Σειρά προβλημάτων 1

Κυρίτσης Βασίλειος 02999 και Σταμούλος Αλέξανδρος 02954

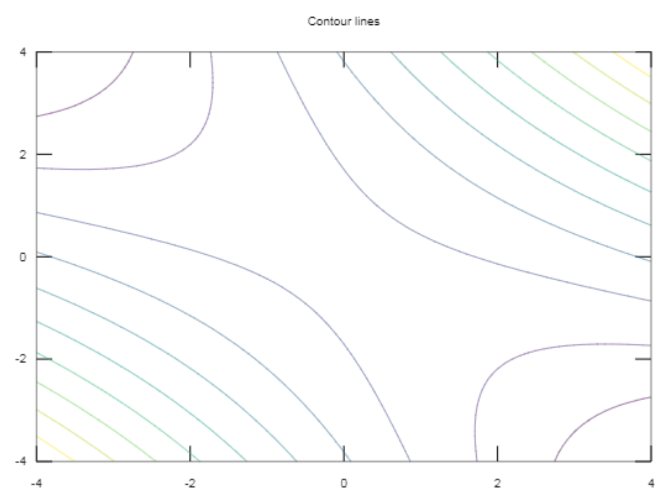
Νευρο-Ασαφής Υπολογιστική 2023-24
Τμήμα Ηλεκτρολόγων Μηχανικών & Μηχανικών Υπολογιστών
Πανεπιστήμιο Θεσσαλίας, Βόλος
{vakyritsis, astamoulos}@e-ce.uth.gr

1 Problem-01

Για την συνάρτηση:

$$f(x,y) = x^2 + 4xy + y^2$$

τα contour lines είναι:



Γενικός τύπος quadratic function:

$$F(x) = c + d^T \hat{x} + \frac{1}{2} \hat{x}^T A \hat{x}$$

Η f δεν έχει σταθερούς ορούς ούτε το x και y αυτούσια οπότε το c = 0 και d = 0. Άρα έχει την μορφή :

$$f(x,y) = \begin{bmatrix} x \ y \end{bmatrix} \begin{bmatrix} a_{11} \ a_{12} \\ a_{21} \ a_{22} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} xa_{11} + ya_{21} \ xa_{12} + ya_{22} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$
$$= a_{11}x^2 + (a_{12} + a_{21})xy + a_{22}y^2$$

Άρα έχουμε $a_{11} = 1$, $a_{22} = 1$ και $a_{12} = a_{21} = 2$. Οπότε η τετραγωνική μορφή της f είναι η

$$f(x,y) = \begin{bmatrix} x \ y \end{bmatrix} \begin{bmatrix} 1 \ 2 \\ 2 \ 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \frac{1}{2} \begin{bmatrix} x \ y \end{bmatrix} \begin{bmatrix} 2 \ 4 \\ 4 \ 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

Ο πίνακάς A = $\begin{bmatrix} 2 \ 4 \\ 4 \ 2 \end{bmatrix}$ ειναι η δευτερη παραγωγος της f δηλαδη ο Hessian matrix. Οι ιδιοτιμές του H μπορουν να βρςθουν χρησιμοποιώντας τη χαρακτηριστική εξίσωση $det(H - I) = 0$ οπου λ οι ιδιοτιμες και I ο μοναδιαιος πινακας:

$$\det\left(\begin{bmatrix} 2-\lambda & 4 \\ 4 & 2-\lambda \end{bmatrix}\right) = (2-\lambda)^2 - 16 = \lambda^2 - 4\lambda - 12 = 0$$

Λυνοντας $\lambda^2 - 4\lambda - 12 = 0$ παιρνουμε τις ιδιοτιμες $\lambda = 6, -2$. Η οριζουσα του Hessian matrix H δινεται απο $\det(H) = (2)(2) - (4)(4) = 4 - 16 = -12$.

Αφού ο Hessian matrix έχει αρνητική ορίζουσα και οι ιδιοτιμές του είναι τόσο θετικές όσο και αρνητικές, αυτό υποδεικνύει ότι το κρίσιμο σημείο είναι ένα σαγματικό σημειο. Συνεπώς, η συνάρτηση $f(x,y) = x^2 + 4xy + y^2$, δεν εχει τοπικά ελάχιστα ή μέγιστα.

2 Problem-02

$$f(x_1, x_2) = (x_1 + 2x_2 - 7)^2 + (2x_1 + x_2 - 5)^2$$

Απλοποιούμε ανοίγοντάς την ταυτότητά

$$\begin{aligned} f(x_1,x_2) &= x_1^2 + 4x_2^2 + 49 + 2(2x_1x_2 + x_1(-7) + 2x_2(-7)) \\ &\quad + 4x_1^2 + x_2^2 + 25 + 2(2x_1x_2 + 2x_1(-5) + x_2(-5)) \\ &= x_1^2 + 4x_2^2 + 49 + 4x_1x_2 - 14x_1 - 28x_2 \\ &\quad + 4x_1^2 + x_2^2 + 25 + 4x_1x_2 - 20x_1 - 10x_2 \\ &= 5x_1^2 - 34x_1 + 8x_1x_2 + 5x_2^2 - 38x_2 + 74. \end{aligned}$$

$$\nabla f(x_1,x_2) = \begin{pmatrix} \dfrac{\partial f}{\partial x_1} \\ \dfrac{\partial f}{\partial x_2} \end{pmatrix} = \begin{pmatrix} 10x_1 + 8x_2 - 34 \\ 10x_2 + 8x_1 - 38 \end{pmatrix}$$

$$||\nabla f(x_1,x_2)|| = \sqrt{\nabla^T f(x_1,x_2) \nabla f(x_1,x_2)}$$

$$= \sqrt{\left(\begin{pmatrix}10x_1 + 8x_2 - 34 & 10x_2 + 8x_1 - 38\end{pmatrix}\right)\left(\begin{pmatrix}10x_1 + 8x_2 - 34 \\ 10x_2 + 8x_1 - 38\end{pmatrix}\right)}$$

$$= \sqrt{(10x_1 + 8x_2 - 34)^2 + (10x_2 + 8x_1 - 38)^2}$$

Iteration 1

$$\hat{x_0} = \begin{pmatrix}-9.5 & 9.5\end{pmatrix}$$

με αντικατάσταση έχουμε

$$||\nabla f(\hat{x_0})|| = \sqrt{(10(-9.5) + 8(9.5) - 34)^2 + (10(9.5) + 8(-9.5) - 38)^2} = \sqrt{3170}$$

και

$$\nabla f(\hat{x_0}) = \begin{pmatrix}10(-9.5) + 8(9.5) - 34 \\ 10(9.5) + 8(-9.5) - 38\end{pmatrix} = \begin{pmatrix}-53 \\ -19\end{pmatrix}$$

Άρα η κατεύθυνση καθόδου:

$$\frac{-\nabla f(\hat{x_0})}{||\nabla f(\hat{x_0})||} = -\frac{1}{\sqrt{3170}}\begin{pmatrix}-53 \\ -19\end{pmatrix} = \frac{1}{\sqrt{3170}}\begin{pmatrix}53 \\ 19\end{pmatrix}$$

Συνεπώς

$$\hat{x_1} = \begin{pmatrix}-9.5 & 9.5\end{pmatrix}^T - \lambda_0 \frac{1}{\sqrt{3170}}\begin{pmatrix}53 & 19\end{pmatrix}^T \Rightarrow \hat{x_1} = \left(-9.5 - \lambda_0 \frac{53}{\sqrt{3170}} \quad 9.5 - \lambda_0 \frac{19}{\sqrt{3170}}\right)^T$$

Αντικαθιστούμε στην f το $\hat{x_1}$

$$f(\hat{x_1}) = 5(-9.5 - \lambda_0 \frac{53}{\sqrt{3170}})^2 - 34(-9.5 - \lambda_0 \frac{53}{\sqrt{3170}}) + 8(-9.5 - \lambda_0 \frac{53}{\sqrt{3170}})(9.5 - \lambda_0 \frac{19}{\sqrt{3170}})$$
$$+ 5(9.5 - \lambda_0 \frac{19}{\sqrt{3170}})^2 - 38(9.5 - \lambda_0 \frac{19}{\sqrt{3170}}) + 74$$

Για να βρούμε το βρούμε το optimal λ_0, θα βρούμε την παράγωγο την προηγούμενης συνάρτησης ως προς λ_0 και θα την θέσουμε ίση με το 0.

$$\begin{aligned}
\frac{df(\hat{x_1})}{d\lambda_0} =& 2 \cdot 5 \cdot \frac{-53}{\sqrt{3170}}(-9.5 - \lambda_0 \frac{53}{\sqrt{3170}}) - 34 \cdot \frac{-53}{\sqrt{3170}} \\
& + 8\left[(\frac{-53}{\sqrt{3170}})(9.5 - \lambda_0 \frac{19}{\sqrt{3170}}) + (-9.5 - \lambda_0 \frac{53}{\sqrt{3170}})(\frac{-19}{\sqrt{3170}})\right] \\
& + 2 \cdot 5 \cdot \frac{-19}{\sqrt{3170}}(9.5 - \lambda_0 \frac{19}{\sqrt{3170}}) - 38 \cdot \frac{-19}{\sqrt{3170}} \\
=& \frac{-530}{\sqrt{3170}}(-9.5 - \lambda_0 \frac{53}{\sqrt{3170}}) + \frac{1802}{\sqrt{3170}} + \frac{-190}{\sqrt{3170}}(9.5 - \lambda_0 \frac{19}{\sqrt{3170}}) + \frac{722}{\sqrt{3170}} \\
& + 8(\frac{-503.5}{\sqrt{3170}} + \lambda_0 \frac{1007}{(\sqrt{3170})^2} + \frac{180.5}{\sqrt{3170}} + \lambda_0 \frac{1007}{(\sqrt{3170})^2}) \\
=& \frac{5035}{\sqrt{3170}} + \lambda_0 \frac{28090}{(\sqrt{3170})^2} - \frac{1805}{\sqrt{3170}} + \lambda_0 \frac{3610}{(\sqrt{3170})^2} \\
& + \frac{2524}{\sqrt{3170}} + 8\left(-\frac{323}{\sqrt{3170}} + \lambda_0 \frac{2014}{(\sqrt{3170})^2}\right) \\
=& \lambda_0 \frac{31700}{(\sqrt{3170})^2}) + \frac{5754}{\sqrt{3170}} - \frac{2584}{\sqrt{3170}} + \lambda_0 \frac{16112}{(\sqrt{3170})^2}) \\
=& \lambda_0 \frac{47812}{(\sqrt{3170})^2}) + \frac{3170}{\sqrt{3170}}
\end{aligned}$$

Άρα θα έχουμε:

$$\begin{aligned}
&\frac{df(\hat{x_1})}{d\lambda_0} = 0 \\
&\lambda_0 \frac{47812}{(\sqrt{3170})^2} + \frac{3170}{\sqrt{3170}} = 0 \\
&\lambda_0 \frac{47812}{(\sqrt{3170})^2} = -\frac{3170}{\sqrt{3170}} \\
&47812\lambda_0 = -3170 \cdot (\sqrt{3170}) \\
&\lambda_0 = \frac{-3170 \cdot (\sqrt{3170})}{47812} \\
&\lambda_0 = -3.7329
\end{aligned}$$

Αντικαθιστούμε το λ0 στο x1 και για να βρούμε το καινούργιο σημείο.

$$\hat{x_1} = \left(-9.5 - \frac{-3170 \cdot (\sqrt{3170})}{47812} \frac{53}{\sqrt{3170}} \;\; 9.5 - \frac{-3170 \cdot (\sqrt{3170})}{47812} \frac{19}{\sqrt{3170}}\right)^T$$

$$\hat{x_1} = \left(-9.5 - \frac{-3170 \cdot 53}{47812} \;\; 9.5 - \frac{-3170 \cdot 19}{47812}\right)^T = \left(-9.5 + 3.514 \;\; 9.5 + 1.26\right)^T$$

$$\hat{x_1} = \left(-5.986 \;\; 10.76\right)^T$$

Iteration 2 με αντικατάσταση έχουμε

$$||\nabla f(\hat{x_1})|| = \sqrt{(10(-5.986) + 8(10.76) - 34)^2 + (10(10.76) + 8(-5.986) - 38)^2} = \frac{\sqrt{33246209}}{250}$$

