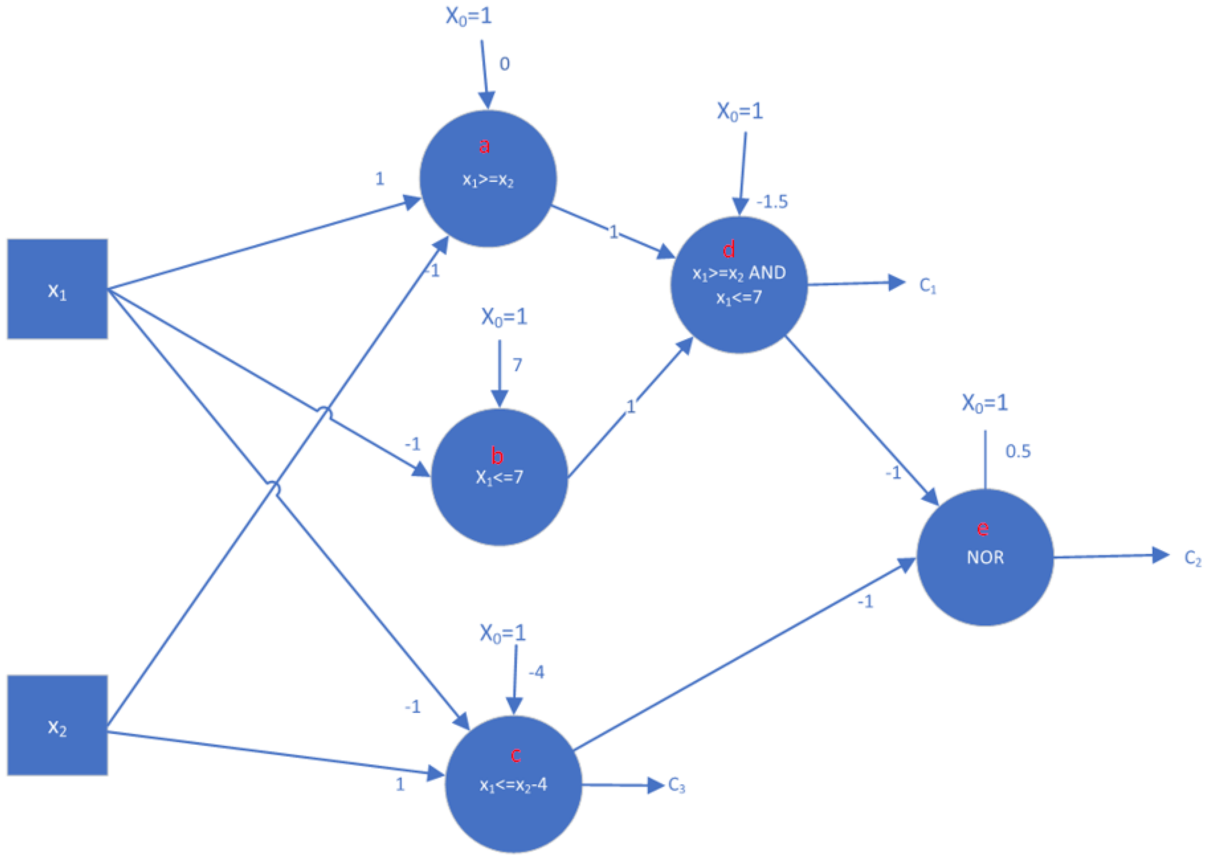
- Για δεύτερο μέλος της δεύτερης που ανισότητας που πρέπει να ικανοποιείται για το χ_1 να είναι το χ μικρότερο ή ίσο από το χ_2-4:

$x_1 \le x_2 - 4 \Leftrightarrow 0 \le -x_1 + x_2 - 4$ άρα η εξίσωση ευθείας του νευρώνα είναι: $w_1 \cdot x_1 + w_2 \cdot x_2 + w_0 = 0$ τότε: $(-1) \cdot x_1 + 1 \cdot x_2 + (-4) \cdot 1 = 0$



Δ. [*3 μονάδες*] Να σχεδιάσετε την αρχιτεκτονική του δικτύου και να αποτυπώσετε τα βάρη του.

Απάντηση:



Η λογική σε αυτό είναι όταν η ανισότητα $x_1 \geq x_2$ AND $x_1 \leq 7$ η έξοδος είναι θετική τότε βγαίνει το C1. Όταν η ανισότητα $x_1 \leq x_2 - 4$ η έξοδος είναι θετική τότε βγαίνει το C2.
Όταν $x_1 \geq x_2$ AND $x_1 \leq 7$ όπως και η ανισότητα $x_1 \leq x_2 - 4$ όπου ή η μια ανισότητα ή η άλλη ανισότητα ή και οι δύο ανισότητες βγάζουν αρνητική τιμή οι έξοδοι αυτών των νευρώνων συνδέονται με έναν νευρώνα που υλοποιεί μια πύλη NOR ώστε να παραχθεί η εναλλακτική έξοδος C3.

