SRN | | | | | | | | | | | | |

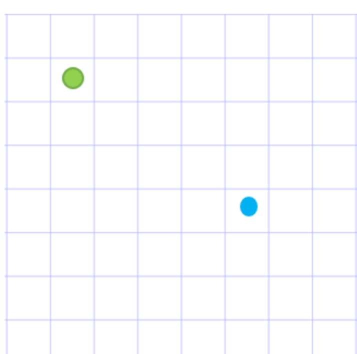
PES UNIVERSITY	**PES University, Bengaluru** (Established under Karnataka Act No. 16 of 2013)	**UE20CS904**

November 2022 : END SEMESTER ASSESSMENT (ESA)

M TECH DATA SCIENCE AND MACHINE LEARNING_ SEMESTER I

UE20CS904 - Mathematical Foundation

Time: 3 Hrs	Answer All Questions	Max Marks: 80

		Section A (20 marks)	
1	a)	Find X if $\det\begin{pmatrix}2 & 4\\5 & 1\end{pmatrix} = det\begin{pmatrix}2x & 4\\6 & x\end{pmatrix}$	2
	sol	Given that, $\begin{vmatrix}2 & 4\\5 & 1\end{vmatrix} = \begin{vmatrix}2x & 4\\6 & x\end{vmatrix}$ $2 - 20 = 2x^2 - 24$ Now, bring the x term on the L.H.S $2x^2 = 6$ $x^2 = 6/2$ To remove the square root on the L.H.S, $x = \pm\sqrt{3}$ Thus, the value of x is $\pm\sqrt{3}$	
	b)	Which distance metric is suitable for calculating the least number of squares moved between the starting position (Green Point) and ending position (Blue point) on the chessboard (each square of unit length) for the Queen (Queen can move either diagonally or vertically or horizontally)? Give formula for the same. 	2
	sol	Solution: Manhattan distance or Euclidean distance and Chebyshev distance can be used to calculate the distance between two points.	

SRN

green = (1,6) blue = (5,3)

The Manhattan distance or L1 norm is,

L1 Norm : $||u|| = |x_1| + |y_1| = 4 + 3 = 7$

The Euclidean distance or L2 norm is,

L2 Norm : $||u|| = \sqrt{x^2 + y^2} = \sqrt{4^2 + 3^2} = \sqrt{9 + 16} = \sqrt{25} = 5$

Chebyshev Distance

L_∞ **Norm:** $||u|| = Max\{|X|, |y|\} = Max\{3,4\} = 4$

c) Find out whether the function is concave or convex , $f(x) = -8x^2 + 15$ — 2

sol

To find the concavity, look at the second derivative. If the function is positive at our given point, it is concave. If the function is negative, it is convex.

To find the second derivative we repeat the process, but using $-16x$ as our expression.

$$\frac{d}{dx}(-16x) = 1 * -16x^{1-1}$$

$$\frac{d}{dx}(-16x) = -16x^0$$

$$\frac{d}{dx}(-16x) = -16$$

As you can see, our second derivative is a constant. It doesn't matter what point we plug in for x; our output will always be negative. Therefore our graph will always be convex.

Combine our two pieces of information to see that at the given point, the graph is decreasing and convex.

d) Find the vector projection of the vector a = {5; 5} on b = {8; 2} — 2

Vector projection of a on b =((a . b) / (b . b)) b

= (50/68) (8,2)

= (5.9, 1.5)

e) What is the effect of higher learning rate in Gradient descent algorithm? — 2

When the learning rate is a higher value, it can cause the model to converge too quickly to a suboptimal solution.

