ניסוי 56 – למידה עמוקה – דו"ח הכנה 1

ליאור וובצ'וק (207584715), אלמוג אדטו (318782976)

שאלה 1:

א. זיהוי דובר מקטע אודיו: בעיית סיווג.
מרחב הדוגמאות – קטעי אודיו שונים המתוייגים לפי הדובר.
מרחב התיוג – זהות הדוברים השונים.

ב. חיזוי שער מניה לאחר מספר ימים: בעיית רגרסיה.
מרחב הדוגמאות – גרפים המציגית התנהגות של מניה למשך זמן מסויים, לצד שער המניה לאחר מספר ימים.
מרחב התיוג – המספרים האי שליליים (שערי המניה האפשריים).

ג. זיהוי טקסט מתמונה: בעיית סיווג.
מרחב הדוגמאות – תמונות המכילות טקסט, לצד הטקסט שמופיע בתמונה.
מרחב התיוג – תווים אפשריים (אותיות, מספרים, סימנים..).

ד. מסנן דואר זבל: בעיית סיווג.
מרחב הדוגמאות – הודעות דואר שונות לצד סיווגן הבינארי כדואר זבל או לא.
מרחב התיוג – דואר זבל או לא דואר זבל (2 ערכים).

ה. גילוי מספר אנשים שנמצא בתמונה: בעיית סיווג.
מרחב הדוגמאות – תמונות לצד מספר האנשים בהן.
מרחב התיוג – המספרים השלמים האי-שליליים.

ו. חיזוי משך נסיעה בין שני יעדים: בעיית רגרסיה.
מרחב הדוגמאות – זמני נסיעה שונים בין זוגות נקודות שונות על גבי מפה, בזמנים ומצבים תחבורתיים שונים.
מרחב התיוג – המספרים האי שליליים (זמני הנסיעה האפשריים).

שאלה 2:

1. $f(x,a) = \ln(ax)$
$$\frac{\partial f}{\partial x} = \frac{1}{x}$$
$$\frac{\partial f}{\partial a} = \frac{1}{a}$$

2. $f(x,a) = (x-a)^2$
$$\frac{\partial f}{\partial x} = 2(x-a)$$
$$\frac{\partial f}{\partial a} = -2(x-a) = 2(a-x)$$

3. $f(x,a,b) = \max(b, ax)$

$$\frac{\partial f}{\partial x} = \begin{cases} a & ax > b \\ 0 & ax < b \end{cases}$$

$$\frac{\partial f}{\partial a} = \begin{cases} x & ax > b \\ 0 & ax < b \end{cases}$$

$$\frac{\partial f}{\partial b} = \begin{cases} 0 & ax > b \\ 1 & ax < b \end{cases}$$

4. $f(x,a,b) = \sqrt{\mathrm{ax+b}}$

$$\frac{\partial f}{\partial x} = \frac{a}{2\sqrt{\mathrm{ax+b}}}$$

$$\frac{\partial f}{\partial a} = \frac{x}{2\sqrt{\mathrm{ax+b}}}$$

$$\frac{\partial f}{\partial b} = \frac{1}{2\sqrt{\mathrm{ax+b}}}$$

5. $f(x) = \tanh(x)$

$$\frac{\partial f}{\partial x} = 1 - \tanh(x) = 1 - f(x)$$

שאלה 3:

$$\frac{\partial\sigma(z)}{\partial z} = \frac{\partial}{\partial z}\frac{1}{1+e^{-z}} = \frac{e^{-z}}{(1+e^{-z})^2} = \sigma(z)\cdot\frac{1+e^{-z}-1}{1+e^{-z}} = \sigma(z)\big(1-\sigma(z)\big)$$

שאלה 4:

$$\frac{\partial f}{\partial x}^{[m]} = \sum_j \frac{\partial f}{\partial (A^T x + b)_j}^{[1]} \cdot \frac{\partial (A^T x + b)_j}{\partial x}^{[m]} = \sum_j \frac{\partial f}{\partial y_j}^{[1]} \cdot A_j^{T\,[m]} = \frac{\partial f}{\partial y}^{[n]} \cdot A^{T\,[n*m]}$$