και

$$\nabla f(\hat{x_1}) = \begin{pmatrix} 10(-5.986) + 8(10.76) - 34 \\ 10(10.76) + 8(-5.986) - 38 \end{pmatrix} = \begin{pmatrix} -7.78 \\ 21.712 \end{pmatrix}$$

Άρα η κατεύθυνση καθόδου:

$$\frac{-\nabla f(\hat{x_1})}{||\nabla f(\hat{x_1})||} = -\frac{1}{\frac{\sqrt{33246209}}{250}} \begin{pmatrix} -7.78 \\ 21.712 \end{pmatrix} = \frac{250}{\sqrt{33246209}} \begin{pmatrix} 7.78 \\ -21.712 \end{pmatrix}$$

Συνεπώς

$$\hat{x_2} = \left(-5.986 \;\; 10.76\right)^T - \lambda_1 \frac{250}{\sqrt{33246209}} \begin{pmatrix} 7.78 \\ -21.712 \end{pmatrix}^T \Rightarrow$$

$$\hat{x_2} = \left(-5.986 - \lambda_1 \frac{1945}{\sqrt{33246209}} \;\; 10.76 + \lambda_1 \frac{5428}{\sqrt{33246209}}\right)^T$$

Αντικαθιστούμε στην f το $\hat{x_2}$

$$\begin{aligned} f(\hat{x_2}) = 5(-5.986 - \lambda_1 \frac{1945}{\sqrt{33246209}})^2 - 34(-5.986 - \lambda_1 \frac{1945}{\sqrt{33246209}}) \\ + 8(-5.986 - \lambda_1 \frac{1945}{\sqrt{33246209}})(10.76 + \lambda_1 \frac{5428}{\sqrt{33246209}}) \\ + 5(10.76 + \lambda_1 \frac{5428}{\sqrt{33246209}})^2 - 38(10.76 + \lambda_1 \frac{5428}{\sqrt{33246209}}) + 74 \end{aligned}$$

Για να βρούμε το βρούμε το optimal λ_1, θα βρούμε την παράγωγο την προηγούμενης συνάρτησης ως προς λ_1 και θα την θέσουμε ίση με το 0.

$$\begin{aligned}
\frac{df(\hat{x_2})}{d\lambda_1} =& 2 \cdot 5 \cdot \frac{-1945}{\sqrt{33246209}}(-5.986 - \lambda_1 \frac{1945}{\sqrt{33246209}}) - 34 \cdot \frac{-1945}{\sqrt{33246209}} \\
&+ 8\left[(\frac{-1945}{\sqrt{33246209}})(10.76 + \lambda_1 \frac{5428}{\sqrt{33246209}}) + (-13.015 - \lambda_1 \frac{1945}{\sqrt{33246209}})(\frac{5428}{\sqrt{33246209}})\right] \\
&+ 2 \cdot 5 \cdot \frac{5428}{\sqrt{3170}}(10.76 + \lambda_1 \frac{5428}{\sqrt{33246209}}) - 38 \cdot \frac{5428}{\sqrt{33246209}} \\
=& \frac{-19450}{\sqrt{33246209}}(-5.986 - \lambda_1 \frac{1945}{\sqrt{33246209}}) + \frac{66130}{\sqrt{33246209}} \\
&+ \frac{54280}{\sqrt{33246209}}(10.76 + \lambda_1 \frac{5428}{\sqrt{33246209}}) + \frac{206264}{\sqrt{33246209}} \\
&+ 8\left(-\frac{20928,2}{\sqrt{33246209}} - \lambda_1 \frac{10557460}{(\sqrt{33246209})^2} - \frac{32492.008}{\sqrt{33246209}} - \lambda_1 \frac{15136160}{(\sqrt{10557460})^2}\right) \\
=& \frac{116427.7}{\sqrt{33246209}} + \lambda_1 \frac{37830250}{(\sqrt{33246209})^2} + \frac{272394}{\sqrt{33246209}} + \lambda_1 \frac{294631840}{(\sqrt{33246209})^2} \\
&+ \frac{58405.28}{\sqrt{33246209}} + 8\left(-\frac{53420.208}{\sqrt{33246209}} - \lambda_1 \frac{21114920}{(\sqrt{33246209})^2}\right) \\
=& \lambda_1 \frac{332462090}{(\sqrt{33246209})^2} + \frac{447226.98}{\sqrt{33246209}} - \frac{427361.664}{\sqrt{33246209}} - \lambda_1 \frac{168919424}{(\sqrt{33246209})^2} \\
=& \lambda_1 \frac{163542666}{(\sqrt{33246209})^2} + \frac{19865.316}{\sqrt{33246209}}
\end{aligned}$$

Άρα θα έχουμε:

$$\begin{aligned}
&\frac{df(\hat{x_2})}{d\lambda_1} = 0 \\
&\lambda_1 \frac{163542666}{(\sqrt{33246209})^2} + \frac{19865.316}{\sqrt{33246209}} = 0 \\
&\lambda_1 \frac{163542666}{(\sqrt{33246209})^2} = -\frac{19865.316}{\sqrt{33246209}} \\
&163542666\lambda_1 = -19865.316 \cdot (\sqrt{33246209}) \\
&\lambda_1 = \frac{-19865.316 \cdot (\sqrt{33246209})}{163542666} \\
&\lambda_1 = -0.7
\end{aligned}$$

Αντικαθιστούμε το λ1 στο x2 και για να βρούμε το καινούργιο σημείο.

$$\hat{x_2} = \begin{pmatrix} -5.986 - \frac{-19865.316 \cdot (\sqrt{33246209})}{163542666} \frac{1945}{\sqrt{33246209}} \\ 10.76 + \frac{19865.316 \cdot (\sqrt{33246209})}{163542666} \frac{5428}{\sqrt{33246209}} \end{pmatrix}^T$$

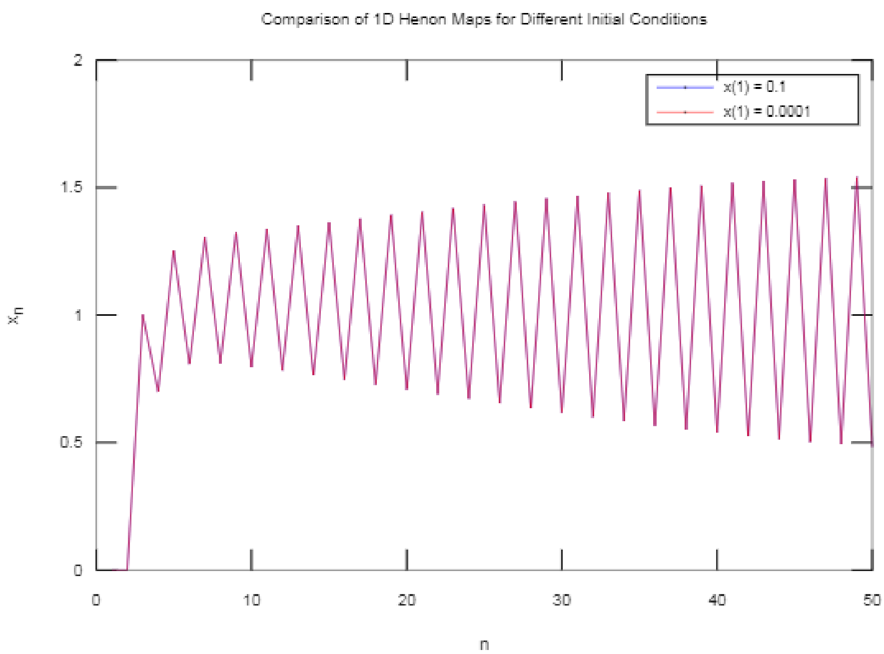
$$\hat{x_2} = \begin{pmatrix} -5.986 - \frac{-19865.316 \cdot 1945}{163542666} & 10.76 + \frac{-19865.316 \cdot 5428}{163542666} \end{pmatrix}^T$$

$$= \begin{pmatrix} -5.986 + 2.362 & 10.76 - 6.6 \end{pmatrix}^T = \begin{pmatrix} -3.624 & 4.16 \end{pmatrix}^T$$

3 Problem-03

```
% Parameters
a = 0.3;
b = 0.4;

% Initial conditions - First set
n = 50;
x1 = zeros(1, n);
x1(1) = 0;
x1(2) = 0;

% Henon map iterations - First set
for i = 3:n
    x1(i) = 1 - a * x1(i-1)^2 + b * x1(i-2);
end

% Initial conditions - Second set
x2 = zeros(1, n);
x2(1) = 0.00001;
x2(2) = 0.00001;

% Henon map iterations - Second set
for i = 3:n
    x2(i) = 1 - a * x2(i-1)^2 + b * x2(i-2);
end

% Plotting both sets together
figure;
plot(1:n, x1, 'b', 1:n, x2, 'r');
```

```
title('Comparison of 1D Henon Maps for Different Initial Conditions');
xlabel('n');
ylabel('x_n');
legend('x(1) = 0.1', 'x(1) = 0.0001');
axis ("auto[x]");
```



Παρατηρούμε ότι για (a,b)=(0.3, 0.4) το σύστημά είναι περιοδικό αφού για μια πολύ μικρή αλλαγή στην αρχικές τιμές (0.00001) το σύστημά βγάζει την ίδια έξοδο.

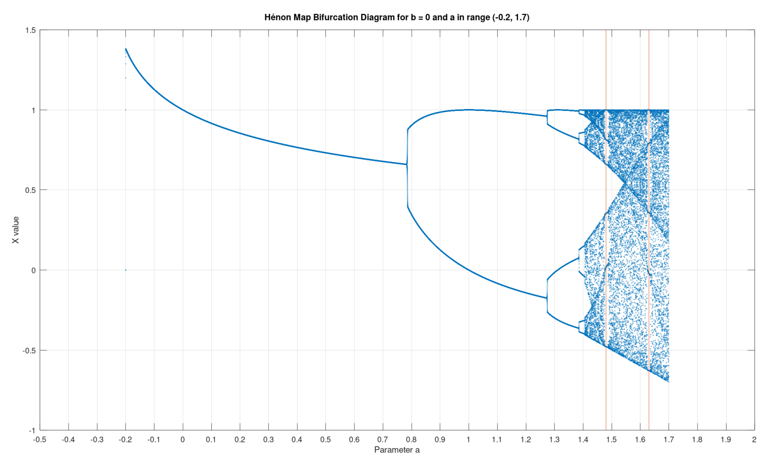
A.

Για να μην κάνουμε πάρα πολλά plot θα κάνουμε ένα bifurcation diagram για b=0 και a = [-0.2, 1.7] έτσι θα δούμε την συμπεριφορά του συστήματος για πολλά α

```
%octave doesn't have xline function so we implement it
xline = @(xval, varargin) line([xval xval], ylim, varargin{:});

a_values = linspace(-0.2, 1.7, 100000);
b = 0;
x = [0, 0]; % Initializing x with 0, 0
```

```
for i = 3:length(a_values)
    x(i) = 1 - a_values(i) * x(i - 1)^2 + b * x(i - 2);
end

plot(a_values, x, 'o', 'MarkerSize', 1);
grid on;
xlabel('Parameter a');
ylabel('X value');
title('Hénon Map Bifurcation Diagram for b = 0 and a in range (-0.2, 1.7)');
xline(1.63);
xline(1.48);
% Setting x-axis ticks for every 0.1
xticks(-1:0.1:2)
```

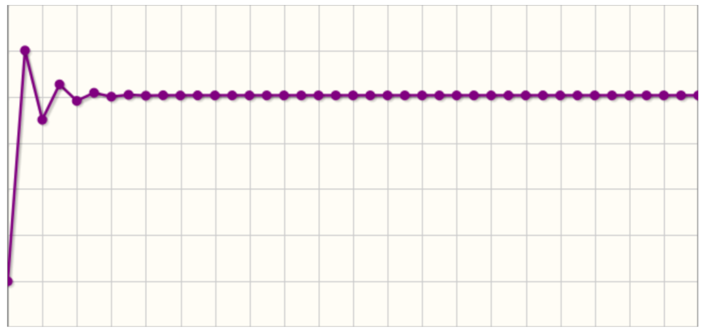
Και παίρνουμε το παρακάτω διάγραμμα



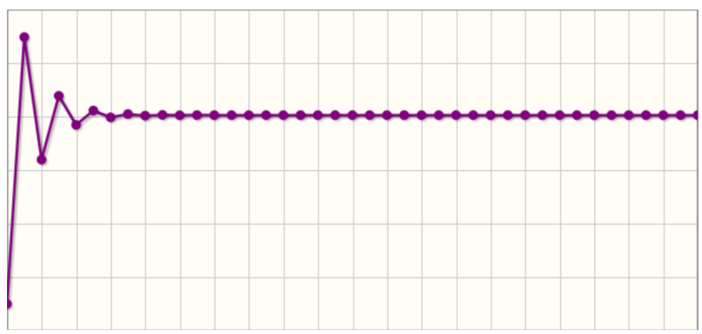
Από το διάγραμμα παρατηρούμε για α από -0.2 μέχρι περίπου 0.79 το σύστημα είναι περιοδικό και συγκλίνει σε μια τιμή, για α από 0.79 μέχρι 1.26 το σύστημα είναι περιοδικό και έχει δυο fixed points, για α 1.26 μέχρι 1.38 το σύστημα έχει 4 fixed points για α 1.38 μέχρι 1.41 έχει 8 fixed points. Για τιμές μεγαλύτερες του 1.4 το σύστημα βλέπουμε ότι είναι χαοτικό με κάποια intervals που το σύστημα είναι περιοδικό π.χ για α=1.48 και α=1.63 όπου έχουν 10 και 12 fixed points αντίστοιχα.

