A Not-for-Profit University

Department of Electrical Engineering
School of Science and Engineering

EE212 Mathematical Foundations for Machine Learning and Data Science

ASSIGNMENT 4 - Solution

Due Date: 23:55, Monday, December 13, 2021. (Submit online on LMS)
Format: 4 problems, for a total of 100 marks
Instructions:

- You are allowed to collaborate with your peers but copying your colleague's solution is strictly prohibited. This is not a a group assignment. Each student must submit his/her own assignment, scanned in **a single PDF document.**
- Solve the assignment on blank A4 sheets and either scan the document using a scanner or use CamScanner proficiently.
- Upload the solved assignment on LMS in the "Assignments" tab under Assignment 01.
- Naming convention should be as follows: "Name_RollNumber_Assignment_1.pdf"
- Feel free to contact the instructor or the teaching assistants if you have any concerns.

- You represent the most competent individuals in the country, do not let plagiarism come in between your learning. In case any instance of plagiarism is detected, the disciplinary case will be dealt with according to the university's rules and regulations.

Problem 1 (15 marks)

(a) **[10 marks]** You are given the following set of data points: $(-1, 0), (0, 2), (1, 4), (2, 5)$. Find the least square regression line defined by $y = ax + b$ where:

$$a = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}$$

$$b = \frac{1}{n}(\sum_{i=1}^{n} y_i - a\sum_{i=1}^{n} x_i)$$

(b) [**5 marks**] Plot the data points and the regression line from part (a) on the same set of rectangular axes.

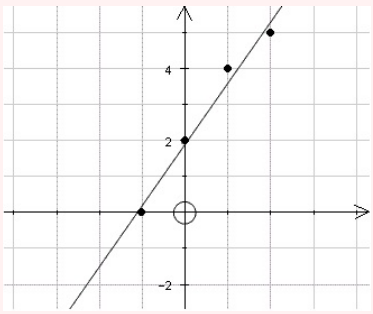
Solution.

(a)

x	y	xy	x^2
-1	0	0	1
0	2	0	0
1	4	4	1
2	5	10	4
$\sum x = 2$	$\sum y = 11$	$\sum xy = 14$	$\sum x^2 = 6$

Using the formulas provided for a and b, we get:
$a = ((4)(14) - \frac{(2)(11)}{(4)(6)-2^2} = \frac{17}{10} = 1.7$
$b = (\frac{1}{4})(11 - (1.7)(2)) = 1.9.$

(b) The regression line and given data points are depicted in the graph below:



Problem 2 (25 marks)

You are tasked with designing a straight road route that heads through Southern Punjab and provides access to four rural villages along the way. You decide to reduce the locations of these villages down to an xy-coordinate plane so that you can devise the optimal straight road that runs through them and minimizes the off-road distance from all of the villages.

In your xy-coordinate plane, Villages A, B, C, and D are located at the following data points, respectively: $(2, 1), (5, 2), (7, 3)$, and $(8, 3)$.

Find the least-squares line in the form $y = \beta_0 + \beta_1 x$ that best fits these data points.

Hint: You can write the system as $X\beta = y$, where:

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ . & . \\ . & . \\ 1 & x_n \end{bmatrix}, \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \text{ and } y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ . \\ . \\ y_n \end{bmatrix}.$$

Remember that fitting a line to experimental data is done as depicted in the figure below.

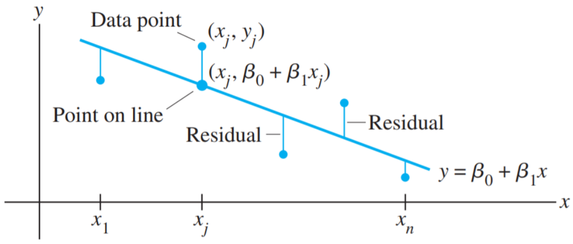


Figure 1: Least-Squares Line Fitting

Solution. Use the x-coordinates of the data to build the design matrix X and the y-coordinates to build the observation vector y.