Ε. [*2 μονάδες*] Ποιες είναι οι έξοδοι του δικτύου που κατασκευάσατε εάν εφαρμόσουμε ως εισόδους τα παραδείγματα (2,1), (2,3) και (2,7);

Απάντηση:
Για το παράδειγμα $x_1 = 2$ και $x_2 = 1$ έχουμε

$$u(x) = \sum_{i=1}^{d} w_i \cdot x_i + w_0 u1$$

$$u(a) = 2 \cdot 1 + 1 \cdot (-1) + 0 = 1 \textit{άρα η έξοδος είναι } 1$$

$$u(b) = 2 \cdot (-1) + 7 = 5 \textit{ άρα η έξοδος είναι } 1$$

$$u(d) = 1 \cdot 1 + 1 \cdot 1 + 1 \cdot (-1.5) = 4.5 \textit{άρα η έξοδος είναι } 1$$

Το u(a)=1 AND u(b)=1 τότε u(d) =1 η έξοδος είναι C1.

Για το παράδειγμα x_1 = 2 και x_2 = 3 έχουμε

$$u(x) = \sum_{i=1}^{d} w_i \cdot x_i + w_0 u1$$

$$u(a) = 2 \cdot 1 + 3 \cdot (-1) + 0 = -1 \; \acute{\alpha}\rho\alpha\; \eta\; \acute{\varepsilon}\xi o\delta o\varsigma\; \varepsilon\acute{\iota}\nu\alpha\iota\; 0$$

$$u(b) = 2 \cdot (-1) + 7 = 5 \; \acute{\alpha}\rho\alpha\; \eta\; \acute{\varepsilon}\xi o\delta o\varsigma\; \varepsilon\acute{\iota}\nu\alpha\iota\; 1$$

$$u(d) = 1 \cdot 0 + 1 \cdot 1 + 1 \cdot (-1.5) = -0.5 \acute{\alpha}\rho\alpha\; \eta\; \acute{\varepsilon}\xi o\delta o\varsigma\; \varepsilon\acute{\iota}\nu\alpha\iota 0$$

$$u(c) = 2 \cdot (-1) + 3 \cdot 1 + 1 \cdot (-4) = -3 \acute{\alpha}\rho\alpha\; \eta\; \acute{\varepsilon}\xi o\delta o\varsigma\; \varepsilon\acute{\iota}\nu\alpha\iota 0$$

$$u(e) = 0(-1) + 0(-1) + 1 \cdot 0.5 = 5 \acute{\alpha}\rho\alpha\; \eta\; \acute{\varepsilon}\xi o\delta o\varsigma\; \varepsilon\acute{\iota}\nu\alpha\iota 1$$

Το u(a)=0 AND u(b)=0 τότε u(d)=0. Επειδή και το u(c)=0 τότε πηγαίνει στον νευρώνα u(e) όπου η έξοδος είναι 1 άρα ή έξοδος είναι C2.

Για το παράδειγμα x_1 = 2 και x_2 = 7 έχουμε

$$u(x) = \sum_{i=1}^{d} w_i \cdot x_i + w_0 u1$$

$$u(a) = 2 \cdot 1 + 7 \cdot (-1) + 0 = -5 \; \acute{\alpha}\rho\alpha\; \eta\; \acute{\varepsilon}\xi o\delta o\varsigma\; \varepsilon\acute{\iota}\nu\alpha\iota\; 0$$

$$u(b) = 2 \cdot (-1) + 7 = 5 \; \acute{\alpha}\rho\alpha\; \eta\; \acute{\varepsilon}\xi o\delta o\varsigma\; \varepsilon\acute{\iota}\nu\alpha\iota\; 1$$

$$u(d) = 1 \cdot 0 + 1 \cdot 1 + 1 \cdot (-1.5) = -0.5 \acute{\alpha}\rho\alpha\; \eta\; \acute{\varepsilon}\xi o\delta o\varsigma\; \varepsilon\acute{\iota}\nu\alpha\iota 0$$

$$u(c) = 2 \cdot (-1) + 7 \cdot 1 + 1 \cdot (-4) = 1 \acute{\alpha}\rho\alpha\; \eta\; \acute{\varepsilon}\xi o\delta o\varsigma\; \varepsilon\acute{\iota}\nu\alpha\iota\; 1$$