B.

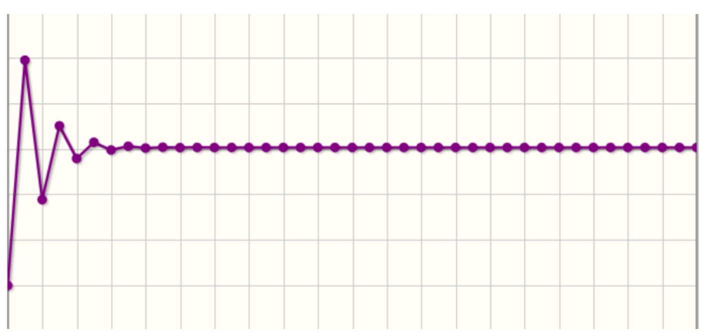
Δοκιμάζουμε για διάφορες τιμές x0 και b ανάμεσα στο 0 και στο 0.32 και κρατώντας το α σταθερό στο 0.3



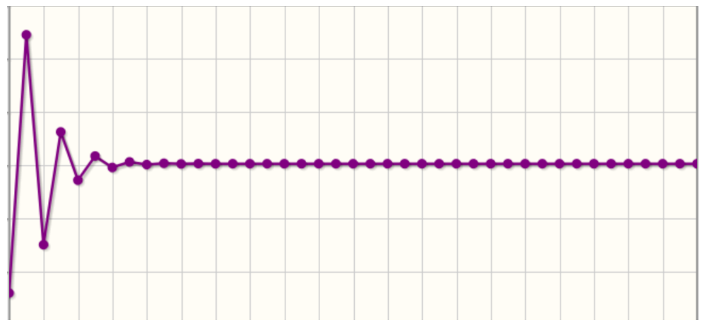
Σχήμα 1. b=0 x=0



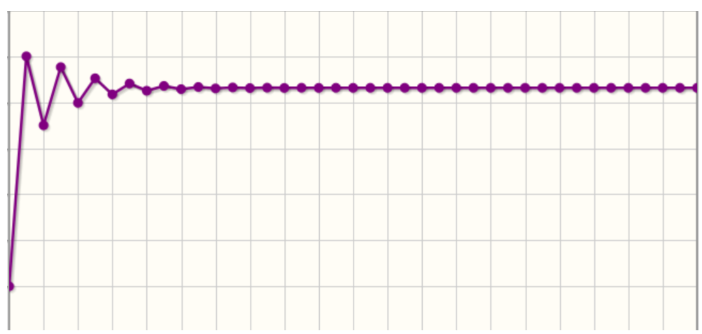
Σχήμα 2. b=0 x=0.1



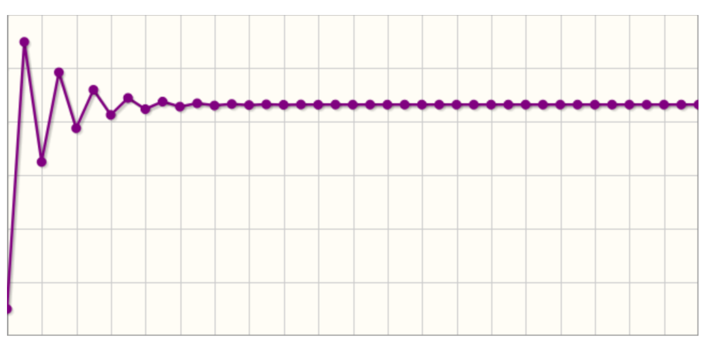
Σχήμα 3. b=0 x=0.2



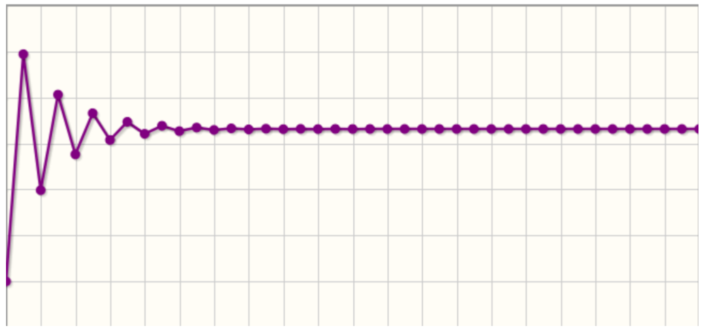
Σχήμα 4. b=0 x=0.32



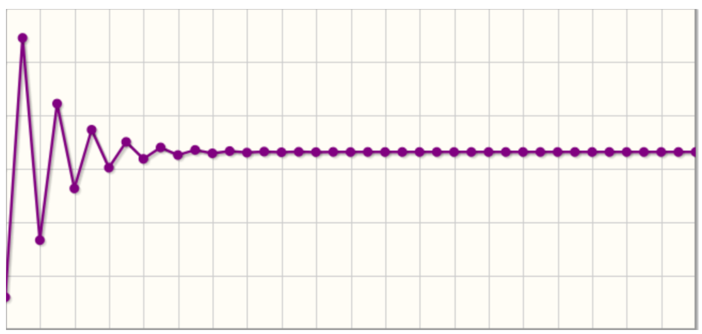
Σχήμα 5. b=0.1 x=0



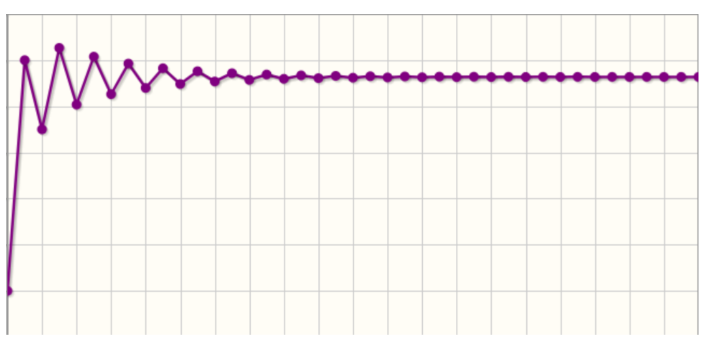
Σχήμα 6. b=0.1 x=0.1



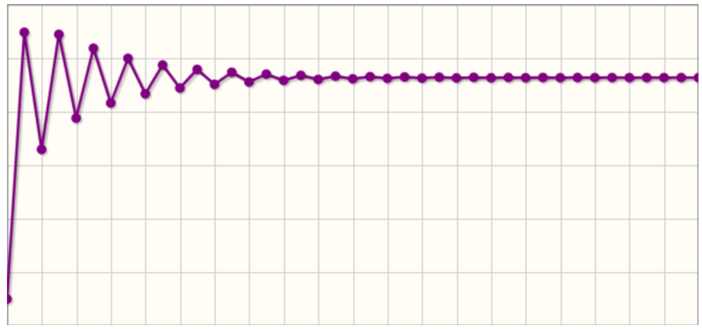
Σχήμα 7. b=0.1 x=0.2



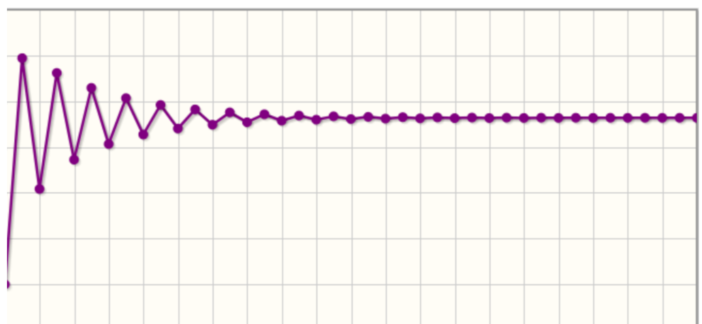
Σχήμα 8. b=0.1 x=0.32



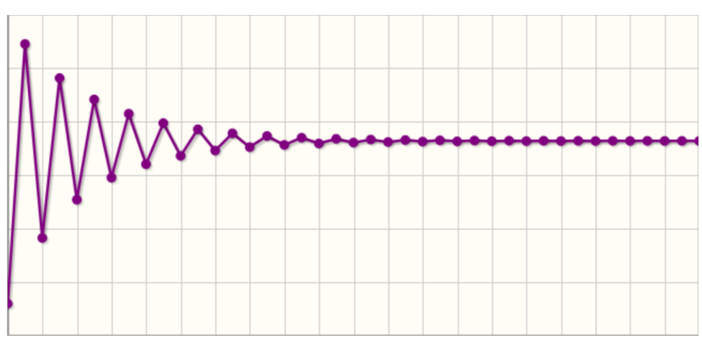
Σχήμα 9. b=0.2 x=0



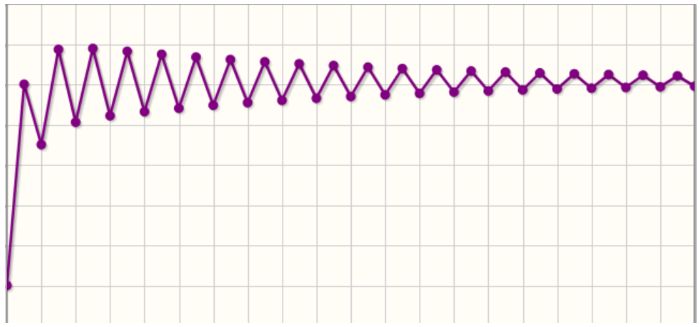
Σχήμα 10. b=0.2 x=0.1



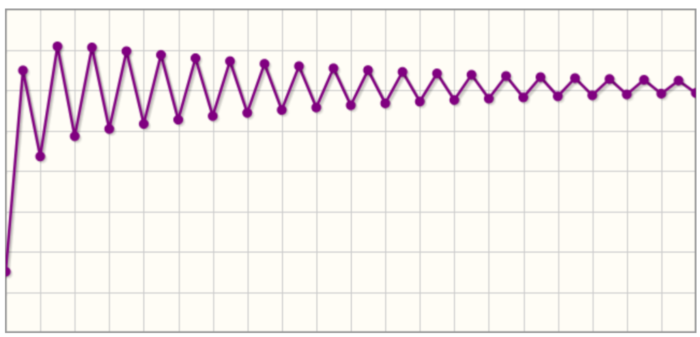
Σχήμα 11. b=0.2 x=0.2



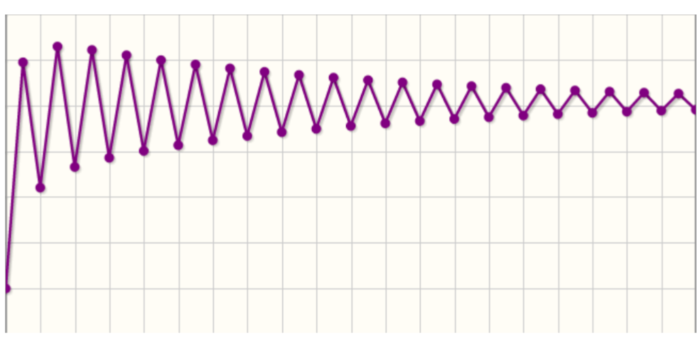
Σχήμα 12. b=0.2 x=0.32



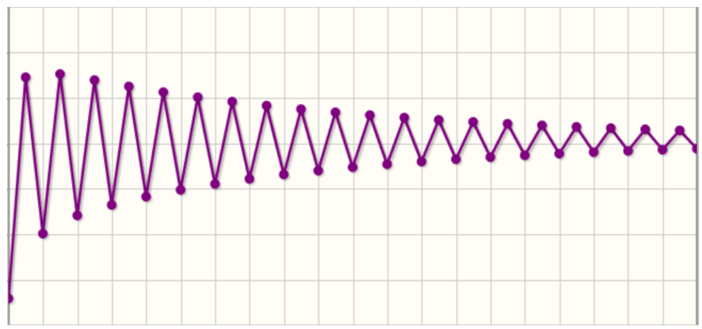
Σχήμα 13. b=0.32 x=0



Σχήμα 14. b=0.32 x=0.1



Σχήμα 15. b=0.32 x=0.2



Σχήμα 16. b=0.32 x=0.32

Παρατηρούμε ότι η συμπεριφορά του συστήματος δεν επηρεάζεται πολύ από τις αρχικές τιμές χ άλλα από τον παράγοντά b