$$X = \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}.$$

For the least-squares solution of $X\beta = y$, obtain the normal equations (with the new notation):

$$X^T X\beta = X^T y$$

That is, compute

$$X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix} = \begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix};$$

$$\mathrm{X}^T y = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 9 \\ 57 \end{bmatrix}$$

The normal equations are

$$\begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 9 \\ 57 \end{bmatrix}$$

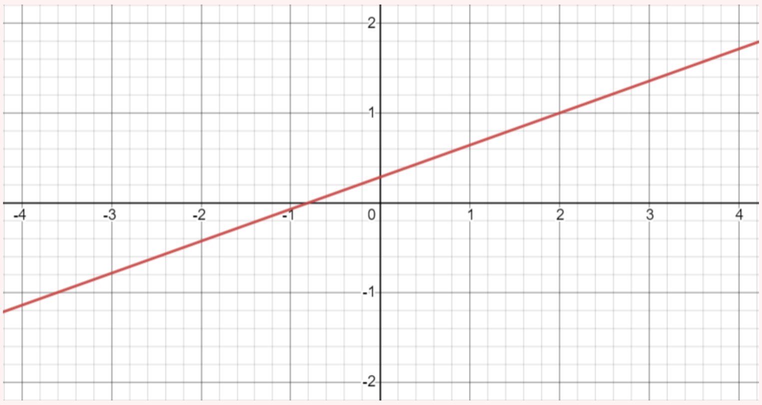
Hence

$$\begin{bmatrix} \beta_0 \\ beta_1 \end{bmatrix} = \begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix}^{-1} \begin{bmatrix} 9 \\ 57 \end{bmatrix} = \frac{1}{84} \begin{bmatrix} 142 & -22 \\ -22 & 4 \end{bmatrix} \begin{bmatrix} 9 \\ 57 \end{bmatrix} = \frac{1}{84} \begin{bmatrix} 24 \\ 30 \end{bmatrix} = \begin{bmatrix} 2/7 \\ 5/14 \end{bmatrix}$$

Thus, the least-squares line has the equation

$$y = \frac{5}{14}x + \frac{2}{7}$$

The graph below depicts the line.



Problem 3 (20 marks)

You are given the following single perceptron classifier which is fed with a binary input.

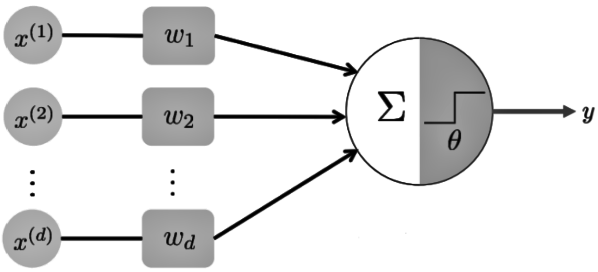


Figure 2: Perceptron Classifier

The output of the classifier is dictated by the following relationship:

$$y = \begin{cases} 1 & \sum_{i=1}^{d} w_i x^{(i)} \geq \theta \\ 0 & \sum_{i=1}^{d} w_i x^{(i)} < \theta \end{cases}$$

(a) **[10 marks]** Suppose that the threshold value $\theta = -3$ and $d = 2$. What would be the values of the weights w_1 and w_2 if the aforementioned perceptron was used to create a NAND function.

The truth table of a NAND function is given below:

A	B	A **NAND** B
0	0	1
0	1	1
1	0	1
1	1	0

(b) **[10 marks]** Suppose that the threshold value $\theta = -3$ and $d = 3$. What would be the values of the weights w_1, w_2, and w_3 if the aforementioned perceptron was used to create a NOR function.

The truth table of a NOR function is given below:

A	B	A **NOR** B
0	0	1
0	1	0
1	0	0
1	1	0

Solution.

(a) One possible solution is that $w_1 = -2$ and $w_2 = -2$.

(b) One possible solution is that $w_1 = -4$ and $w_2 = -4$ and $w_3 = -4$.