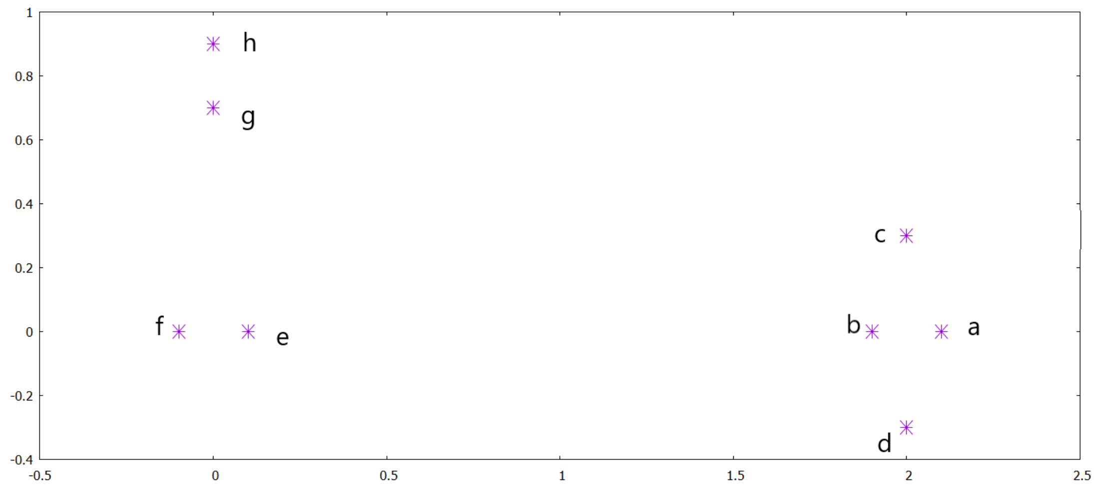
$$u(e) = 0(-1) + 1(-1) + 1 \cdot 0.5 = -0.5 \acute{\alpha}\rho\alpha\; \eta\; \acute{\varepsilon}\xi o\delta o\varsigma\; \varepsilon\acute{\iota}\nu\alpha\iota 0$$

Το u(a)=0 AND u(b)=0 τότε u(d)=0. Επειδή και το u(c)=1 τότε ή έξοδος είναι C3.

Θέμα 4: Ο αλγόριθμος *k*-means [*15 μονάδες*]

Δίνεται το σύνολο οκτώ παραδειγμάτων X={a,b,c,d,e,f,g,h} διάστασης d=2, των οποίων οι συνιστώσες (x_1, x_2) ορίζονται στο παρακάτω πίνακα:

	x_1	x_2
a	2.1	0
b	1.9	0
c	2	0.3
d	2	-0.3
e	0.1	0
f	-0.1	0
g	0	0.7
h	0	0.9



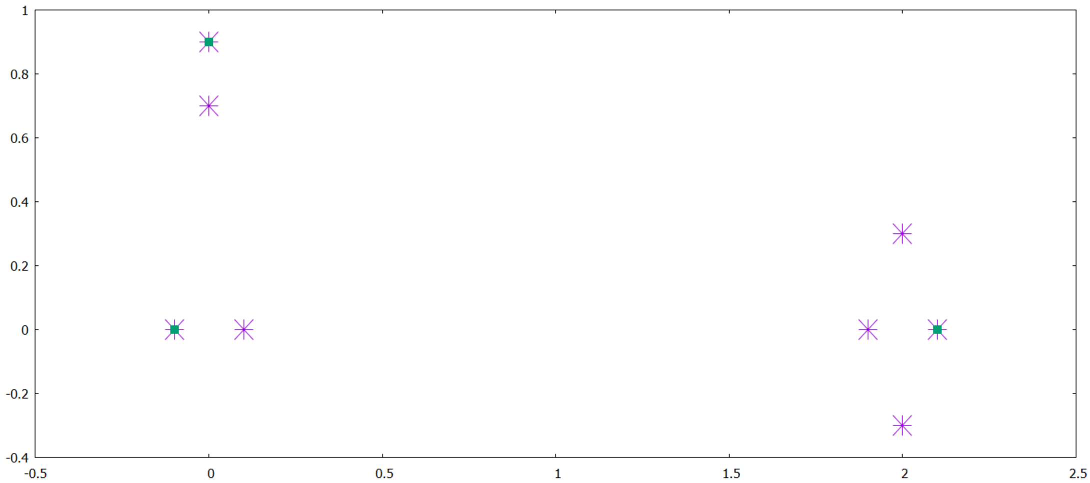
Μπορείτε να παρατηρήσετε οτι τα παραδείγματα ομαδοποιούνται σε 3 ομάδες: O_1^*={a,b,c,d}, O_2^*={e,f}, O_3^*={g,h}.

Θέλουμε να χρησιμοποιήσουμε τον αλγόριθμο k-means για να ομαδοποιήσουμε τα παραδείγματα του συνόλου X σε **τρεις** ομάδες O_1, O_2 και O_3. Ο αλγόριθμος k-means θεωρεί ότι κάθε ομάδα περιγράφεται από ένα κέντρο και λειτουργεί μεταβάλοντας τις θέσεις κέντρων $w_1=(w_{11}, w_{12})$, $w_2=(w_{21}, w_{22})$ και $w_3=(w_{31}, w_{32})$ των τριών ομάδων O_1, O_2 και O_3 αντίστοιχα.

Ιδανικά θα θέλαμε ο αλγόριθμος k-means να δώσει ως λύση τις ομάδες O_1^*, O_2^*, O_3^*.

Για να εκτελέσουμε τον αλγόριθμο k-means πρέπει πρώτα να καθορίσουμε τις αρχικές θέσεις των τριών κέντρων w_1, w_2 και w_3. Αυτό συνήθως γίνεται επιλέγοντας τυχαία τρία από τα παραδείγματα του συνόλου X.