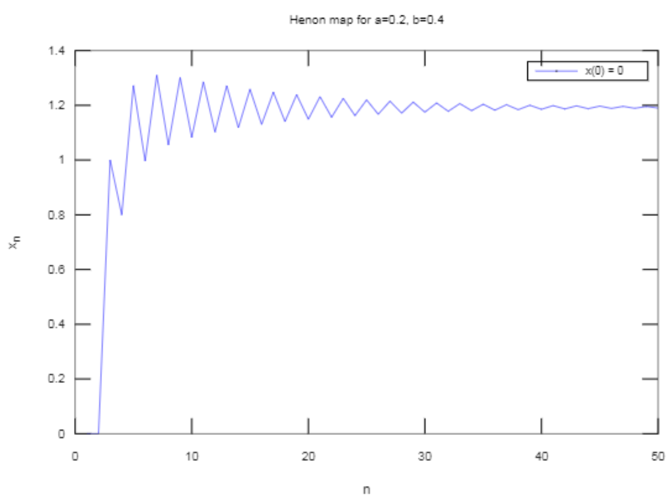
C.

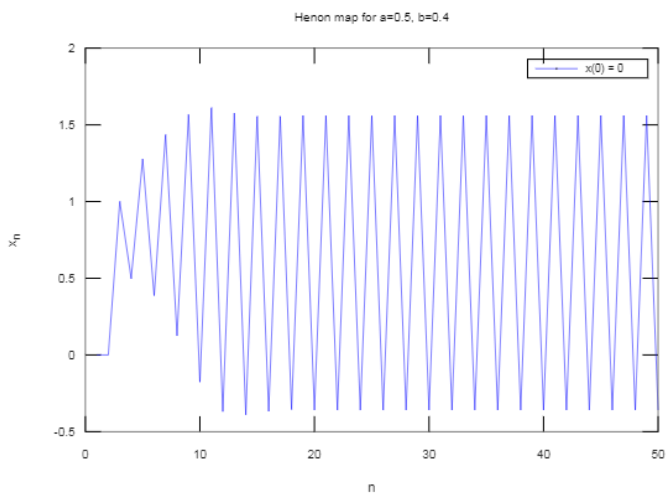
Παρατηρήσαμε ότι για μικρότερες τιμές του a το σύστημα έχει ένα fixed point, για (a,b)=(0.3675, 0.3) έχουμε το πρώτο bifurcation και το σύστημα έχει δυο fixed points. Οι τιμές αυτές είναι μια ειδική περίπτωσή όπου όταν

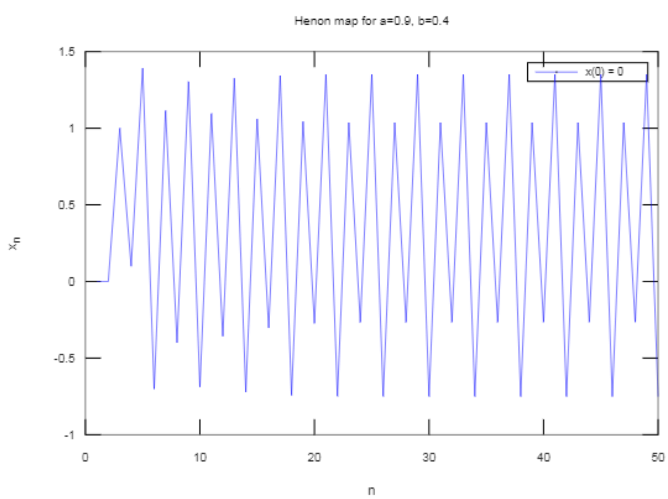
$$a = \frac{3(b-1)^2}{4}$$

έχουμε period doubling bifurcation κάτι που επιβεβαιώνουμε.

D.







Παρατηρούμε ότι για όλα τα ζεύγη τιμών το σύστημα είναι περιοδικό αλλά διπλασιάζονται τα fixed points. Για α=0.2 έχουμε 1 fixed point , για α=0.5 έχουμε 2 και για α-0.9 4 fixed points. Αυτό το φαινόμενο στα δυναμικά συστήματα ονομάζεται period-doubling bifurcation.

4 Problem-04

$$
\begin{aligned}
\frac{d}{dx}S &= \frac{d}{dx}\left(\frac{1}{1+e^{-x}}\right)\\
&= \frac{d}{dx}\left(1+\mathrm{e}^{-x}\right)^{-1}\\
&= -(1+e^{-x})^{-2}(-e^{-x})\\
&= \frac{e^{-x}}{\left(1+e^{-x}\right)^2}\\
&= \frac{1}{1+e^{-x}}\cdot\frac{e^{-x}}{1+e^{-x}}\\
&= \frac{1}{1+e^{-x}}\cdot\frac{(1+e^{-x})-1}{1+e^{-x}}\\
&= \frac{1}{1+e^{-x}}\cdot\left(\frac{1+e^{-x}}{1+e^{-x}}-\frac{1}{1+e^{-x}}\right)\\
&= \frac{1}{1+e^{-x}}\cdot\left(1-\frac{1}{1+e^{-x}}\right)\\
&= S\cdot(1-S)
\end{aligned}
$$

$$
\begin{aligned}
\frac{d}{dx}S &= \frac{d}{dx}\left(\frac{e^x-e^{-x}}{e^x+e^{-x}}\right)\\
&= \frac{e^x+e^{-x}}{\left(e^x+e^{-x}\right)^2}\frac{d}{dx}(e^x-e^{-x}) - \frac{e^x-e^{-x}}{\left(e^x+e^{-x}\right)^2}\frac{d}{dx}(e^x+e^{-x})\\
&= \frac{\left(e^x+e^{-x}\right)\left(e^x+e^{-x}\right)}{\left(e^x+e^{-x}\right)^2} - \frac{\left(e^x-e^{-x}\right)\left(e^x-e^{-x}\right)}{\left(e^x+e^{-x}\right)^2}\\
&= \frac{\left(e^x+e^{-x}\right)^2}{\left(e^x+e^{-x}\right)^2} - \frac{\left(e^x-e^{-x}\right)^2}{\left(e^x+e^{-x}\right)^2}\\
&= 1-S^2
\end{aligned}
$$

Από το πρώτο ερώτημά έχουμε την sigmoid σ(x) = $\frac{1}{1+e^{-x}}$ με $\frac{d\sigma}{dx} = \sigma\cdot(1-\sigma)$ αρά

$$
\begin{aligned}
\frac{d}{dx}S &= \frac{d}{dx}\left(\frac{x}{1+e^{-x}}\right) = \frac{d}{dx}(x\cdot\sigma\ )\\
&= \frac{dx}{dx}\ \sigma\ + x\frac{d\sigma}{dx} =\ \sigma\ + x\ \sigma\ (1-\sigma) = \sigma + x\ \sigma - x\ \sigma^2\\
&= \sigma + S - x\ \sigma^2 = S +\ \sigma\ (1-x\sigma) = S +\ \sigma(\mathrm{x})\ (1-S)
\end{aligned}
$$

$$\begin{aligned}\frac{d}{dx}S &= \frac{d}{dx}(x \cdot tanh(ln(1+e^x)))\\ &= tanh(ln(1+e^x) + x(1 - tanh^2(ln(1+e^x)) \cdot \frac{d}{dx}(ln(1+e^x))\\ &= \frac{S}{x} + x \cdot \left(1 - \left(\frac{S}{x}\right)^2\right) \cdot \frac{e^x}{1+e^x}\\ &= \frac{S}{x} + swish(x) \cdot \left(1 - \left(\frac{S}{x}\right)^2\right)\end{aligned}$$

αφού

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

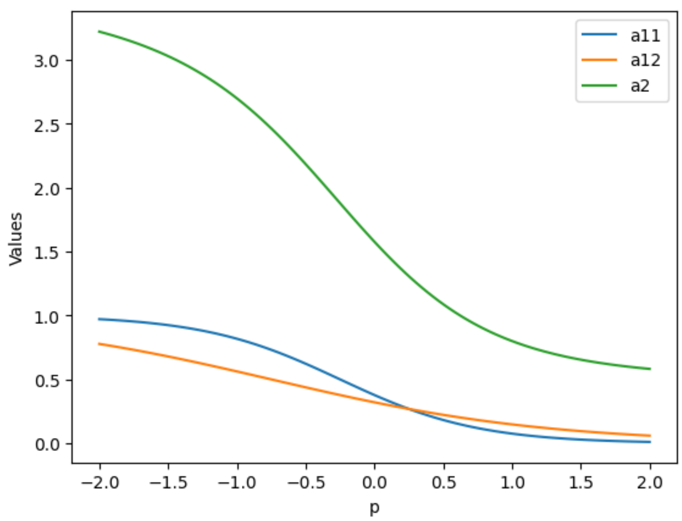
5 Problem-05

Ο παρεχόμενος κώδικας ορίζει μια απλή κλάση νευρωνικού δικτύου με μια μέθοδο αρχικοποίησης για τον καθορισμό βαρών, biases και activation functions. Περιλαμβάνει την συνάρτηση forward pass η οποία κάνει ένα πέρασμα από το νευρωνικό και επιστρέφει μια λίστα με τις εξόδους κάθε layer. Τέλος ορίζει τρεις συναρτήσεις ενεργοποίησης: purelin, sigmoid και swish. Τα αρχικά βάρη και biases ορίζονται για ένα δίκτυο δύο επιπέδων με 2 νευρώνες εισόδου και 1 νευρώνα εξόδου όπως δείχνει το σχήμα.

```python
import numpy as np
import matplotlib.pyplot as plt

class NeuralNetwork:
    def __init__(self, weights, biases, activation_functions):
        self.weights = weights
        self.biases = biases
        self.activation_functions = activation_functions

    def forward_pass(self, input_data):
        outputs = []
        out = input_data
        for i, layer in enumerate(layers):
            out = self.activation_functions[i](np.dot(self.weights[i], out) + self.biases[i])
            outputs.append(out)
        return outputs
```