Problem 4 (40 marks)

You are tasked with creating an autonomous vehicle (a car that can drive itself without human involvement) using a neural network. The specifications of this network are as follows:

- Input is a set of grayscale images of the steering wheel of size 128×128 pixel.
- Each image comes with two labels; the first denotes the steering wheel's angle and the second denotes the vehicle's speed in km/h.
- The neural network has a hidden layer with 1024 units. The activation function used in this hidden layer is the ReLU function.
- There is no activation function at the output.

(a) [**5 marks**] Draw and label the neural network.
Hint: Each node must connect with every node of the following layer.

(b) [**5 marks**] Calculate the total number of weights in the neural network.
Note: Be mindful of the bias.

The loss function chosen is given by: $J = \frac{1}{2}|\boldsymbol{y} - \hat{\boldsymbol{y}}|$, where $\boldsymbol{y}$ is the true training label vector, and $\hat{\boldsymbol{y}}$ is the predicted output vector.
You are given the following transformations:

- $\boldsymbol{x}$ is the flattened image vector with bias appended.
- $\boldsymbol{z}$ is the vector of values at hidden layer, before activation function.
- $W^{[1]}$ is the weight matrix mapping from the input layer to the hidden layer, i.e., $\boldsymbol{z} = W^{[1]}\boldsymbol{x}$
- $g(.)$ is the ReLU activation function.
- $\boldsymbol{a}$ is the vector of values at hidden layer, after passing through activation function, i.e., $\boldsymbol{a} = g(\boldsymbol{z})$
- $W^{[2]}$ is the weight matrix mapping from hidden layer to the output layer, i.e., $\hat{\boldsymbol{y}} = W^{[2]}\boldsymbol{a}$

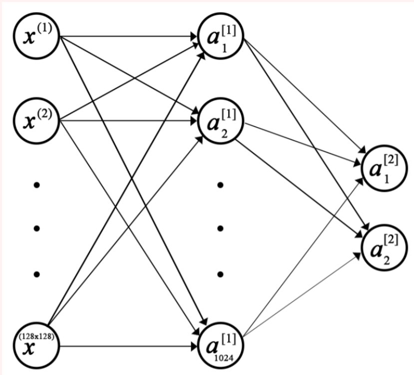
(c) **[10 marks]** Derive $\frac{\partial J}{\partial W_{ij}^{[2]}}$

(d) **[10 marks]** Now write $\frac{\partial J}{\partial W^{[2]}}$, i.e., use vector products.

(e) **[10 marks]** Using your result from previous parts, derive $\frac{\partial J}{\partial W_{ij}^{[1]}}$

Solution.

(a) The neural network is shown below:



(b) $\big(d \times 1024 + 1024\big) + \big((1024 \times 2) + 2\big)$, where $d = 128 \times 128$.

(c)

$$\begin{aligned}\frac{\partial J}{\partial W_{ij}^{[2]}} &= (\hat{\boldsymbol{y}} - \boldsymbol{y})^T \frac{\partial \hat{\boldsymbol{y}}}{\partial W_{ij}^{[2]}} \\ &= (\hat{y}_i - y_i)a_j\end{aligned}$$

(d)

$$\frac{\partial J}{\partial W^{[2]}} = (\hat{\boldsymbol{y}} - \boldsymbol{y})^T \boldsymbol{a}$$

(e)

$$\begin{aligned}\frac{\partial J}{\partial W_{ij}^{[1]}} &= (\hat{\boldsymbol{y}} - \boldsymbol{y})^T \frac{\partial \hat{\boldsymbol{y}}}{\partial W_{ij}^{[1]}} \\ &= (\hat{\boldsymbol{y}} - \boldsymbol{y}) W^{[2]} \frac{\partial \boldsymbol{a}}{\partial W_{ij}^{[1]}} \\ &= (\hat{\boldsymbol{y}} - \boldsymbol{y}) W^{[2]} [0, \ldots, g^{'}(z_i)x_j, \ldots, 0]^T \\ &= ((\hat{\boldsymbol{y}} - \boldsymbol{y})^T W^{[2]})_i g^{'}(z_i) x_j\end{aligned}$$

The solution for part (e) is similar to the solution for part (c) except that it includes the activation function.

— End of Assignment —