A. [*6 μονάδες*] Εστω ότι η τυχαία επιλογή δίνει την αρχικοποίηση των κέντρων: w_1=a, w_2=f και w_3=h.



Ξεκινάμε την εκτέλεση του αλγορίθμου.

Α1. [*3 μονάδες*] **1η επανάληψη, βήμα 1:** Να κατατάξετε τα όλα τα παραδείγματα του συνόλου X σε μια από τις τρεις ομάδες O_1, O_2 και O_3 με βάση την απόσταση από το πλησιέστερο κέντρο. Ως απόσταση $d(x,w_i)$ ενός παραδείγματος $x=(x_1, x_2)$ από κάποιο κέντρο $w_i=(w_{i1}, w_{i2})$ να θεωρήσετε το τετράγωνο της Ευκλείδειας απόστασης: $d(x,w_i)=||x-w_i||^2=(x_1-w_{i1})^2+(x_2-w_{i2})^2$.

Να συμπληρώσετε τον παρακάτω πίνακα:

Παράδειγμα	**Πλησιέστερο κέντρο**	**Απόσταση από πλησιέστερο κέντρο**	**Ανάθεση στην ομάδα**	**Παράδειγμα**	**Πλησιέστερο κέντρο**	**Απόσταση από πλησιέστερο κέντρο**	**Ανάθεση στην ομάδα**
a	w_1	0	O_1	e	w_2	0.04	O_2
b	w_1	0.04	O_1	f	w_2	0	O_2
c	w_1	0.1	O_1	g	w_3	0.04	O_3
d	w_1	0.1	O_1	h	w_3	0	O_3

<u>Απάντηση:</u>

οι πράξεις εκτελέστηκαν ως εξής:

$$\gamma\iota\alpha\ \mathrm{O}_1=\{a,b,c,d\}\ \kappa\alpha\iota\ w_1=a$$

$$d(a,w1)\Rightarrow(2.1-2.1)^2+(0-0)^2=0$$

$$d(b,w1)\Rightarrow(1.9-2.1)^2+(0-0)^2=0.04$$

$$d(c,w1)\Rightarrow(2-2.1)^2+(0.3-0)^2=0.1$$

$$d(d,w1)\Rightarrow(2-2.1)^2+((-0.3)-0)^2=0.1$$

$$\gamma\iota\alpha\ \mathrm{O}_2=\{e,f\}\ \kappa\alpha\iota\ w_2=f$$

$$\gamma\iota\alpha\ d(e,w2)\Rightarrow(0.1-(-0.1))^2+(0-0)^2=0.04$$

$$\gamma\iota\alpha\ d(f,w2)\Rightarrow((-0.1)-(-0.1))^2+(0-0)^2=0$$

$$\gamma\iota\alpha\ \mathrm{O}_3=\{g,h\}\ \kappa\alpha\iota\ w_3=h$$

$$\gamma\iota\alpha\ d(g,w3)\Rightarrow(0-0)^2+(0.7-0.9)^2=0.04$$

$$\gamma\iota\alpha\ d(h,w3)\Rightarrow(0-0)^2+(0.9-0.9)^2=0$$

1η επανάληψη, βήμα 2: Με βάση την παραπάνω κατάταξη των παραδειγμάτων σε ομάδες, να υπολογίσετε τα **νέα κέντρα** των ομάδων ως τον μέσο όρο των παραδειγμάτων της ομάδας τους:

w_1		w_2		w_3	
w_{11}	w_{12}	w_{21}	w_{22}	w_{31}	w_{32}
2	0	0	0	0	0.8

Ο υπολογισμός έγινε ως εξής:

$$\Gamma\iota\alpha O_1\{a,b,c,d\}\ \tau\acute{o}\tau\varepsilon\ w_{11} = \frac{2.1+1.9+2+2}{4} = 2$$

$$\Gamma\iota\alpha O_1\{a,b,c,d\}\ \tau\acute{o}\tau\varepsilon\ w_{12} = \frac{0+0+0.3+(-0.3)}{4} = 0$$

$$\Gamma\iota\alpha O_2\{e,f\}\ \tau\acute{o}\tau\varepsilon\ w_{21} = \frac{0.1+(-0.1)}{2} = 0$$

$$\Gamma\iota\alpha O_2\{e,f\}\ \tau\acute{o}\tau\varepsilon\ w_{22} = \frac{0+0}{2} = 0$$

$$\Gamma\iota\alpha O_3\{g,h\}\ \tau\acute{o}\tau\varepsilon\ w_{31} = \frac{0+0}{2} = 0$$

$$\Gamma\iota\alpha O_3\{g,h\}\ \tau\acute{o}\tau\varepsilon\ w_{32} = \frac{0.7+0.9}{2} = 0.8$$

Ελεγχος τερματισμού: Έχουν αλλάξει τα κέντρα; Αν ναι συνεχίζουμε στην επόμενη επανάληψη, αλλιώς σταματάμε.

Απάντηση:
Βάση του υπολογισμού τα κέντρα είναι σε νέες συντεταγμένες σε σύγκριση από την αρχική αρχικοποίηση των κέντρων άρα συνεχίζουμε τον αλγόριθμο.