2 a) Calculate the angle between two given vectors. The two vectors are, — 2

a = $\vec{i} + 2\vec{j}$ and

b = $9\vec{i} + 3\vec{j}$

sol Solution :

$$\cos\Theta = \frac{\vec{a} . * \vec{b}}{||\vec{a}||\ ||\vec{b}||} = \frac{(1*9)+(2*3)}{\sqrt{1+4}\sqrt{81+9}}$$

SRN | | | | | | | | | | | | |

		$= \frac{15}{\sqrt{5}\,\sqrt{90}} \quad = \frac{15}{5\sqrt{18}}$ $= \frac{3}{3\sqrt{2}} \quad = \frac{1}{\sqrt{2}}$ Hence $\Theta = 45°$	
	b)	**What will happen when eigenvalues are roughly equal?** A. PCA will perform outstandingly B. PCA will perform badly C. Can't Say D. None of above Sol: B. PCA will perform badly Since All Pcs are same When eigenvalues are roughly equal, this indicates that each principal component explains a similar amount of variance in the data. In such cases, dropping any principal components would result in a significant loss of information. This makes PCA less effective for dimensionality reduction because it implies that the original data is uniformly spread across all dimensions and there are no significantly dominant directions of variance to prioritize.	2
	c)	Calculate the Jacobian matrix for the following function $f_1(x,y) = x^3y$ $f_2(x,y) = \frac{x^2}{y} + y^2$ Solution :	2

SRN

$$J(x,y) = \begin{pmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} \end{pmatrix} = \begin{pmatrix} 3x^2 y & x^3 \\ \frac{2x}{y} & -\frac{x^2}{y^2} + 2y \end{pmatrix}$$

d) In simple linear regression for a single data point (x_1, y_1) we define loss as

$L(w_0, w_1) = \left(\widehat{y_1} - (w_0 + w_1 x_1)\right)^2$ where $\widehat{y_1}$ is predicted value for y_1 find $\frac{\partial L}{\partial w_0}$ and $\frac{\partial L}{\partial w_1}$

d) 2

$L(w_0, w_1) = \left(\widehat{y_1} - (w_0 + w_1 x_1)\right)^2$

$\frac{\partial L}{\partial w_0} = 2\left(\widehat{y_1} - (w_0 + w_1 x_1)\right)(-1)$

$\frac{\partial L}{\partial w_0} = -2\left(\widehat{y_1} - (w_0 + w_1 x_1)\right)$

$\frac{\partial L}{\partial w_1} = 2\left(\widehat{y_1} - (w_0 + w_1 x_1)\right)(-x_1)$

$\frac{\partial L}{\partial w_1} = -2\left(\widehat{y_1} - (w_0 + w_1 x_1)\right)x_1$

e) We have an rgb image saved as img. An RGB image has length and width 63.We are creating a new image by concatenating img[:,:63,1] , img[:,63:126,:2] & img[:,126:,0]. Wha kind of changes can we observe in the new image as compared to the original image (img). 2

Solution :

img[row_start:row_end, column_start:column_end , channels]

red_channel = img[:,:,0]

img[:,:63,1]:

SRN

This slices the original image to take all rows (:), the first 63 columns (:63), and only the green channel (1) of the RGB image. This effectively extracts a green-channel-only view of the left third of the original image.

img[:,63:126,:2]:

There seems to be a small confusion in the slicing. Given that the original image width is 63 pixels, img[:,63:126,:2] suggests an attempt to slice beyond the image's width.

img[:,126:,0]:

Similar to the second operation, this suggests taking a slice beyond the described dimensions of the image.

Section B (30 marks)

3 a Consider a firm operating two plants in two different locations. They both produce the same output (say, **10 units**) using the same type of inputs. Although the amounts of inputs vary between the plants the output level is the same. 5

1. The firm management suspects that the production cost in Plant 2 is higher than in Plant 1. Verify?

The following information was collected from the managers of these plants.

