UNIVERSITY OF THESSALY
SCHOOL OF ENGINEERING

DEPARTMENT OF ELECTRICAL
& COMPUTER ENGINEERING

Νευρο-Ασαφής Υπολογιστική
Χειμερινό Εξάμηνο 2023-2024
Δημήτριος Κατσαρός

<u>Σειρά προβλημάτων: 1η: **ΟΜΑΔΙΚΕΣ** (2-ΑΤΟΜΩΝ) **ΕΡΓΑΣΙΕΣ**</u>

Ημέρα ανακοίνωσης: Wednesday, November 22, 2023
Προθεσμία παράδοσης: Παρασκευή, Δεκέμβριος 22, 2023

SECTION 0: Warming-up with linear algebra, calculus, and automatic control

Problem-01

Plot the contour lines of the following function: $f(x, y) = x^2 + 4xy + y^2$. Write it in quadratic form. Then, find and characterize the (local) minima/maxima of this function (Show your analytic calculations).

Problem-02

Execute two iterations of the Gradient Descent to the function $f(x_1, x_2)=(x_1+2x_2-7)^2+(2x_1+x_2-5)^2$ with initial point $\mathbf{x_0}=(-9.5, 9.5)$. Show your analytic calculations using the optimal λ in each iteration.

Problem-03

Consider the dynamical system called Henon map defined by the following recursive equation:

$$\begin{aligned} x_{k+1} &= f(x_k) \\ &= 1-ax_k^2+bx_{k-1} \end{aligned}$$

Many dynamical systems transition into chaos as we increase a control or gain parameter such as x_0. Select (a,b)=(0.3,) and use the two choices of initial conditions, $x_0=0$ and $x_0=0.00001$, to generate x_1,…. Plot the two trajectories. Are they aperiodic (chaotic) or periodic? For the series of the following questions, we need to plot the sequence of generated values and also record your observations.

A. [$x_0=0$] Test the generated sequence of x_i with b=0, and various values for a. What do you observe? chaos or some fixed-point(s)?
B. Test for various values of x_0 and b, but a being between 0 and 0.32.
C. [$x_0=0$] Test the case of values (a,b)=(0.3675, 0.3). What do you observe?
D. [$x_0=0$] Test the case of values (a,b)=(0.2, 0.4), (a,b)=(0.5, 0.4) and (a,b)=(0.9, 0.4). What do you observe?

SECTION 1: Introduction to neural networks

Problem-04

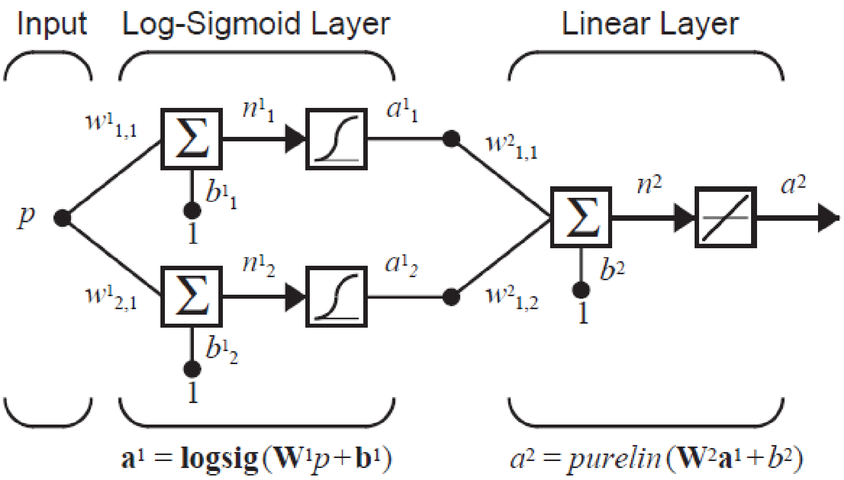
Express the derivative dS/dx, denoted as S', of the following activation functions S in terms of the original function S, i.e., determine φ such that S' = φ(S,x). [The first three functions comprise well-established activation functions, known as logsig, tansig, and

Google's Swish, respectively.] In particular, for the fourth activation function, recognize the relation of its derivative to the Swish activation function.

- $S = \dfrac{1}{1+e^{-x}}$
- $S = \dfrac{e^{x}-e^{-x}}{e^{x}+e^{-x}}$
- $S = \dfrac{x}{1+e^{-x}}$
- $S = x \times tanh(ln(1+e^{x}))$

Problem-05

Consider the following neural network



with $w^1_{1,1}=-2$, $w^1_{2,1}=-1$, $b^1_1=-0.5$, $b^1_2=-0.75$, $w^2_{1,1}=2$, $w^2_{1,2}=1$, $b^2=0.5$.
Sketch the following responses (plot the indicated variable versus p for $-2 < p < 2$).

A. a^1_1
B. a^1_2
C. a^2

Then, change the *logsig* activation function with the *Swish* activation function and sketch again the aforementioned responses.

Problem-06

We are interested in comparing the gradient descent optimizer for several values of the batch size, i.e., $n_b=1 \rightarrow SGD$, $n_b=n \rightarrow GD$, $n_b=\varphi(n) \ll n \rightarrow BGD$. While applying these variants on a given problem with an objective function *F*, we observed that when $n_b=1$, we get better convergence than $n_b=n$, while increasing the batch size from $n_b=1$ to $n_b=n/10$ consistently improves the results, in that the method converges faster and to a smaller value of the objective function *F*. We then observe that the convergence slows down as we increase the batch size from $n/10$ to n.
Explain these observations.

Problem-07

You are given the following activation function:

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

where, k, L and m are trainable parameters, i.e., hyperparameters.