A2. [*3 μονάδες*] Ακολουθώντας το πρότυπο περιγραφής του ερωτήματος A1, να περιγράψετε τα επόμενα βήματα εκτέλεσης του αλγορίθμου μέχρι τον τερματισμό του.

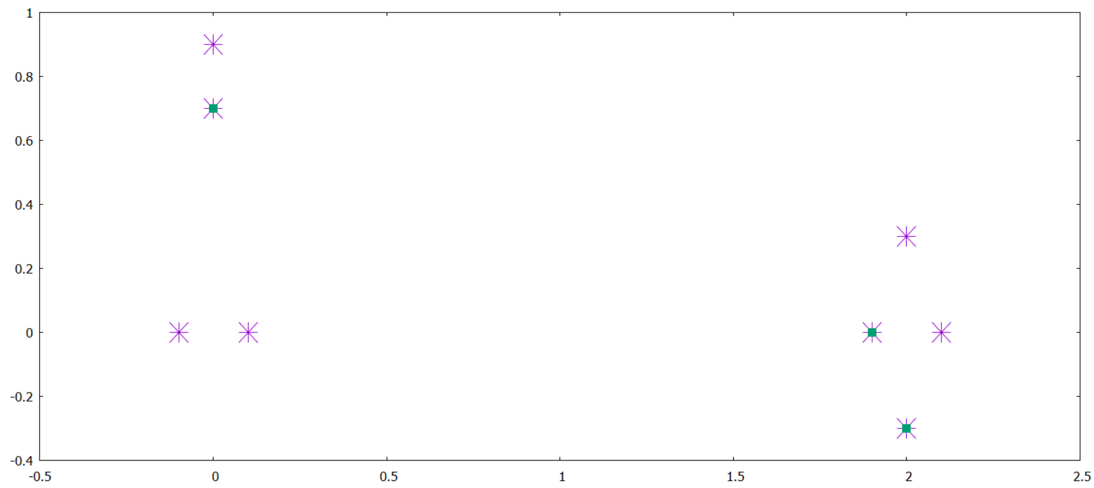
Απάντηση:
Με βάση τα νέα κέντρα ξανά υπολογίζω όλα τα παραδείγματα του συνόλου χ και για τις τριες ομάδες και με βάση την απόσταση τους από το κέντρο.

Στα επόμενα βήματα επανάληψης υπολογίζω την απόσταση του κάθε συνόλου από τα κέντρα και παρατηρώ ότι το κέντρο w_1 θα είναι στο κέντρο του παραδείγματος συνόλου χ στην ομαδοποίηση O_1 = {a,b,c,d}. Το κέντρο w_2 θα είναι στο κέντρο του παραδείγματος συνόλου χ στην ομαδοποίηση O_2 = {e,f} και τέλος το κέντρο w_1 θα είναι στο κέντρο του παραδείγματος συνόλου χ στην ομαδοποίηση O_3 = {g,h}.
Τέλος δεν αλλάζει το κέντρο άρα ο αλγόριθμος τερματίζει

Βρίσκει ο αλγόριθμος την ομαδοποίηση O_1^*, O_2^*, O_3^*;

<u>Απάντηση:</u>
Ο αλγόριθμος k-mean βρίσκει την ομαδοποίηση.

Β. [*6 μονάδες*] Έστω ότι η τυχαία επιλογή δίνει μια **διαφορετική αρχικοποίηση των κέντρων**: w_1=b, w_2=d και w_3=g.



Να περιγράψετε τα βήματα εκτέλεσης του αλγορίθμου μέχρι τον τερματισμό ακολουθώντας το πρότυπο περιγραφής του ερωτήματος Α. Βρίσκετε την ίδια λύση με το ερώτημα Α;

<u>Απάντηση:</u>

Αρχικοποιώντας τα κέντρα βάση της εκφώνησης της άσκησης, στο 1ο βήμα της πρώτης επανάληψης γίνεται ο υπολογισμός απόστασης των συνόλων από το πλησιέστερο κέντρο όπως αναγράφεται και στο υπό ερώτημα Α1 και βρίσκουμε για όλα τα σύνολα την απόσταση από το πλησιέστερο κέντρο και καταλήγουμε ότι:

- το w_1 θα έχει ως κοντινότερη απόσταση τα παραδείγματα a,b,c
- το w_2 θα έχει ως κοντινότερη απόσταση τα παραδείγματα d
- το w_3 θα έχει ως κοντινότερη απόσταση τα παραδείγματα e,f,g,h

έπειτα από τον υπολογισμό της 1η επανάληψης στο βήμα 2
γίνεται κατάταξη των παραδειγμάτων με υπολογισμό τα νέα κέντρα των ομάδων ως τον μέσο όρο των παραδειγμάτων της ομάδας του. Άρα:

w_1		w_2		w_3	
w_{11}	w_{12}	w_{21}	w_{22}	w_{31}	w_{32}
2	0.1	2	-0.3	0	0.4

Τα κέντρα άλλαξαν άρα ο αλγόριθμος συνεχίζει και τελικώς καταλήγει:

- το w_1 θα έχει ως κοντινότερη απόσταση τα παραδείγματα a,b,c το οποίο θα είναι και στο κέντρο αυτών.
- το κέντρο w_2 θα έχει ως κοντινότερη απόσταση τα παραδείγματα d δεν έχει γίνει αλλαγή διότι δεν υπάρχει άλλο παράδειγμα σε ομάδα ώστε να γίνει το κέντρο τους. Άρα w_2=d
- το κέντρο w_3 θα έχει ως κοντινότερη απόσταση τα παραδείγματα e,f,g,h το οποίο θα είναι και στο κέντρο αυτών.