כאשר A_j^T הוא השורה ה-j של A^T.

$$\frac{\partial f}{\partial b}^{[n]} = \frac{\partial f}{\partial (A^T x + b)}^{[n]} \cdot \frac{\partial (A^T x + b)}{\partial b}^{[1]} = \frac{\partial f}{\partial y}^{[n]}$$

$$\frac{\partial f}{\partial A_{ij}}^{[1]} = \sum_K \frac{\partial f}{\partial (A^T x + b)_k}^{[1]} \cdot \frac{\partial (A^T x + b)_k}{\partial A_{ij}}^{[1]} = \sum_K \frac{\partial f}{\partial y_k}^{[1]} \cdot x_j \delta_{i,k}^{[1]} = \frac{\partial f}{\partial y_i}^{[1]} \cdot x_j^{[1]} = \left(\frac{\partial f}{\partial y}\cdot x\right)_{ij}^{[1]}$$

$$\rightarrow \quad \frac{\partial f}{\partial A}^{[n*m]} = \frac{\partial f}{\partial y}^{[n]} \cdot x^{[m]}$$

$$\frac{\partial F}{\partial X_i}^{[m]} = \sum_j \frac{\partial f(Y_j)}{\partial X_i}^{[m]} = \sum_j \frac{\partial f(Y_j)}{\partial X_i}^{[m]} \cdot \delta_{ij} = \frac{\partial f(Y_i)}{\partial X_i}^{[m]} = \sum_j \frac{\partial f(Y_i)}{\partial Y_{ji}}^{[1]} \cdot A_j^{T\,[m]}$$

$$= \sum_j \frac{\partial F(Y)}{\partial Y_{ji}}^{[1]} \cdot A_j^{T\,[m]} = \sum_j \left(\frac{\partial F}{\partial Y}\right)_{ji}^{[1]} A_j^{T\,[m]} = \left(A^{T\,[m*n]} \cdot \frac{\partial F}{\partial Y}^{[n*k]}\right)_i^{[m]}$$

$$\rightarrow \quad \frac{\partial F}{\partial X}^{[m*k]} =\cdot A^{T\,[m*n]} \frac{\partial F}{\partial Y}^{[n*k]}$$

$$\frac{\partial F}{\partial b}^{[n]} = \sum_j \frac{\partial f(Y_j)}{\partial b}^{[n]} = \sum_j \frac{\partial f(Y_j)}{\partial Y_j}^{[n]} = \sum_j \frac{\partial F(Y)}{\partial Y_j}^{[n]} = \sum_j \left(\frac{\partial F}{\partial Y}\right)_j^{[n*k]} = \frac{\partial F}{\partial Y}^{[n*k]} \cdot \mathbb{I}^{[k]}$$

$$\frac{\partial F}{\partial A}^{[n*m]} = \sum_j \frac{\partial f(Y_j)}{\partial A}^{[n*m]} = \sum_j \frac{\partial f(Y_j)}{\partial Y_j}^{[n]} \cdot X_j^{[m]} = \sum_j \frac{\partial F(Y)}{\partial Y_j}^{[n]} X_j^{[m]} = \frac{\partial F}{\partial Y}^{[n]} \cdot X^{[m]}$$

כאשר בכל איבר צויינו והוסברו מימדי המטריצות/ווקטורים במפורש.

מימד האיבר הראשון בשורה הוא המימד לו אנו מצפים מהתוצאה של אותה שורה – ואכן בכל שורה קיבלנו תוצאה התואמת לציפיותינו.

שאלה 5:

$$\left(\frac{\partial p_i}{\partial o_j}\right)_{i=j} = \frac{\partial p_i}{\partial o_i} = \frac{e^{o_i} \cdot \sum_j e^{o_j} - e^{2o_i}}{\left(\sum_j e^{o_j}\right)^2} = p_i \frac{\sum_j e^{o_j} - e^{o_i}}{\sum_j e^{o_j}} = p_i(1-p_i)$$

$$\left(\frac{\partial p_i}{\partial o_j}\right)_{i\neq j} = \frac{0 - e^{o_i+o_j}}{\left(\sum_j e^{o_j}\right)^2} = -p_i p_j$$