PLANT 1

Input	Price	Amount used
Input 1	3	9
Input 2	5	10
Input 3	7	8

PLANT 2

Input	Price	Amount used
Input 1	4	8
Input 2	7	12
Input 3	3	9

SRN | | | | | | | | | | | |

- Question 1. Does this information confirm the suspicion of the firm management'
- Plant 1: Price Vector $p_1 = \begin{bmatrix} 3 \\ 5 \\ 7 \end{bmatrix}$ and Quantity Vector $q_1 = \begin{bmatrix} 9 \\ 10 \\ 8 \end{bmatrix}$
- Plant 2 : Price Vector $p_2 = \begin{bmatrix} 4 \\ 7 \\ 3 \end{bmatrix}$ and Quantity Vector $q_1 = \begin{bmatrix} 8 \\ 12 \\ 9 \end{bmatrix}$
- In Plant 1 the total cost is $\mathbf{c_1} = \mathbf{p_1}.\mathbf{q_1} = 133$, which implies that $\boldsymbol{unit\ cost\ is}\ \mathbf{13.3}$.
- In Plant 2, cost of production is $\boldsymbol{c_2} = \boldsymbol{p_2}.\boldsymbol{q_2} = 143$, which gives $\boldsymbol{unit\ cost\ as}\ \mathbf{14.3}$ which is higher than the first plant. That is, the suspicion is reasonable.

b

(ii) A stone is dropped into a **quiet lake** and waves move in circles at a speed of 5cm per second. At the instant, when the radius of the circular wave is 8 cm, how fast is the enclosed area increasing? Solution :

5

we'll use the relationship between the area of a circle and its radius to find out how fast the enclosed area is increasing when the radius of the circular wave is 8 cm. The area A of a circle is given by the formula A=πr2, where r is the radius of the circle. To find how fast the enclosed area is increasing, we need to find the rate of change of the area with respect to time, denoted as dA/dt.

1. **Identify what we know**:
   - The speed of the radius increasing ($\frac{dr}{dt}$) is 5 cm per second.
   - The radius (r) at the instant we are interested in is 8 cm.
2. **Identify what we want to find**:
   - We want to find $\frac{dA}{dt}$ when $r = 8$ cm.
3. **Use the area formula of a circle to relate A and r**:
   - $A = \pi r^2$.
4. **Differentiate both sides of the area formula with respect to time (t) to find $\frac{dA}{dt}$**:
   - $\frac{dA}{dt} = \frac{d}{dt}(\pi r^2)$.
   - By applying the chain rule, we get $\frac{dA}{dt} = 2\pi r\frac{dr}{dt}$.
5. **Substitute the known values into the differentiated equation to find $\frac{dA}{dt}$**:
   - Given $\frac{dr}{dt} = 5$ cm/s and $r = 8$ cm, substitute these values into $\frac{dA}{dt} = 2\pi r\frac{dr}{dt}$.

Let r be the radius of the circular wave and A the enclosed area at time t.

$\therefore \quad A = \pi r^2 \quad \Rightarrow \quad \frac{dA}{dt} = \pi . 2r\frac{dr}{dt} = 2\pi r\frac{dr}{dt}$

From given condition, $\frac{dr}{dt} = 5$ cm / sec

$\therefore \quad$ when r = 8 cm, $\frac{dA}{dt} = 2\pi \times 8 \times 5\ cm^2/s = 80\ cm^2/s$

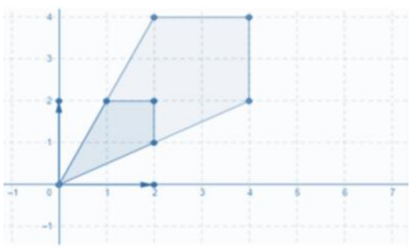
$\therefore$ the enclosed area is increasing at the rate of 80 cm^2/s, when r = 8 cm.

SRN

c) Compute the following convolution for coloured cell, what kind of output the following convolution will have on an image? **5**

$$\begin{bmatrix}138 & 134 & 101\\ 119 & 99 & 83\\ 84 & 80 & 79\end{bmatrix} * \begin{bmatrix}0 & -1 & 0\\ -1 & 5 & -1\\ 0 & -1 & 0\end{bmatrix} = \begin{bmatrix} & & \\ & 79 & \\ & & \end{bmatrix}$$

d) In the plot shown below the dark shaded portion represents the original coordinates of an object and the same after transformation is represented by the lightly shaded object. Write the coordinates, the transformation matrix and the coordinates after transformation. **5**