Ο αλγόριθμος τερματίζει μιας και δεν υπάρχει άλλη αλλαγή στο κέντρο.

Παρατηρώ ότι λόγω της αρχικοποίησης των κέντρων δεν βρέθηκε η ίδια λύση με το ερώτημα Α.
Στο ερώτημα Α η αρχικοποίηση γινόταν αρχικά όσο πιο κοντά στις πραγματικές ομάδες ενώ στο ερώτημα Β βρέθηκαν δύο κέντρα και έγινε διαχωρισμός αυτής της ομάδας.

Γ. [*3 μονάδες*] Το **σφάλμα ομαδοποίησης** (διασπορά) μιας λύσης του k-means προκύπτει υπολογίζοντας καταρχήν για κάθε παράδειγμα x την απόστασή του $d(x,w_i)$ από το κέντρο w_i της ομάδας O_i στην οποία έχει τοποθετηθεί (το d(x,w) ορίζεται στο ερώτημα Α1). Στην συνέχεια αθροίζουμε τις αποστάσεις για όλα τα παραδείγματα.

Να υπολογίσετε το σφάλμα ομαδοποίησης για τις λύσεις των ερωτημάτων Α και Β και να σχολιάσετε τη σχέση του με την ποιότητα της λύσης.

<u>Απάντηση:</u>

<u>Για το ερώτημα Α:</u>

$$\begin{aligned}
&\gamma\iota\alpha\ \mathrm{O}_1 = \{\alpha, b, c, d\}\ \kappa\alpha\iota\ w_1(2,0)\\
&d(a,w1) \Rightarrow (2.1-2)^2 + (0-0)^2 = 0.01\\
&d(b,w1) \Rightarrow (1.9-2)^2 + (0-0)^2 = 0.01\\
&d(c,w1) \Rightarrow (2-2)^2 + (0.3-0)^2 = 0.09\\
&d(d,w1) \Rightarrow (2-2.1)^2 + ((-0.3)-0)^2 = 0.09
\end{aligned}$$

$$\begin{aligned}
&\gamma\iota\alpha\ \mathrm{O}_2 = \{e, f\}\ \kappa\alpha\iota\ w_2(0,0)\\
&\gamma\iota\alpha\ d(e,w2) \Rightarrow (0.1-0)^2 + (0-0)^2 = 0.01\\
&\gamma\iota\alpha\ d(f,w2) \Rightarrow ((-0.1)-0)^2 + (0-0)^2 = 0.01
\end{aligned}$$

$$\begin{aligned}
&\gamma\iota\alpha\ \mathrm{O}_3 = \{g, h\}\ \kappa\alpha\iota\ w_3(0,0.8)\\
&\gamma\iota\alpha\ d(g,w3) \Rightarrow (0-0)^2 + (0.7-0.8)^2 = 0.01\\
&\gamma\iota\alpha\ d(h,w3) \Rightarrow (0-0)^2 + (0.9-0.8)^2 = 0.01
\end{aligned}$$

Αθροίζοντας όλες τις αποστάσεις το σφάλμα ομαδοποίησης είναι: 0.24

<u>Για το ερώτημα Β:</u>

$$\gamma\iota\alpha\ \mathrm{O}_1 = \{\alpha, b, c\}\ \kappa\alpha\iota\ w_1(2, 0.1)$$
$$d(a, w1) \Rightarrow (2.1-2)^2 + (0-0.1)^2 = 0.02$$
$$d(b, w1) \Rightarrow (1.9-2)^2 + (0-0.1)^2 = 0.02$$
$$d(c, w1) \Rightarrow (2-2)^2 + (0.3-0.1)^2 = 0.04$$

$$\gamma\iota\alpha\ \mathrm{O}_2 = \{d\}\ \kappa\alpha\iota\ w_2(2, -0.3)$$
$$d(d, w1) \Rightarrow (2-2)^2 + ((-0.3)-(-0.3))^2 = 0$$