A. Show (graphically and analytically) how this activation function can approximate, Swish, sigmoid, ReLU, and ELU, i.e., by proper selection of hyperparameter values.

B. Calculate the derivatives of S with respect to x, k, L and m.

SECTION 2: Working with ADALINE neural networks

Problem-08

Suppose that we have the following three reference patterns and their targets:

$$\left\{\mathbf{p}_1 = \begin{bmatrix} 2 \\ 4 \end{bmatrix}, t_1 = \begin{bmatrix} 26 \end{bmatrix}\right\}, \left\{\mathbf{p}_2 = \begin{bmatrix} 4 \\ 2 \end{bmatrix}, t_2 = \begin{bmatrix} 26 \end{bmatrix}\right\}, \left\{\mathbf{p}_3 = \begin{bmatrix} -2 \\ -2 \end{bmatrix}, t_3 = \begin{bmatrix} -26 \end{bmatrix}\right\}.$$

The probability of vector $\mathbf{p}_1$ is 0.20, the probability of vector $\mathbf{p}_2$ is 0.70, and the probability of vector $\mathbf{p}_3$ is 0.10.

A. Draw the network diagram for an ADALINE network with no bias that could be trained on these patterns.

B. Sketch the contour plot of the mean square error performance index.

C. Show the optimal decision boundary (for the weights that minimize mean square error), and verify that it separates the patterns into the appropriate categories.

D. Find the maximum stable learning rate for the LMS algorithm. If the target values are changed from 26 and -26 to 11 and -11, how would this change the maximum stable learning rate?

E. Perform one iteration of the LMS algorithm, starting with all weights equal to zero, and presenting input vector $\mathbf{p}_1$. Use a learning rate of α=0.05.

Problem-09

Suppose that we have the following two reference patterns and their targets:

$$\left\{\mathbf{p}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, t_1 = \begin{bmatrix} -1 \end{bmatrix}\right\}, \left\{\mathbf{p}_2 = \begin{bmatrix} -2 \\ 1 \end{bmatrix}, t_2 = \begin{bmatrix} 1 \end{bmatrix}\right\}.$$

The vectors are equiprobable. We want to train an ADALINE network without a bias on this data set.

A. Sketch the contour plot of the mean square error performance index.

B. Sketch the optimal decision boundary.

C. Sketch the trajectory of the LMS algorithm on your contour plot. Assume a learning rate equal to 0.025, and start with initial weights $\mathbf{W}(0) = [3 \quad 1]$.

Problem-10

The patterns (0,0), (0,1), (1,0), (-1,-1) belong to class A and the patterns (2.1,0), (0,-2,5), (1.6,-1,6) belong to class B. They are equiprobable.

A. Draw the patterns in a 2-D diagram. Are the categories linearly separable?

B. For the aforementioned patterns design the architecture of an ADALINE neural network which will separate the classes.

C. Starting with initial weights/biases equal to 0.5 calculate the final (after convergence) weights/biases of the ADALINE using the LMS training algorithm.

SECTION 3: Introduction to fuzzy logic

Problem-11

"very" is used as an adjective to reduce vagueness on fuzzy set membership. The interpretation is that if the statement "A is true" has truth value equal to x, then the statement "A is very true" has truth value x^2, because the "very true" is more demanding.

A. True or false (explain your answer): Let "S" be a fuzzy set. Then "Very S" is a fuzzy subset of "S".

"more or less" is used as an adjective to increase vagueness - the interpretation is that if "A is true" has truth value x, then "A is more or less true" has truth value $\sqrt{x}$.

B. True or false (explain your answer): Let "S" be a fuzzy set. Then "S" is a fuzzy subset of "more or less S".

C. Using the definitions just given, is it true that "not very S" is a subset of "more or less S", or vice versa, or is it impossible to say?

D. Is "not more or less S" a subset of "very S", or vice versa, or is it impossible to say?

Problem-12

Υποθέστε ότι η A(x) έχει την εξής συνάρτηση αληθείας:

$A(x) = 1$ για $x <= 2$
$A(x) = 1 - (x-2)/3$ για $2 < x < 5$
$A(x) = 0$ για $x >= 5$

και η B(x) έχει την εξής συνάρτηση αληθείας:

$B(x) = 0$ για $x <= 3$
$B(x) = (x-3)/4$ για $3 < x < 7$
$B(x) = 1$ για $x >= 7$

Βρείτε τις τιμές του x για τις οποίες έχει την μέγιστη δυνατή τιμή αληθείας η πρόταση: "not (A(x) OR B(x))".

Χρηστικές πληροφορίες:

Η προθεσμία παράδοσης είναι αυστηρή. Είναι δυνατή η παροχή παράτασης (μέχρι 4 ημέρες), αλλά μόνο αφού δώσει ο διδάσκων την έγκρισή του και αυτή η παράταση στοιχίζει 10% ποινή στον τελικό βαθμό της συγκεκριμένης Σειράς Προβλημάτων. Η παράδοση γίνεται με email στο dkatsar@e-ce.uth.gr του αρχείου λύσεων σε μορφή pdf (typeset). Το subject του μηνύματος **αυστηρά** πρέπει να είναι: CE418-Problem set 01: AEM1-AEM2

Ερμηνεία συμβόλων:

Δεν απαιτεί την χρήση υπολογιστή ή/και την ανάπτυξη κώδικα.

Απαιτεί την ανάπτυξη κώδικα σε όποια γλώσσα προγραμματισμού (C/C++, Java, …) ή pyTorch ή Matlab ή ... επιθυμείτε. Το παραδοτέο θα περιέχει:

- Την λύση της άσκησης
- Τον πηγαίο κώδικα υλοποίησης