The original coordinates (A, B, C, D) are

$$\begin{pmatrix}0 & 1 & 2 & 2\\ 0 & 2 & 2 & 1\end{pmatrix}$$

The transformation matrix is

$$\begin{pmatrix}2 & 0\\ 0 & 2\end{pmatrix}$$

The transformed coordinates (A', B', C', D') are

$$\begin{pmatrix}0 & 2 & 4 & 4\\ 0 & 4 & 4 & 2\end{pmatrix}$$

e) The Following table lists the weight and heights of 5 boys Find the covariance matrix for the data. **5**

Boy	1	2	3	4	5
Weight(lb)	120	125	125	135	145
Height(in.)	61	60	64	68	72

$M = \begin{pmatrix}130\\ 65\end{pmatrix}$. Subtract M from the observation vectors and obtain

$$B = \begin{pmatrix}-10 & -5 & -5 & 5 & 15\\ -4 & -5 & -1 & 3 & 7\end{pmatrix}$$

Then the sample covariance matrix is

$$S = \frac{1}{5-1}\begin{pmatrix}-10 & -5 & -5 & 5 & 15\\ -4 & -5 & -1 & 3 & 7\end{pmatrix}\begin{pmatrix}-10 & -4\\ -5 & -5\\ -5 & -1\\ 5 & 3\\ 15 & 7\end{pmatrix}$$

SRN | | | | | | | | | | | |

		$= \frac{1}{4}\begin{pmatrix} 400 & 190 \\ 190 & 100 \end{pmatrix} = \begin{pmatrix} 100 & 47.5 \\ 47.5 & 25 \end{pmatrix}$	
	f)	Find the Eigen values of A : $A = \begin{bmatrix} 4 & 2 \\ 1 & 3 \end{bmatrix}$ First, we solve the characteristic equation $\lvert A - \lambda I \rvert = 0$, where I is the identity matrix and λ represents the eigenvalues. $\begin{vmatrix} 4-\lambda & 2 \\ 1 & 3-\lambda \end{vmatrix} = 0$ $(4-\lambda)(3-\lambda) - (2)(1) = 0$ $\lambda^2 - 7\lambda + 10 = 0$ Solving this quadratic equation, we find the eigenvalues λ: $\lambda = \frac{7 \pm \sqrt{(-7)^2 - 4(1)(10)}}{2(1)}$ $\lambda = \frac{7 \pm \sqrt{49-40}}{2}$ $\lambda = \frac{7 \pm \sqrt{9}}{2}$ $\lambda_1 = 5, \quad \lambda_2 = 2$	5

SRN

For each eigenvalue, we solve $(A - \lambda I)v = 0$ to find the eigenvector v.

For $\lambda_1 = 5$:

$$(A - 5I)v = \begin{bmatrix} -1 & 2 \\ 1 & -2 \end{bmatrix} v = 0$$

Let's solve the system:

$-v_1 + 2v_2 = 0$

$v_1 = 2v_2$

Choosing $v_2 = 1$ (for simplicity), we get $v_1 = 2$. So, one eigenvector is $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$.

For $\lambda_2 = 2$:

$$(A - 2I)v = \begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix} v = 0$$

Let's solve the system:

$2v_1 + 2v_2 = 0$

$v_1 + v_2 = 0$

Choosing $v_2 = -1$ (for simplicity), we get $v_1 = 1$. So, another eigenvector is $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$.