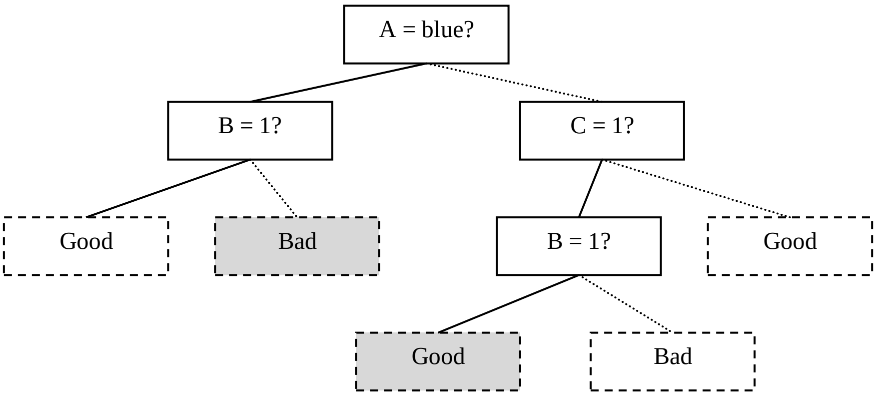
$$\gamma\iota\alpha\ \mathrm{O}_3 = \{e, f, g, h\}\ \kappa\alpha\iota\ w_3(0, 0.4)$$
$$\gamma\iota\alpha\ d(e, w2) \Rightarrow (0.1-0)^2 + (0-0.4)^2 = 0.17$$
$$\gamma\iota\alpha\ d(f, w2) \Rightarrow ((-0.1)-0)^2 + (0-0.4)^2 = 0.17$$
$$\gamma\iota\alpha\ d(g, w3) \Rightarrow (0-0)^2 + (0.7-0.4)^2 = 0.09$$
$$\gamma\iota\alpha\ d(h, w3) \Rightarrow (0-0)^2 + (0.9-0.4)^2 = 0.25$$

Αθροίζοντας όλες τις αποστάσεις το σφάλμα ομαδοποίησης είναι: 0.76

Παρατηρώ ότι το σφάλμα ομαδοποίησης στο ερώτημα Β είναι μεγαλύτερο διότι δεν παράγει σωστή λύση στις ομάδες. Ενώ στο ερώτημα Α το σφάλμα είναι μικρότερο διότι παράγει πιο σωστή λύση στις ομάδες.

Θέμα 5: Δέντρα Απόφασης [*15 μονάδες*]

Σας δίνεται το ακόλουθο δυαδικό δέντρο απόφασης:



Οι εσωτερικοί κόμβοι φαίνονται ως ορθογώνια με συνεχείς γραμμές ενώ τα φύλλα φαίνονται ως ορθογώνια με διάστικτες γραμμές.

Οι απαντήσεις ΝΑΙ φαίνονται με συνεχείς γραμμές ενώ οι απαντήσεις ΟΧΙ φαίνονται με διάστικτες γραμμές.

Οι δυνατές τιμές για το Α είναι {blue,green}, για το Β είναι {1,2} και για το C είναι {1,2}.

Θέλουμε να κατασκευάσουμε ένα σύνολο δεδομένων εκπαίδευσης, από το οποίο θα μπορούσε να είχε προέλθει το παραπάνω δέντρο, που να είναι ελάχιστο ως προς τον αριθμό των δεδομένων (πλήθος παραδειγμάτων και πλήθος χαρακτηριστικών).

Α. [*3 μονάδες*] Ποιός είναι ο ελάχιστος αριθμός παραδειγμάτων που θα χρειαστούμε; Ποιά είναι η, κατά το δυνατόν, πιο λιτή μορφή κάθε παραδείγματος;

Απάντηση:
Ο ελάχιστος αριθμός παραδειγμάτων που θα χρειαστούν εφόσον κάθε παράδειγμα ξεκινάει από τη ρίζα περνάει από κάποιον κόμβο και έχει ως target κάποιο από τα φύλλα του δέντρου τότε τα παραδείγματα που θα χρειαστεί θα είναι όσα και τα φύλλα τελικής απόφασης target του δέντρου. Δηλαδή θα είναι 5.

Η πιο λιτή μορφή κάθε παραδείγματος: Εφόσον το κάθε σύνολο απόφασης ανήκει σε κάθε παράδειγμα θα χρειαστούμε τέτοια λιτή μορφή παραδείγματος όσοι είναι και οι διαφορετικοί κόμβοι επιλογής Δηλαδή: 3 κόμβοι στο συγκεκριμένο δέντρο.
Ανεξάρτητα από το πλήθος των φύλλων όπου ανήκει σε ένα χαρακτηριστικό μιας και είναι η τελική απόφαση target. Όλα τα φύλλα έχουν το ίδιο σύνολο τιμών.

Color	choiceB	choiceC	target

Β. [*2 μονάδες*] Πως θα είναι ένα παράδειγμα που θα φτάσει στο σκιασμένο φύλλο Good;

Απάντηση:

	χαρακτηριστικά			
Example	**Color**	**choiceB**	**choiceC**	**Target**
e_1	Green	1	1	Good

$e_1 \rightarrow$ color={Green}, choiceB={1}, choiceC={1}, target={Good}

Γ. [*2 μονάδες*] Πως θα είναι ένα παράδειγμα που θα φτάσει στο σκιασμένο φύλλο Bad;
Απάντηση:

	χαρακτηριστικά			
Example	**Color**	**choiceB**	**choiceC**	**Target**
e_1	Blue	2	1	Bad

e_1 → color={Blue}, 21hoice={2}, 21hoice={1}, target={Bad}

επειδή ή επιλογή 21hoice δεν ελέγχεται στο συγκεκριμένο παράδειγμα είναι αδιάφορη η συγκεκριμένη τιμή.