שאלה 6:

a. $f'(x) = 2x$
   $x_1 = -1 + 0.1 \cdot 2 = -0.8$
   $x_2 = -0.8 + 0.1 \cdot 1.6 = -0.64$
   $x_3 = -0.64 + 0.1 \cdot 1.28 = -0.512$
b. $f'(x) = 2x$
   $x_1 = -1 + 3 \cdot 2 = 5$
   $x_2 = 5 - 3 \cdot 5 = -10$
   $x_3 = -10 + 3 \cdot 20 = 50$

ג. חשוב לקחת גודל צעד גדול מספיק כדי להתכנס לנקודת האקסטרמום אחרי מספר לא גדול של צעדים, אך אין לקחת ערך גדול מדי על מנת להימנע מ-Overshoot כלומר מצב בו לא נתכנס לערך האקסטרימום.

ד. גודל הצעד המירבי הוא עד 1 (לא כולל).

נוכיח. סדרת האיטרציות שלנו היא :
$$\begin{cases} x_0 = -1 \\ x_{k+1} = x_k - \eta \frac{\partial}{\partial x} x^2 | (x = x_k) = x_k - 2\eta x_k = (1-2\eta)x_k \end{cases}$$

זהו טור גיאומטרי בעל $q = 1 - 2\eta$. לכן הטור יתכנס עבור :

$$|q| < 1 \quad \rightarrow \quad -1 < 1 - 2\eta < 1 \quad \rightarrow \quad 0 < \eta < 1$$

שאלה 7:

כל פונקציה שהרשת הפשוטה יכולה לממש, יכולה להתממש גם ברשת הארוכה יותר בכך שהשכבה הלינארית הראשונה תהיה זהה לשכבה הלינארית היחידה של הרשת הפשוטה ושאר השכבות יהיו פונקציית יחידה.

כל פונקציה שתתממש על ידי שלוש השכבות הלינאריות של הרשת הארוכה תהיה שרשור של פעולות לינאריות, ולכן בעצמה פעולה לינארית, ומכאן תהיה ניתנת למימוש על ידי שכבה לינארית בודדת.

שאלה 8:

	מעבר קדמי	נגזרת לפי פרמטרים	נגזרת לפי הכניסה
שכבה לינארית	$w^T x$	$\frac{\partial(w^T x)}{\partial w} = x$	$\frac{\partial(w^T x)}{\partial x} = w$
סיגמואיד	$\sigma(x) = \frac{1}{1+e^{-x}}$	X	$\frac{\partial \sigma(x)}{\partial x} = \sigma(x)(1-\sigma(x))$
NLL	$NLL(x) = -\sum_{i=1}^{n} (1-y_i)\log(1-x) + y_i \log(x)$	X	$\frac{\partial NLL}{\partial x} = -\sum_{i=1}^{n} \frac{1-y_i}{\mathrm{x}-1} + \frac{y_i}{x} = -\sum_{i=1}^{n} \frac{\mathrm{x}-\mathrm{y_i}}{x^2 - x}$

בכל שורה, המשתנה x הוא המוצא של השורה שלפניה, כך שלדוגמה, x בשורה השלישית שווה ל$\sigma(x)$ בשורה השנייה.

שאלה 9:

Editor - C:\Users\almoga\OneDrive - Hailo Technologies LTD\Documents\MATLAB\SIPL\irisAdaline.m