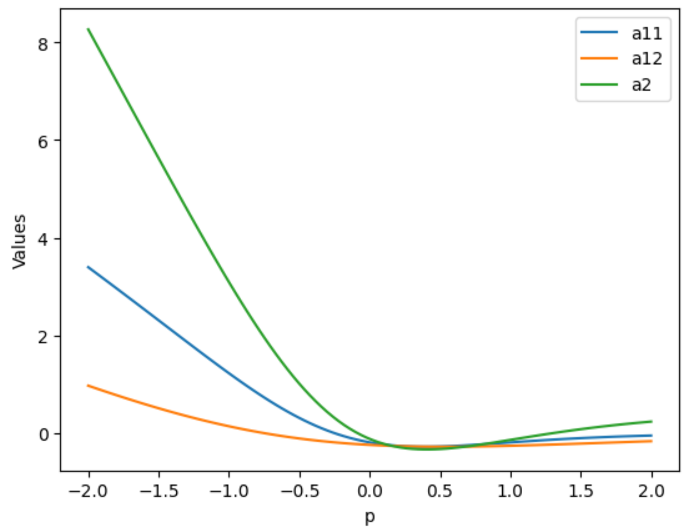
```python
def pureline(x):
  return x

def sigmoid(x):
  return 1 / (1 + np.exp(-x))

def swish(x):
  return x / (1 + np.exp(-x))

weights = [
    [-2, -1],
    [[2, 1]]
]

biases = [
    [-0.5, -0.75],
    [0.5]
]

layers = [2, 1]

activation = [sigmoid, pureline]

# Create neural network instance
neural_net = NeuralNetwork(weights, biases, activation)

p_values = np.linspace(-2, 2, 100)
a11 = []
a12 = []
a2 = []

for p in p_values:
  # Perform forward pass
  layer_outputs  = neural_net.forward_pass(p)
  layer_outputs_as_lists = [layer.tolist() for layer in layer_outputs]
  a11.append(layer_outputs_as_lists[0][0])
  a12.append(layer_outputs_as_lists[0][1])
  a2.append(layer_outputs_as_lists[1])

# Plot
```

```
plt.plot(p_values, a11, label='a11')
plt.plot(p_values, a12, label='a12')
plt.plot(p_values, a2, label='a2')
plt.xlabel('p')
plt.ylabel('Values')
plt.legend()
plt.show()
```



Σχήμα 17. Output with sigmoid activation function

Για να χρησιμοποιήσουμε την swish άπλα αλλάζουμε την λίστα activation σε activation = [swish, pureline]



Σχήμα 18. Output with swish activation function

6 Problem-06

Το Stochastic Gradient Descent (SGD) χρησιμοποιεί batch size 1, οπότε ο υπολογισμός του gradient γίνεται πολύ πιο γρήγορα από τον υπολογισμό του συνολικού Gradient Descent (GD). Το γεγονός ότι το SGD είναι καλύτερος από τον GD σημαίνει ότι το σύνολο δεδομένων επιτρέπει αποτελεσματική μάθηση ακόμα και με nb = 1, δηλαδή το dataset έχει κάλο correlation. Αυξάνοντας σταδιακά το μέγεθος του batch από 1 έως ένα μικρό υποσύνολο των δεδομένων (n/10), οδηγεί σε ταχύτερη σύγκλιση, καθώς βελτιώνει την προσέγγιση του gradient με περισσότερα δεδομένα, χωρίς το μεγάλο κόστος του υπολογισμού του ολικού gradient. Καθώς το μέγεθος του batch πλησιάζει στο σύνολο των δεδομένων (nb $\approx$ n), ο αλγόριθμος αρχίζει συμπεριφέρεται περισσότερο σαν τον GD άρα η επαναλήψεις παίρνουν κοντά στον χρόνο του gd, αλλά έχουν και θόρυβο από τα δεδομένα που λείπουν με αποτέλεσμα να χειροτερεύει η απόδοση.

7 Problem-07

A.

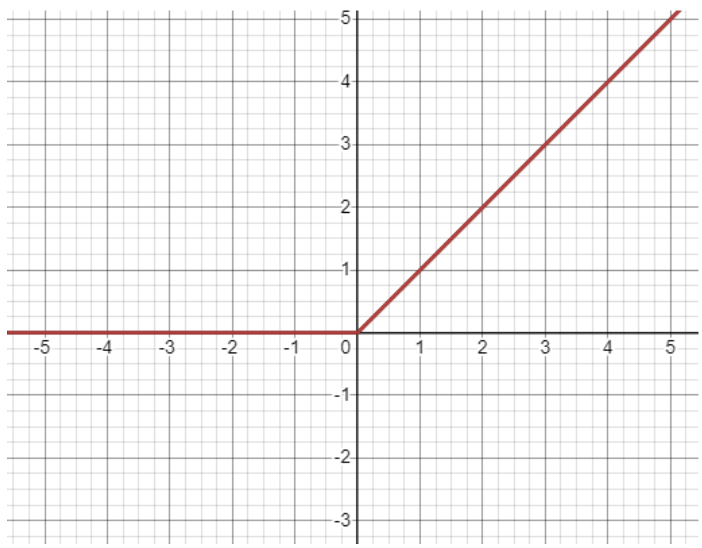
$$S(x) = \begin{cases} x^k & x > L \\ x^k \dfrac{(L+x)^m}{(L+x)^m + (L-x)^m} & |x| \leq L \\ 0 & x < -L \end{cases}$$

$$\text{ReLU} = \begin{cases} x & x > 0 \\ 0 & x \leq 0 \end{cases}$$

Εύκολα μπορούμε να δούμε ότι αν στην S(x) αντικαταστήσουμε: L = 0, k = 1 και m = 1, τότε μπορούμε να προσεγγίσουμε την ReLU:

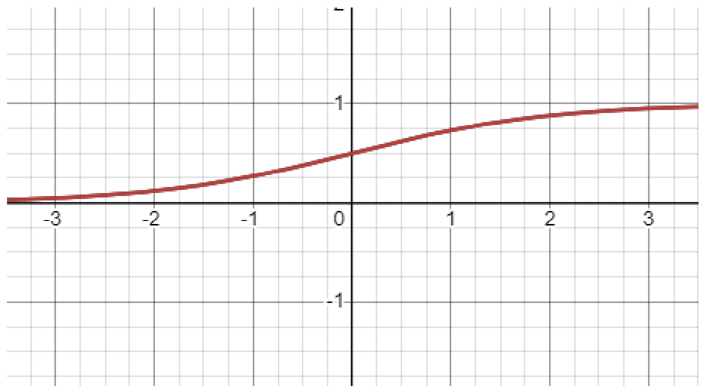
$$S(x) = \begin{cases} x^1 & x > 0 \\ x^1 \dfrac{(0+x)^1}{(0+x)^1 + (0-x)^1} & |x| \leq 0 \\ 0 & x < 0 \end{cases} = \begin{cases} x & x > 0 \\ x \dfrac{x}{x-x} & |x| \leq 0 \\ 0 & x < 0 \end{cases} = \begin{cases} x & x > 0 \\ undefined & x = 0 \\ 0 & x < 0 \end{cases}$$

Συνεπώς με αυτές τις τιμές των hyperparameters η S προσεγγίζει την ReLU για όλο το R εκτός από το 0, όπου η S δεν ορίζεται. Επιπλέον μπορούμε να διαπιστώσουμε την προσέγγιση εάν κοιτάξουμε τις γραφικές παραστάσεις των δύο συναρτήσεων.



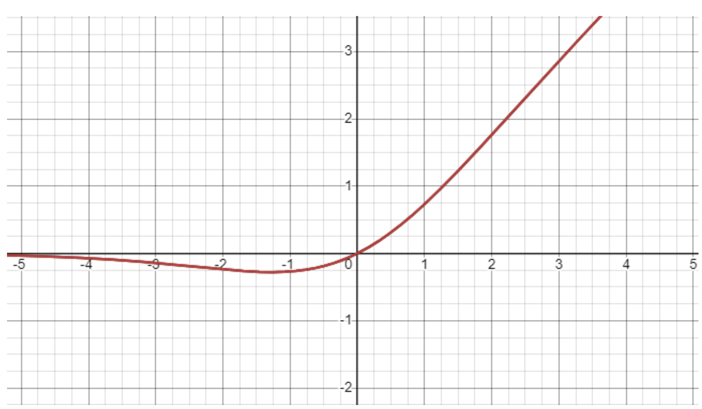
$$\text{Sigmoid} = \frac{1}{1+e^{-x}}$$

Στην περίπτωση της sigmoid μπορούμε από την γραφική παράσταση δοκιμάζοντας διάφορες τιμές των παραμέτρων της S να συμπεράνουμε ότι για k = 0 γίνεται μια προσομοίωση της ασύμπτωτης της sigmoid στην y=1. Η παράμετρος L θα καθορίσει το διάστημα που θα γίνει η προσομοίωση του 'S' ,που παρατηρούμε στην sigmoid, πειραματικά από την γραφική παράσταση διαπιστώσαμε ότι μια καλή προσέγγιση γίνεται όταν τηρείτε η αναλογία L = 2m και όταν οι τιμές αυτών των παραμέτρων είναι στα διαστήματα 0<m<115 και 0<L<230. Για τις υψηλότερες τιμές τους (m=115 και L=230) έχουμε την καλύτερη προσέγγιση. Τέλος να σημειωθεί ότι η S(x) δεν θα ορίζεται για για x=-L και x=L.



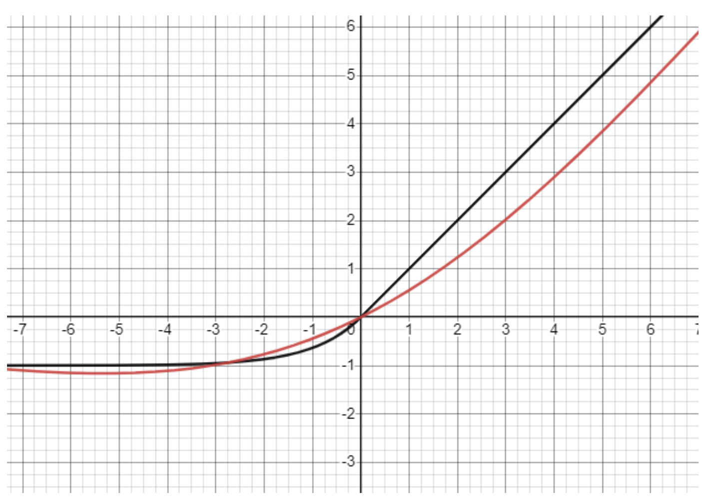
$$\text{Swish} = \frac{x}{1+e^{-x}} = x \cdot sigmoid(x)$$

Ομοίως όπως με την sigmoid, έτσι και με την swish βάση της γραφικής παράστασης παρατηρούμε ότι για θετικές τιμές η swish γίνεται y=x. Για να το προσημειώσουμε αυτό χρειαζόμαστε το k=1 ώστε s(x) = x. Για το υπόλοιπο κομμάτι της συνάρτησης μπορούμε να διαπιστώσουμε ότι αν θέσουμε L=2m για τις ίδιες τιμές όπως και πριν με τη sigmoid (m=115 και L=230) τότε θα έχουμε μια καλή αναπαράσταση swish χρησιμοποιώντας την S.



$$\text{ELU} = \begin{cases} x & x > 0 \\ e^x - 1 & x \leq 0 \end{cases}$$

Για την ELU παρατηρούμε ότι για αρνητικές τιμές εισόδου βγάζει αρνητικές εξόδους κοντά στο -1, ενώ για θετικές έχουμε y=x. Η S αριστερά από διάστημα -L<x<L έχει τιμή 0, πράγμα που καθιστά αδύνατη την ακριβή αναπαράσταση της ELU, ενώ δεξιά του διαστήματος αυτού θα αναπαριστά τη y=x (για k=1) αφού και η ELU για x>0: x. Στο διάστημα -L<x<L και συγκεκριμένα στο αρνητικό κομμάτι για m>0 η S γίνεται αρνητική παρατηρώντας μια καμπύλη που αυξομειώνεται αναλόγως το m, όσο το m μεγαλώνει τόσο μικραίνει το εμβαδόν της S στον αρνητικό άξονα και τόσο μειώνεται η του ελαχίστου. Δεν υπάρχει συνδυασμός m,L που να δημιουργεί ασύμπτωτη στο -1(η σε οποιαδήποτε τιμή). Για τις τιμές m και L καταλήξαμε σε m=30 και L=250



B.