SRN | | | | | | | | | | | | |

Section C (30 marks)			
4	a)	A headphone manufacturer determines that in order to sell x units of a new headphone, the price per unit, in dollars, must be p(x) = 1000 - x. The manufacturer also determines that the total cost of producing x units is given by C(x)= 3000 + 20x. i) Find the total revenue R(x) ii) Find the total profit P(x). iii) How many units must the company produce and sell in order to maximize profit? iv) What is the maximum profit? v) What price per unit must be charged in order to make this maximum profit? Sol:- **a)** $R(x) = \text{Total revenue}$ $= (\text{Number of units}) \cdot (\text{Price per unit})$ $= x \cdot p$ $= x(1000 - x) = 1000x - x^2$ **b)** $P(x) = \text{Total revenue} - \text{Total cost}$ $= R(x) - C(x)$ $= (1000x - x^2) - (3000 + 20x)$ $= -x^2 + 980x - 3000$ **c)** To find the maximum value of $P(x)$, we first find $P'(x)$: $P'(x) = -2x + 980.$ This is defined for all real numbers, so the only critical values will come from solving $P'(x) = 0$: $P'(x) = -2x + 980 = 0$ $-2x = -980$ $x = 490.$ There is only one critical value. We can therefore try to use the second derivative to determine whether we have an absolute maximum. Note that $P''(x) = -2$, a constant. Thus, $P''(490)$ is negative, and so profit is maximized when 490 units are produced and sold. **d)** The maximum profit is given by $P(490) = -(490)^2 + 980 \cdot 490 - 3000$ $= \$237,100.$ Thus, the stereo manufacturer makes a maximum profit of \$237,100 by producing and selling 490 stereos. **e)** The price per unit needed to make the maximum profit is $p = 1000 - 490 = \$510.$	10
	b)	Find singular Value decomposition of $A = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}$	10

SRN | | | | | | | | | | | |

SVD decomposes a matrix A into three matrices U, Σ, and V^T such that $A = U\Sigma V^T$:

- U is an orthogonal matrix whose columns are the eigenvectors of AA^T.
- Σ is a diagonal matrix containing the square roots of the eigenvalues from both AA^T and A^TA, known as the singular values of A.
- V^T is the transpose of an orthogonal matrix V whose columns are the eigenvectors of A^TA.

Find the SVD of $\begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}$.

$$\begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}^T = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}$$

$AA^T =$

$$\begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix} \cdot \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix} = \begin{bmatrix} 13 & 12 \\ 12 & 13 \end{bmatrix}$$

$$\begin{vmatrix} 13-\lambda & 12 \\ 12 & 13-\lambda \end{vmatrix} = 0$$

$$(13-\lambda)^2 - 12 \cdot 12 = 0$$

$$\lambda^2 - 26\lambda + (13^2 - 12^2) = 0$$

$$\lambda^2 - 26\lambda + 25 = 0$$

Solving for λ, we get:

$$\lambda = \frac{26 \pm \sqrt{26^2 - 4 \cdot 25}}{2}$$

$$\lambda = \frac{26 \pm \sqrt{676 - 100}}{2}$$

$$\lambda = \frac{26 \pm \sqrt{576}}{2}$$

$$\lambda = \frac{26 \pm 24}{2}$$

So the eigenvalues are $\lambda_1 = 25$ and $\lambda_2 = 1$.

SRN | | | | | | | | | | | |

Eigenvalue: 25, **eigenvector:** $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

Eigenvalue: 1, **eigenvector:** $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$.

Find the square roots of the nonzero eigenvalues (σ_i):

$\sigma_1 = 5$

$\sigma_2 = 1$

The Σ matrix is a zero matrix with σ_i on its diagonal: $\Sigma = \begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix}$.

The columns of the matrix U are the normalized (unit) vectors: $U = \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}$

, $v_i = \frac{1}{\sigma_i} \cdot \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}^T \cdot u_i$:

$$v_1 = \frac{1}{\sigma_1} \cdot \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}^T \cdot u_1 = \frac{1}{5} \cdot \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix} \cdot \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix}$$

$$v_2 = \frac{1}{\sigma_2} \cdot \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}^T \cdot u_2 = \frac{1}{1} \cdot \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix} \cdot \begin{bmatrix} -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix} = \begin{bmatrix} -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix}$$

e, $V = \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}$.

	c)		10

SRN

Consider the data given below and fit a linear regression line y = ax+b using gradient descent.