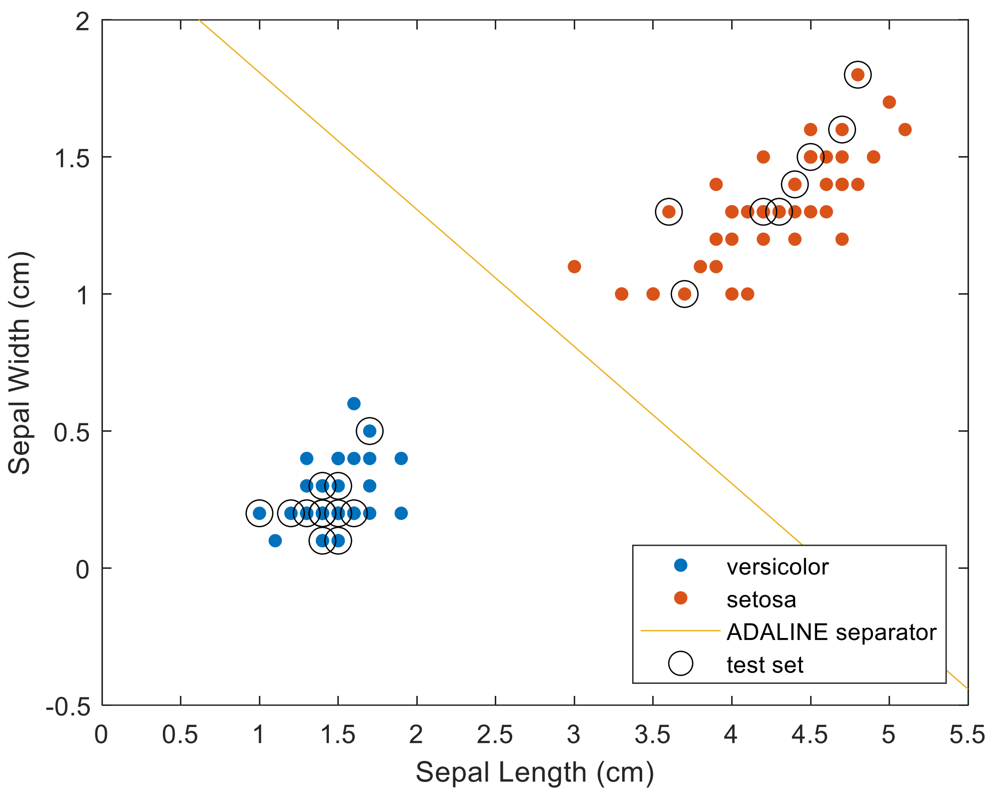
irisDataPrep.m | adaline.m | irisAdaline.m | plotIris.m | untitled.m

```matlab

%% Prepare data

trainRatio = 0.8;
[Xtrain, Ytrain, Xtest, Ytest] = irisDataPrep(meas, species, trainRatio,"sign");
X = [Xtrain,Xtest];
Y = [Ytrain,Ytest];

%% Set ADALINE's hyperparameters

numEpochs = 10;
lr0 = 0.02;

%% Train ADALINE to solve the task

[W, error] = adaline(Xtrain, Ytrain, numEpochs, lr0);
disp(['Train error rate = ', num2str(error), '%'])

%% Evaluate Test Error

testSize = size(Ytest,2);

% TODO: calculate the zero-one loss.
% i.e. on what percent of the data do we make an error
Ypred = W(1:2)'*Xtest + W(3)
error = ((Ytest - Ypred)*(Ytest - Ypred)')/size(Ytest,2);

disp(['Test error rate = ', num2str(error), '%'])
plotIris(X, Y, W, Xtest);

```

Editor - C:\Users\almoga\OneDrive - Hailo Technologies LTD\Documents\MATLAB\SIPL\adaline.m

irisDataPrep.m | adaline.m | irisAdaline.m | plotIris.m | untitled.m

```matlab

dataDim = size(X,1);
numSamples = size(X,2);

% Augment X to include a bias term
X = [X; ones(1,numSamples)];

% Initialize weights
W = 0.1*rand(dataDim+1, 1);

for kk = 1:numEpochs
    for ii = 1:numSamples
        % TODO: update W
        lr = lr0 * 0.99 ^ kk;
        W = W - lr*X(:,ii)*(W'*X(:,ii) - Y(ii));

        % You may optionally uncomment the function below to see the change
        % followed by each sample
        % plotIris(X(1:2,:),Y, W)
    end
    plotIris(X(1:2,:),Y, W)
end

% TODO: calculate the zero-one loss.
% i.e. on what percent of the data do we make an error
Ypred = W'*X;
error = ((W'*X - Y)*(W'*X - Y)')/numSamples;

end

```



2
1.5
1
0.5
0
-0.5
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
5.5
Sepal Width (cm)
Sepal Length (cm)
versicolor
setosa
ADALINE separator
test set