Παράγωγος της S ως προς x

Για $x < -L : S(x) = 0 \Leftrightarrow \dfrac{\partial S}{\partial x} = 0$

Για $x > L : S(x) = x^k \Leftrightarrow \dfrac{\partial S}{\partial x} = kx^{k-1}$

Για $-L \leq x \leq L : S(x) = x^k \cdot \dfrac{(L+x)^m}{(L+x)^m + (L-x)^m} \Leftrightarrow$

$$\begin{aligned}\frac{\partial S}{\partial x} &= kx^{k-1}\frac{(L+x)^m}{(L+x)^m + (L-x)^m} \\ &+ x^k \frac{m(L+x)^{m-1}[(L+x)^m + (L-x)^m] - (L+x)^m[m(L+x)^{m-1} - m(L-x)^{m-1}]}{[(L+x)^m + (L-x)^m]^2} \\ &= kx^{k-1}\frac{(L+x)^m}{(L+x)^m + (L-x)^m} \\ &+ x^k \cdot m\frac{(L+x)^{m-1}(L+x)^m + (L+x)^{m-1}(L-x)^m - (L+x)^m(L+x)^{m-1} + (L+x)^m(L-x)^{m-1}}{[(L+x)^m + (L-x)^m]^2} \\ &= kx^{k-1}\frac{(L+x)^m}{(L+x)^m + (L-x)^m} + mx^k\frac{(L+x)^{m-1}(L-x)^m + (L+x)^m(L-x)^{m-1}}{[(L+x)^m + (L-x)^m]^2}\end{aligned}$$

Παράγωγος της S ως προς k

Για $x < -L : S(x) = 0 \Leftrightarrow \dfrac{\partial S}{\partial x} = 0$

Για $x > L : S(x) = x^k \Leftrightarrow \dfrac{\partial S}{\partial x} = x^k \ln x$

Για $-L \leq x \leq L : S(x) = x^k \cdot \dfrac{(L+x)^m}{(L+x)^m + (L-x)^m} \Leftrightarrow$

$$\frac{\partial S}{\partial x} = x^k \ln x \frac{(L+x)^m}{(L+x)^m + (L-x)^m}$$

Παράγωγος της S ως προς L

Για $x < -L : S(x) = 0 \Leftrightarrow \dfrac{\partial S}{\partial x} = 0$

Για $x > L : S(x) = x^k \Leftrightarrow \dfrac{\partial S}{\partial x} = 0$

Για $-L \leq x \leq L : S(x) = x^k \cdot \dfrac{(L+x)^m}{(L+x)^m + (L-x)^m} \Leftrightarrow$

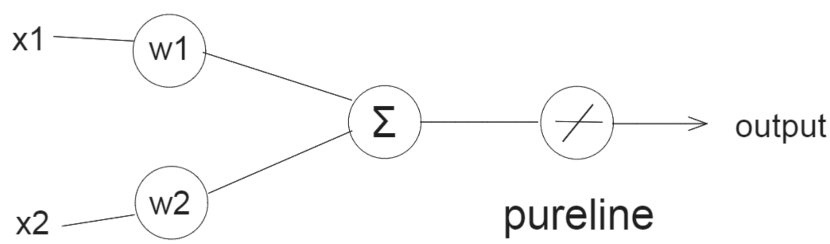
$$\begin{aligned}\frac{\partial S}{\partial x} &= x^k \frac{m(L+x)^{m-1}[(L+x)^m + (L-x)^m] - (L+x)^m[m(L+x)^{m-1} + m(L-x)^{m-1}]}{[(L+x)^m + (L-x)^m]^2} \\ &= x^k \cdot m\frac{(L+x)^{m-1}(L+x)^m + (L+x)^{m-1}(L-x)^m - (L+x)^m(L+x)^{m-1} - (L+x)^m(L-x)^{m-1}}{[(L+x)^m + (L-x)^m]^2}\end{aligned}$$

Παράγωγος της S ως προς m

$$\text{Για } x < -L : S(x) = 0 \Leftrightarrow \frac{\partial S}{\partial x} = 0$$
$$\text{Για } x > L : S(x) = x^k \Leftrightarrow \frac{\partial S}{\partial x} = 0$$
$$\text{Για } -L \leq x \leq L : S(x) = x^k \cdot \frac{(L+x)^m}{(L+x)^m + (L-x)^m} \Leftrightarrow$$
$$\frac{\partial S}{\partial x} = x^k \frac{\ln(L+x)(L+x)^m[(L+x)^m + (L-x)^m] - (L+x)^m[\ln(L+x)(L+x)^m + \ln(L-x)(L-x)^m]}{[(L+x)^m + (L-x)^m]^2}$$
$$= x^k \frac{\ln(L+x)(L+x)^m(L+x)^m + \ln(L+x)(L+x)^m(L-x)^m - \ln(L+x)(L+x)^m(L+x)^m}{[(L+x)^m + (L-x)^m]^2}$$
$$- \frac{\ln(L-x)(L+x)^m(L-x)^m}{[(L+x)^m + (L-x)^m]^2}$$
$$= x^k \frac{\ln(L+x)(L+x)^m[(L-x)^m - \ln(L-x)(L+x)^m(L-x)^m}{[(L+x)^m + (L-x)^m]^2}$$

8 Problem-08

A.



B.

Αρχικά θα υπολογίσουμε το MSE μέσα από την συνάρτηση

$$F(x) = c - 2x^T h + x^T R x$$

όπου

$$c = E[t^2], \quad h = E[tz], \quad R = E[zz^T]$$

$$c = E[t^2] = t_1^2 \cdot prob_1 + t_2^2 \cdot prob_2 + t_3^2 \cdot prob_4 = 26^2 \cdot 0.2 + 26^2 \cdot 0.7 + (-26)^2 \cdot 0.1$$
$$= 26^2(0.2 + 0.7 + 0.1) = 676$$

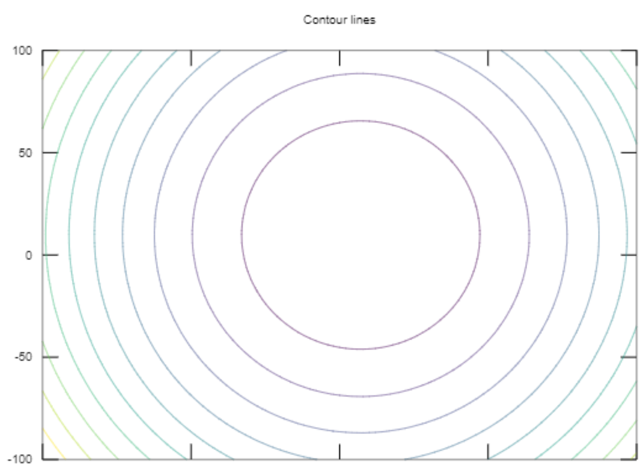
$$h = E[tz] = prob_1 \cdot t_1 \cdot p_1 + prob_2 \cdot t_2 \cdot p_2 + prob_3 \cdot t_3 \cdot p_3$$
$$= 0.2\cdot 26 \begin{bmatrix} 2 \\ 4 \end{bmatrix} + 0.7\cdot 26 \begin{bmatrix} 4 \\ 2 \end{bmatrix} + 0.1\cdot(-26) \begin{bmatrix} -2 \\ -2 \end{bmatrix} = \begin{bmatrix} 10.4 \\ 20.8 \end{bmatrix} + \begin{bmatrix} 72.8 \\ 36.4 \end{bmatrix} + \begin{bmatrix} 5.2 \\ 5.2 \end{bmatrix} = \begin{bmatrix} 88.4 \\ 62.4 \end{bmatrix}$$

$$R = E[zz^T] = p_1\cdot p_1^T \cdot prob_1 + p_2\cdot p_2^T \cdot prob_2 + p_3\cdot p_3^T \cdot prob_3 = \begin{bmatrix} 2 \\ 4 \end{bmatrix} \begin{bmatrix} 2 & 4 \end{bmatrix} 0.2 + \begin{bmatrix} 4 \\ 2 \end{bmatrix} \begin{bmatrix} 4 & 2 \end{bmatrix} 0.7$$
$$+ \begin{bmatrix} -2 \\ -2 \end{bmatrix} \begin{bmatrix} -2 & -2 \end{bmatrix} 0.1 = \begin{bmatrix} 0.8 & 1.6 \\ 1.6 & 3.2 \end{bmatrix} + \begin{bmatrix} 11.2 & 5.6 \\ 5.6 & 2.8 \end{bmatrix} + \begin{bmatrix} 0.4 & 0.4 \\ 0.4 & 0.4 \end{bmatrix} = \begin{bmatrix} 12.4 & 7.6 \\ 7.6 & 6.4 \end{bmatrix}$$

$$F(x) = 6.76 - 2 \begin{bmatrix} w_1 & w_2 \end{bmatrix} \begin{bmatrix} 88.4 \\ 62.4 \end{bmatrix} + \begin{bmatrix} w_1 & w_2 \end{bmatrix} \begin{bmatrix} 12.4 & 7.6 \\ 7.6 & 6.4 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$
$$= 676 - 176.8w_1 - 124.8w_2 + 12.4w_1^2 + 6.4w_2^2 + 15.2w_1w_2$$

Κάνουμε plot τα contour lines

```
[w1, w2] = meshgrid(-100:1:100);
f = 676 - 176.8*w1 - 124.8*w2 + 12.4*w1.^2 + 6.4*w2.^2 + 15.2*w1*w2
contour(w1, w2, f)
title("Contour lines")
```



C.

Για το optimal decision boundary πρέπει να λύσουμε για το ελάχιστο w^*

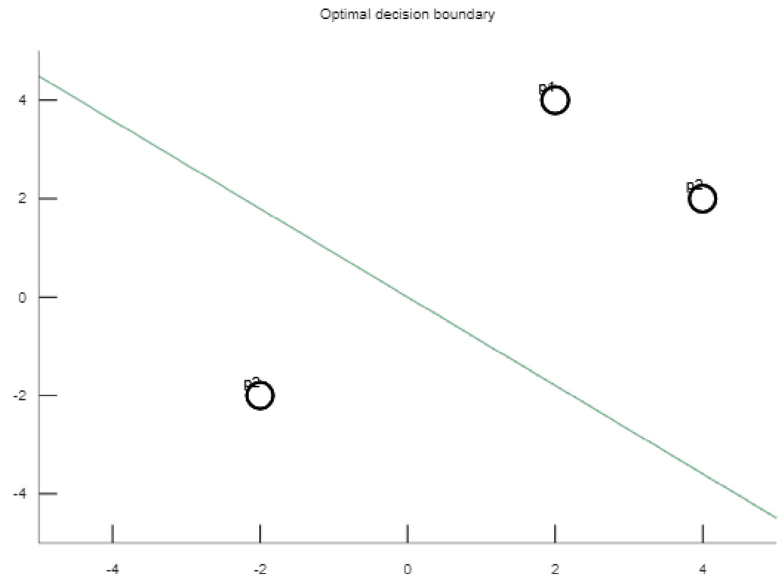
$$x^* = R^{-1}h = \begin{bmatrix} 12.4 & 7.6 \\ 7.6 & 6.4 \end{bmatrix}^{-1} \begin{bmatrix} 88.4 \\ 62.4 \end{bmatrix} = \begin{bmatrix} \frac{6.4}{21.6} & -\frac{7.6}{21.6} \\ -\frac{7.6}{21.6} & \frac{12.4}{21.6} \end{bmatrix} \begin{bmatrix} 88.4 \\ 62.4 \end{bmatrix} = \begin{bmatrix} \frac{91.52}{21.6} \\ \frac{101.92}{21.6} \end{bmatrix}$$

$$w^T \cdot p + b = 0 \Rightarrow \begin{bmatrix} \frac{91.52}{21.6} & \frac{101.92}{21.6} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} + 0 = 0 \Rightarrow \frac{91.52}{21.6} w_1 + \frac{101.92}{21.6} w_2 = 0$$
$$\Rightarrow \frac{101.92}{21.6} w_2 = -\frac{91.52}{21.6} w_1 101.92 * w_2 = -91.52 * w_1 \Rightarrow w_2 = -\frac{91.52}{101.92} w_1$$

Κάνουμε plot το decision boundary και τα δυο σημεία

```
x = linspace(-5, 5);
y = -91.52/101.92 * x;

hold on
axis([-5 5 -5 5])
plot(x, y)
plot(2, 4, "o")
plot(4, 2, "o")
plot(-2, -2, "o")
text(2, 4, 'p1', 'VerticalAlignment', 'bottom', 'HorizontalAlignment', 'right')
text(4, 2, 'p2', 'VerticalAlignment', 'bottom', 'HorizontalAlignment', 'right')
text(-2, -2, 'p2', 'VerticalAlignment', 'bottom', 'HorizontalAlignment', 'right')
title("Optimal decision boundary")
hold off
```

Παρατηρούμε ότι η ευθεία χωρίζει τα patterns στις κατάλληλλες κατηγορίες



D.