Δ. [*4 μονάδες*] Κατασκευάστε ένα ελάχιστο σύνολο δεδομένων εκπαίδευσης. Πόσα διαφορετικά τέτοια σύνολα θα μπορούσατε να φτιάξετε;

Απάντηση:

	χαρακτηριστικά			
Example	**Color**	**choiceB**	**choiceC**	**Target**
e_1	Blue	1	1	Good
e_2	Blue	2	1	Bad
e_3	Green	1	1	Good
e_4	Green	2	1	Bad
e_5	Green	1	2	Good

Το σύνολο που κατασκεύασα έχει 5 παραδείγματα όσα και τα φύλλα απόφασης target. Σε κάποια παραδείγματα υπάρχουν κάποιοι κόμβοι αποφάσεων οι οποίοι δεν με ενδιαφέρουν διότι είχε ολοκληρωθεί το κάθε παράδειγμα ανεξαρτήτως απόφασης των κόμβων αυτών.

Έπειτα από εξαντλητική ανάπτυξη των κόμβων αποφάσεων οι οποίοι δεν συμμετέχουν όταν έχει ένα παράδειγμα ολοκληρωθεί μπορούν να κατασκευαστούν 8 διαφορετικά σύνολα απόφασης.

Ε. [*4 μονάδες*] Προσαρμόστε 2 ελάχιστα σύνολα δεδομένων εκπαίδευσης ώστε να τα τροφοδοτήσετε ως είσοδο στο πρόγραμμα κατασκευής δέντρων απόφασης που βρίσκεται στο AISpace (http://aispace.org/dTree/), χρησιμοποιώντας (στο Aispace) το κέρδος πληροφορίας ως μέθοδο επιλογής σας διαχωριστικής ιδιότητας σε κάθε κόμβο.
Παράγεται το δέντρο απόφασης πσαςσας δόθηκε στην αρχή του θέματος;
Εξηγήστε σύσαςμα τις παρασαςήσεις σας.

Απάντηση:
Δυο ελάχιστα σύνολα δεδομένων εκπαίδευσης είναι ως εξής:

Σύνολο α:

	χαρακτηριστικά			
Example	**Color**	**choiceB**	**choiceC**	**Target**
e_1	Blue	1	1	Good
e_2	Blue	2	1	Bad
e_3	Green	1	1	Good
e_4	Green	2	1	Bad
e_5	Green	1	2	Good

Σύνολο β:

	χαρακτηριστικά			
Example	**Color**	**choiceB**	**choiceC**	**Target**
e_1	Blue	1	2	Good
e_2	Blue	2	1	Bad

e_3	Green	1	1	Good
e_4	Green	2	1	Bad
e_5	Green	1	2	Good

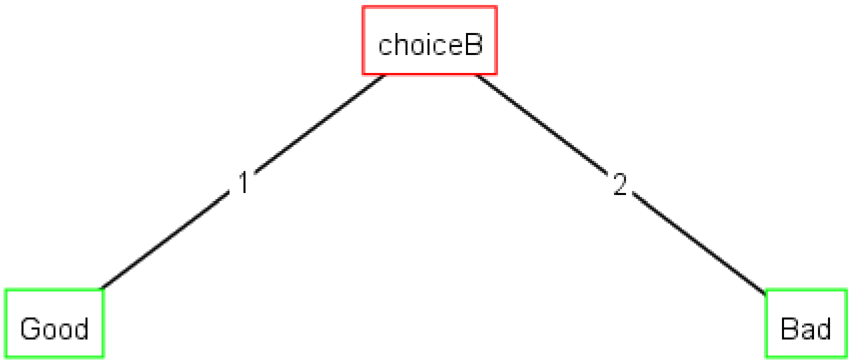
Ο κώδικας που χρησιμοποιήθηκε στο πρόγραμμα κατασκευής δέντρων απόφασης είναι:

```
T: color, choiceB, choiceC, target;

A: Blue,1,1,Good;
A: Blue,2,1,Bad;
A: Green,1,1,Good;
A: Green,2,1,Bad;
A: Green,1,2,Good;

A: Blue,1,2,Good;
A: Blue,2,1,Bad;
A: Green,1,1,Good;
A: Green,2,1,Bad;
A: Green,1,2,Good;
```

Το δέντρο που παράχθηκε είναι το εξής:



Παρατηρώ ότι το πρόγραμμα έκανε μια πιο απλουστευμένη μορφή απόφασης βάση των δεδομένων που εισήγαγα διότι για οποιοδήποτε συνδυασμό των συνόλων για κάθε χαρακτηριστικό το κομβικό σημείο του choiceB όταν είναι 1 τότε το target είναι Good αλλά όταν είναι 2 τότε το target είναι Bad. Άρα όταν ρωτήσουμε τον κόμβο choiceB χωρίς να ρωτήσουμε τους άλλους κόμβους η απόφαση θα στόχος θα είναι η ίδια.