X	0	0.4	0.6	1
Y	0	1	0.48	0.95

Initialize the weights a and b to 0.8, 0.2 respectively. Update the weights such that the error is minimum using gradient descent. Use the function sum of squared errors $\sum (y-\hat{y})^2$ where $\hat{y}$ is the y-predicted value and y is the actual given y. Plot the linear regression line after updating

sol

See the plot of the linear regression line which is close the the new a and b values

Solution:
Let us assume the learning rate as 0.1

X	Y	YP	Y-YP	(Y-YP) squared	SSE	Der. of SSE w.r.t.a	Der. of SSE w.r.t.b
0	0	0.2	-0.2	0.04	0.02	0	0.2
0.4	1	0.52	0.48	0.2304	0.1152	-0.192	-0.48
0.6	0.48	0.68	-0.19613	0.038466597	0.019233299	0.117677419	0.196129032
1	0.95	1	-0.04839	0.002341311	0.001170656	0.048387097	0.048387097
						-0.025935484	-0.035483871

Learning rate = 0.1
New a = 0.8 - 0.1(-0.025) 0.8025
New b = 0.2 -0.1(-0.035) 0.2035

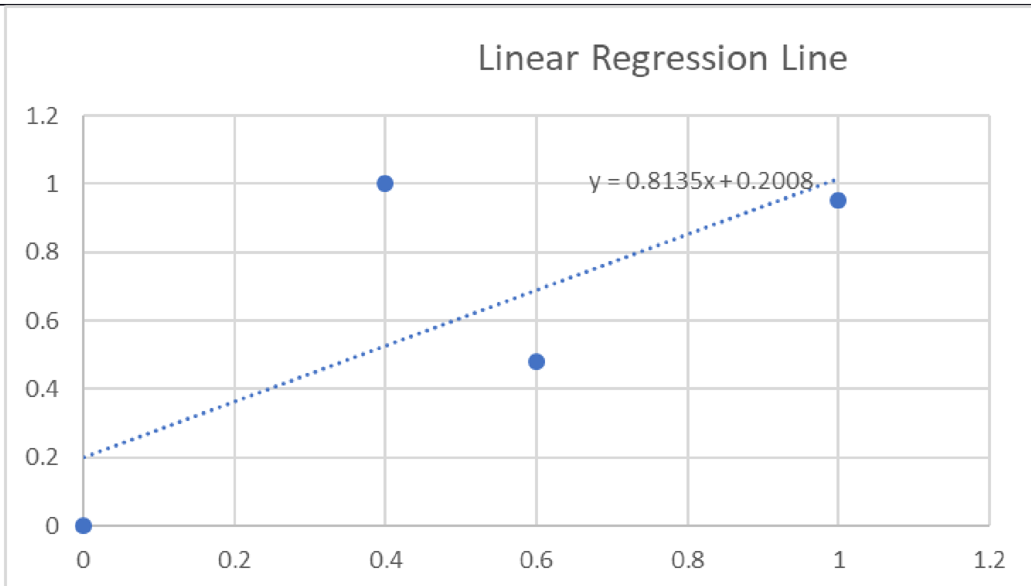
Replace a with 0.8025 and b with 0.2035 and repeat the same.

X	Y	YP	Y-YP	(Y-YP) squared	SSE	Der. of SSE w.r.t.a	Der. of SSE w.r.t.b
0	0	0.2035	-0.2035	0.04141225	0.020706125	0	0.2035
0.4	1	0.5245	0.4755	0.22610025	0.113050125	-0.1902	-0.4755
0.6	0.48	0.685	-0.20113	0.040452888	0.020226444	0.120677419	0.201129032
1	0.95	1.006	-0.05439	0.002957956	0.001478978	0.054387097	0.054387097
						-0.015135484	-0.016483871

Learning rate = 0.1
New a = 0.8 - 0.1(-0.015) 0.804
New b = 0.2 -0.1(-0.016) 0.2051

See the plot of the linear regression line which is close the the new a and b values

SRN | | | | | | | | | | | |



- $$\frac{\partial SSE}{\partial a} = \frac{1}{m}\sum(y_p - y)X = grad\_a$$
- $$\frac{\partial SSE}{\partial b} = \frac{1}{m}\sum(y_p - y) = grad\_b$$

$$a_{new} = a_{old} - r\frac{1}{m}\sum(y_p - y)X$$

$$b_{new} = b_{old} - r\frac{1}{m}\sum(y_p - y)$$

$H = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix}$ $f = x^2 + y^2$ $|H| > 0$ then we minimum