Θα υπολογίσουμε τις ιδιοτιμές του R = $\begin{bmatrix} 12.4 & 7.6 \\ 7.6 & 6.4 \end{bmatrix}$
Από τον ορισμό του ιδιοδιανύσματος της αντίστοιχης ιδιοτιμής έχουμε
$Rv = \lambda v \Rightarrow Rv - \lambda v = 0 \Rightarrow (R - \lambda I) \cdot v = 0.$
Η εξίσωση έχει μη μηδενικές ρίζες εάν και μόνο εάν $det(R - \lambda I) = 0$

$$det(R - \lambda I) = \begin{bmatrix} 12.4 - \lambda & 7.6 \\ 7.6 & 6.4 - \lambda \end{bmatrix} = (12.4 - \lambda) \cdot (6.4 - \lambda) - 7.6 \cdot 7.6 =$$
$$79.36 - 12.4\lambda - 6.4\lambda + \lambda^2 - 57.76 = \lambda^2 - 18.8\lambda + 21.6$$

Άρα έχουμε

$$\lambda^2 - 18.8\lambda + 21.6 = 0 \Rightarrow (\lambda + \frac{\sqrt{1669} - 47}{5})(\lambda - \frac{\sqrt{1669} + 47}{5}) = 0$$

Και για να βρούμε τις ρίζες

$$\lambda_1 + \frac{\sqrt{1669} - 47}{5} = 0 \Rightarrow \lambda_1 = -\frac{\sqrt{1669} - 47}{5} \Rightarrow \lambda_1 = \frac{47 - \sqrt{1669}}{5}$$

$$\lambda_2 - \frac{\sqrt{1669} + 47}{5} = 0 \Rightarrow \lambda_2 = \frac{\sqrt{1669} + 47}{5} \Rightarrow \lambda_2 = \frac{47 + \sqrt{1669}}{5}$$

Για να υπολογίσουμε το maximum stable learning rate θα επιλέξουμε το μέγιστο λ .

$\lambda_{max} = \dfrac{47 + \sqrt{1669}}{5}$
Από θεωρία γνωρίζουμε ότι:

$$0 < \alpha < \frac{1}{\lambda max} \Rightarrow 0 < \; < \frac{1}{\lambda max} \Rightarrow 0 < \alpha < \frac{1}{\frac{47+\sqrt{1669}}{5}}$$

Άρα $a_{max} = \dfrac{5}{47 + \sqrt{1669}}$ = 0,0569
Εάν τα target values αλλάξουν ο μόνος πίνακας που επηρεαστεί είναι ο c, οι πίνακες h και R θα παραμείνουν ίδιοι. Όπως είδαμε για τον υπολογισμό του maximum stable learing rate χρειάστηκε να βρούμε τις ιδιοτιμές του R. Αφού η αλλαγή των target values δεν επηρεάζει τον πίνακα R, τότε δεν επηρεάζει και την τιμή του maximum stable learning rate. Συνεπώς θα μείνει ίδιος.

E.

Για το συγκεκριμένο network η έξοδος είναι: output = pureline(W*b)
Για ένα iteration LMS με είσοδο το p1 και όλα τα βάρη μηδέν έχουμε:

$$output(0) = pureline(W(0) \cdot p1) = pureline(\begin{bmatrix} 0 \; 0 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 4 \end{bmatrix}) = 0$$

$$e(0) = t1 - output(0) = 26 - 0 = 26$$

$$W(1) = W(0) + 2ae(0)p1^T = \begin{bmatrix} 0 \; 0 \end{bmatrix} + 2 \cdot 0.05 \cdot 26 \cdot \begin{bmatrix} 2 \; 4 \end{bmatrix} = \begin{bmatrix} 0 \; 0 \end{bmatrix} + \begin{bmatrix} 5.2 \; 10.4 \end{bmatrix} = \begin{bmatrix} 5.2 \; 10.4 \end{bmatrix}$$

9 Problem-09

A.

Τα διανύσματα είναι equipropable αρα το propability του vector p1 είναι ίδιο με το propability του vector p2 ίσο με $\frac{1}{2} = 0.5$
Απο τις διαφανεις εχουμε

$$F(x) = c - 2x^T h + x^T R x$$

όπου

$$c = E[t^2], \quad h = E[tz], \quad R = E[zz^T]$$

$$c = E[t^2] = t_1^2 \cdot prob_1 + t_2^2 \cdot prob_2 = (-1)^2 \cdot 0.5 + 1^2 \cdot 0.5 = 1$$

$$h = E[tz] = prob_1 \cdot t_1 \cdot p_1 + prob_2 \cdot t_2 \cdot p_2 = 0.5 \cdot (-1) \begin{bmatrix} 1 \\ 2 \end{bmatrix} + 0.5 \cdot 1 \cdot \begin{bmatrix} -2 \\ 1 \end{bmatrix} = \begin{bmatrix} -0.5 \\ -1 \end{bmatrix} + \begin{bmatrix} -1 \\ 0.5 \end{bmatrix}$$
$$= \begin{bmatrix} -1.5 \\ -0.5 \end{bmatrix}$$

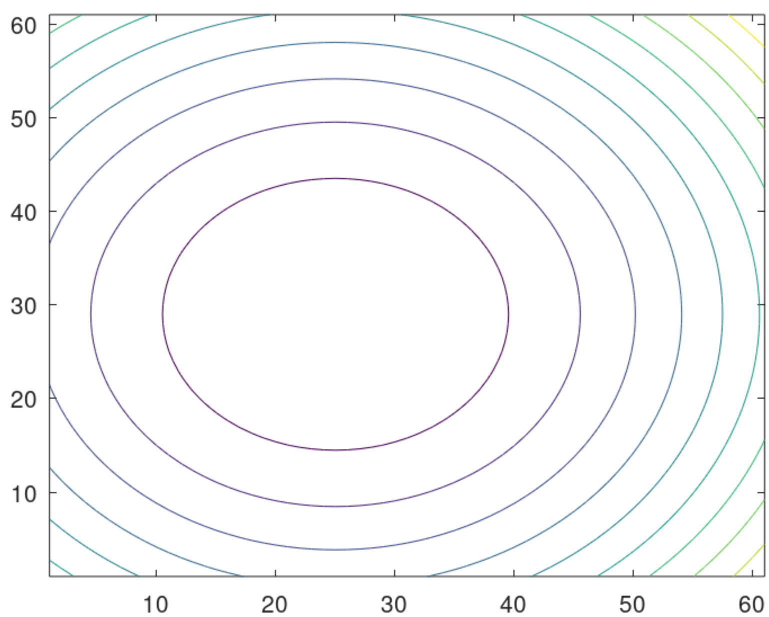
$$R = E[zz^T] = p_1 \cdot p_1^T \cdot prob_1 + p_2 \cdot p_2^T \cdot prob_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \begin{bmatrix} 1 & 2 \end{bmatrix} 0.5 + \begin{bmatrix} -2 \\ 1 \end{bmatrix} \begin{bmatrix} -2 & 1 \end{bmatrix} 0.5$$
$$= \begin{bmatrix} 0.5 & 1 \\ 1 & 2 \end{bmatrix} + \begin{bmatrix} 2 & -1 \\ -1 & 0.5 \end{bmatrix} = \begin{bmatrix} 2.5 & 0 \\ 0 & 2.5 \end{bmatrix}$$

Άρα αντικαθιστούμε

$$F(x) = 1 - 2 \begin{bmatrix} w_{1,1} & w_{2,2} \end{bmatrix} \begin{bmatrix} -1.5 \\ -0.5 \end{bmatrix} + \begin{bmatrix} w_{1,1} & w_{2,2} \end{bmatrix} \begin{bmatrix} 2.5 & 0 \\ 0 & 2.5 \end{bmatrix} \begin{bmatrix} w_{1,1} \\ w_{2,2} \end{bmatrix} = 1 + 3w_{1,1} + w_{1,2} + 2.5w_{1,1}^2 + 2.5w_{1,2}^2$$

Κάνουμε plot τα contour lines

```
[w1, w2] = meshgrid(-4:0.1:4);
 f = 1 + 3*w1 +w2 +2.5*(w1.^2 + w2.^2)
contour(w1, w2, f)
title("Contour lines")
```



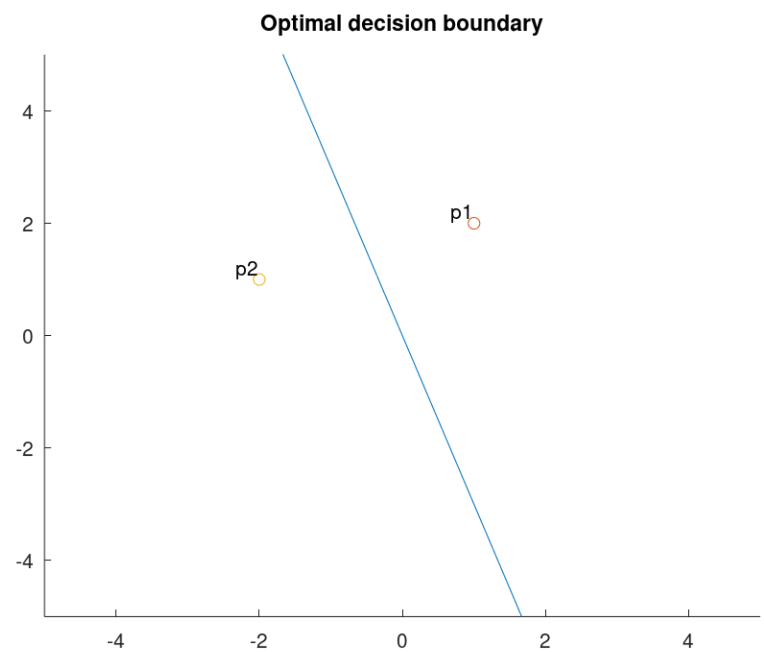
B.

$$x^* = R^{-1}h = \begin{bmatrix} 2.5 & 0 \\ 0 & 2.5 \end{bmatrix}^{-1} \begin{bmatrix} -1.5 \\ -0.5 \end{bmatrix} = \begin{bmatrix} -0.6 \\ -0.2 \end{bmatrix}$$

$$w^T \cdot p + b = 0 \Rightarrow \begin{bmatrix}-0.6 & -0.2\end{bmatrix} \begin{bmatrix}w_{1,1} \\ w_{1,2}\end{bmatrix} + 0 = 0 \Rightarrow -0.6w_{1,1} - 0.2w_{1,2} = 0 \Rightarrow w_{1,2} = -3w_{1,1}$$

Κανουμε plot το decision boundary και τα δυο σημεια

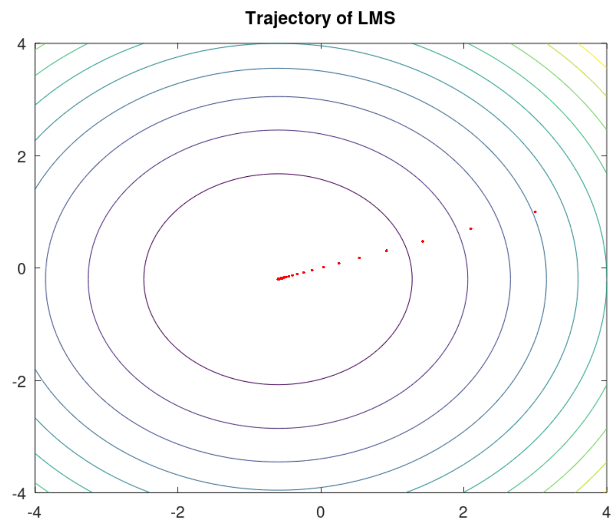
```
x = linspace(-5, 5);
y = -3 * x;

hold on
axis([-5 5 -5 5])
plot(x, y)
plot(1, 2, "o")
plot(-2, 1, "o")
text(1, 2, 'p1', 'VerticalAlignment', 'bottom', 'HorizontalAlignment', 'right')
text(-2, 1, 'p2', 'VerticalAlignment', 'bottom', 'HorizontalAlignment', 'right')
title("Optimal decision boundary")
hold off
```



C.

Παρατηρουμε οτι το trajectory ξεκινα απο (3,1) και ειναι καθετο στα contour lines καθως προσεγιζει το gradient descent και συγκλινει στο ελαχιστο που υπολογισαμε (-0.6, -0.2) μετα απο 78 επαναληψεις (39 επι 2 εισοδους).

```
% Generating grid for contour plot
[w1, w2] = meshgrid(-4:0.1:4);
f = 1 + 3 * w1 + w2 + 2.5 * (w1.^2 + w2.^2);

contour(w1, w2, f);
title("Trajectory of LMS");
hold on;

% Initialization
p = [1 -2; 2 1];
t = [-1 1];
alpha = 0.025;
w = [3; 1];
plot(3, 1, "r.");

% Training loop
for step = 1:40
    for i = 1:2
        a = dot(w', p(:, i));
        error = t(i) - a;
        w = w + 2 * alpha * error * p(:, i);
    end
    plot(w(1), w(2), "r.");
end

disp(w(1));
disp(w(2));
hold off;
```



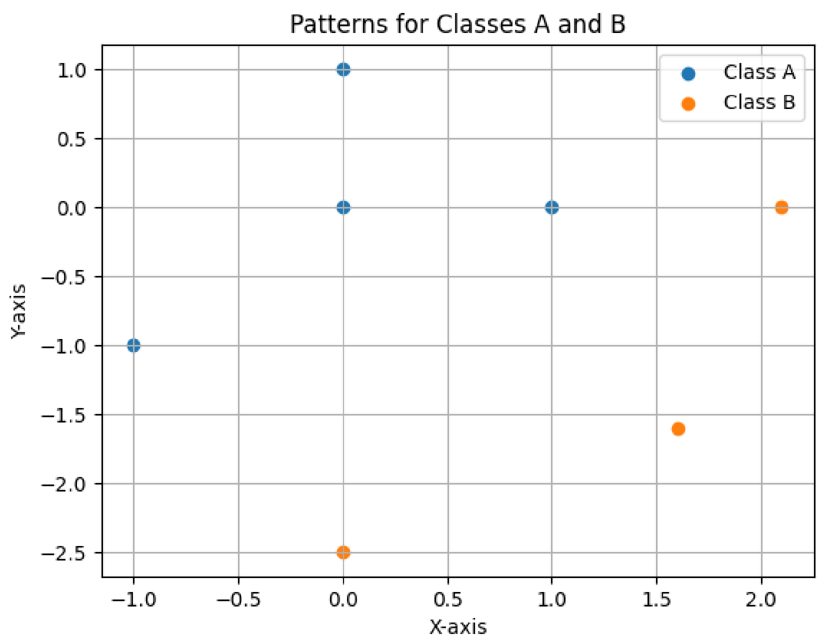
10 Problem-10

A.

Κανουμε plot τα patterns και παρατηρουμε οτι ειναι γραμμικα διαχωρισιμα.

```
import matplotlib.pyplot as plt

# Patterns for class A
class_a = [(0, 0), (0, 1), (1, 0), (-1, -1)]

# Patterns for class B
class_b = [(2.1, 0), (0, -2.5), (1.6, -1.6)]

# Unpack x and y coordinates for plotting
x_a, y_a = zip(*class_a)
x_b, y_b = zip(*class_b)

# Plotting class A and class B points
plt.scatter(x_a, y_a, label='Class A')
plt.scatter(x_b, y_b, label='Class B')
plt.xlabel('X-axis')
plt.ylabel('Y-axis')
plt.legend()
```

```
        plt.title('Patterns for Classes A and B')
        plt.grid(True)
        plt.show()
```



B.

Αφού έχουμε 2 κλάσεις διανυσμάτων εισόδου θα χρειαστουμε ενα adaline με εναν νευυρωνα επειδη S νευρωνες μπορουν να κατηγοριοποιησουν 2^S . Ως εισοδο θα εχουμε τα x,y του pattern και θα χρησιμοποιησουμε target vectors -1 για την κλαση A και 1 για την κλαση B. Επομενως το training set θα ειναι το:

$$\left\{p_1 = \begin{bmatrix}0\\0\end{bmatrix}, t_1 = -1\right\}\left\{p_2 = \begin{bmatrix}0\\1\end{bmatrix}, t_2 = -1\right\}\left\{p_3 = \begin{bmatrix}1\\0\end{bmatrix}, t_3 = -1\right\}$$

$$\left\{p_4 = \begin{bmatrix}-1\\-1\end{bmatrix}, t_4 = -1\right\}\left\{p_5 = \begin{bmatrix}2.1\\0\end{bmatrix}, t_5 = 1\right\}\left\{p_6 = \begin{bmatrix}0\\-2.5\end{bmatrix}, t_3 = 1\right\}$$

$$\left\{p_7 = \begin{bmatrix}1.6\\-1.6\end{bmatrix}, t_3 = 1\right\}$$

C.

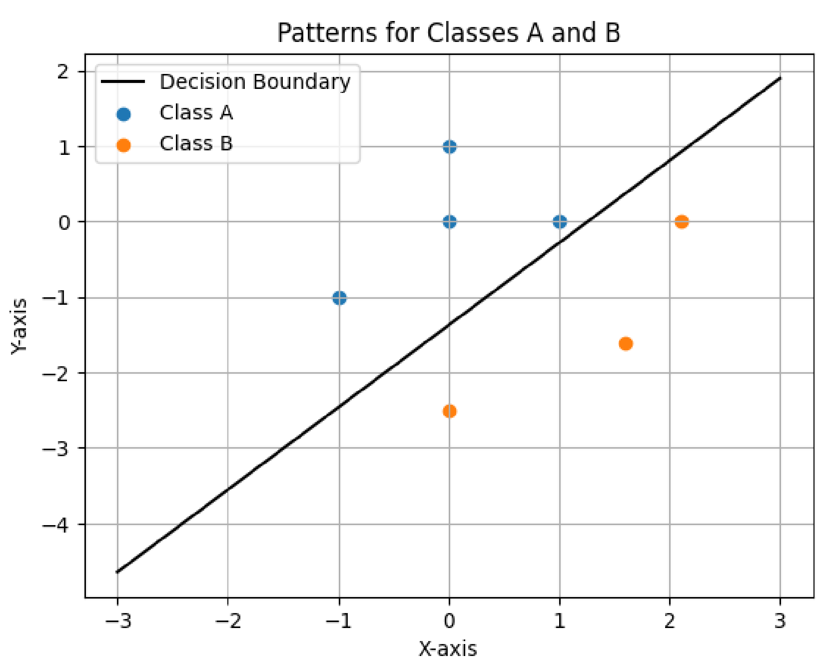
Χρησιμοποιουμε των κωδικα απο την ασκηση 9 για να τρέξουμε τον LMS αλγοριθμο. Παρατηρουμε οτι συγκλινει μετα απο 36 επαναλήψεις. Χρησιμοποιωντας τις τιμες απο τα τα βαρη και το bias που βρηκαμε, σχηματιζουμε μια ευθεία η οποία χωρίζει τις δυο κλασεις αποτελεσματικά. Τελος κανουμε plot και πάλι τα σημεια, μαζι με την ευθεία που το ομαδοποιεί.

```
% Initialization
p = [0 0 1 -1 2.1 0 1.6; 0 1 0 -1 0 -2.5 -1.6];
t = [-1 -1 -1 -1 1 1 1];
alpha = 0.04;
w = [0.5; 0.5];
b = 0.5;

% Training loop
for step = 1:36
    for i = 1:7
        a = dot(w', p(:, i)) + b;
        error = t(i) - a;
        w = w + 2 * alpha * error * p(:, i);
        b = b + 2 * alpha * error;
    end
end

disp(w(1));
disp(w(2));
disp(b);
```

Τα αποτελεσματα που βρηκαμε ειναι $w = \begin{pmatrix} 0.5989 \\ -0.5556 \end{pmatrix}$ και $b = -0.7187$ Επειτα κανουμε plot τα patetrns και το decion boundary που ειναι η ευθεια $y = \dfrac{-w_0 * x - b}{w_1}$

```
import numpy as np
import matplotlib.pyplot as plt

# Patterns for class A
class_a = [(0, 0), (0, 1), (1, 0), (-1, -1)]

# Patterns for class B
class_b = [(2.1, 0), (0, -2.5), (1.6, -1.6)]

# Unpack x and y coordinates for plotting
x_a, y_a = zip(*class_a)
x_b, y_b = zip(*class_b)
```

```
# Plotting decision boundary (a line)
w0 = 0.5989
w1 = -0.5556
b = -0.7187
x_decision = np.linspace(-3, 3, 100)  # X-axis range
y_decision = (-w0 * x_decision - b) / w1  # Decision boundary equation
plt.plot(x_decision, y_decision, label='Decision Boundary', color='black')

# Plotting class A and class B points
plt.scatter(x_a, y_a, label='Class A')
plt.scatter(x_b, y_b, label='Class B')
plt.xlabel('X-axis')
plt.ylabel('Y-axis')
plt.legend()
plt.title('Patterns for Classes A and B')
plt.grid(True)
plt.show()
```



11 Problem-11

A.

Το S ειναι ενα fuzzy set με membership function $\mu_{\tilde{S}(x)} = x$ και Very S με $\mu_{Ve\tilde{r}yS(x)} = x^2$ Για να ισχυει το statement : Let "S" be a fuzzy set. Then "Very S" is a fuzzy subset of "S" πρεπει $\mu_{Ve\tilde{r}yS(x)} \leq \mu_{\tilde{S}(x)}$ για καθε x στο $0 \leq x \leq 1$. Αρα το statement ειναι true αφου

$$x \leq 1$$

πολλαπλασιάζω με x και επειδή $x \geq 0$ παμενει η φορα της ανισωσης

$$x^2 \leq x \Leftrightarrow \mu_{Ve\tilde{r}yS(x)} \leq \mu_{\tilde{S}(x)}$$

B.

Εστω S' το fuzzy set more or less S αρα $\mu_{\tilde{S'}(x)} = \sqrt{x}$. Εχουμε

$$x \leq 1 \Leftrightarrow x^2 \leq x \Leftrightarrow \sqrt{x^2} \leq \sqrt{x} \Leftrightarrow |x| \leq \sqrt{x} \Leftrightarrow x \leq \sqrt{x} \Leftrightarrow \mu_{\tilde{S}(x)} \leq \mu_{\tilde{S'}(x)}$$

αρα $S \subseteq$ "more or less S" αρα το statement ειναι true.

C.

Για να ισχυει το statement πρεπει $1 - x^2 \leq \sqrt{x}$ ομως για x = 0.36 εχουμε $1 - x^2 = 1 - 0.36^2 = 1 - 0.1296 = 0.8704$ Που ειναι μεγαλυτερο απο το $\sqrt{0.36} = 0.6$ αρα το statement ειναι false.

D.

Για να ισχυει το statement πρεπει $1 - \sqrt{x} \leq x^2$ ομως για x = 0.36 εχουμε $1 - \sqrt{0.36} = 1 - 0.6 = 0.4$ που ειναι μεγαλυτερο απο το $0.36^2 = 0.1296$ αρα το statement ειναι false.

12 Problem-12

Η εκφραση not(A(x) OR B(x)) γραφεται 1 - max(A(x), B(x)). Πρωτα υπολογιζουμε το A(x) OR B(x) που ειναι το max(A(x), B(x)) αρα εχουμε:

$$max(A(x), B(x)) = \begin{cases} 1, & x \leq 2 \\ 1 - \dfrac{x-2}{3}, & 2 < x < \dfrac{29}{7} \\ \dfrac{x-3}{4}, & \dfrac{29}{7} \leq x < 7 \\ 1, & x \geq 7 \end{cases}$$

Η εκφραση 1 - max(A(x), B(x)) μεγιστοποιειται στο ελαχιστο της max(A(x), B(x)) που ειναι στο σημειο $\mathbf{x} = \dfrac{\mathbf{22}}{\mathbf{7}}$ αρα η μεγιστη τιμε αληθειας ειναι η

$$1 - \frac{\frac{22}{7} - 3}{4} = \frac{\mathbf{1}}{\mathbf{28}}$$