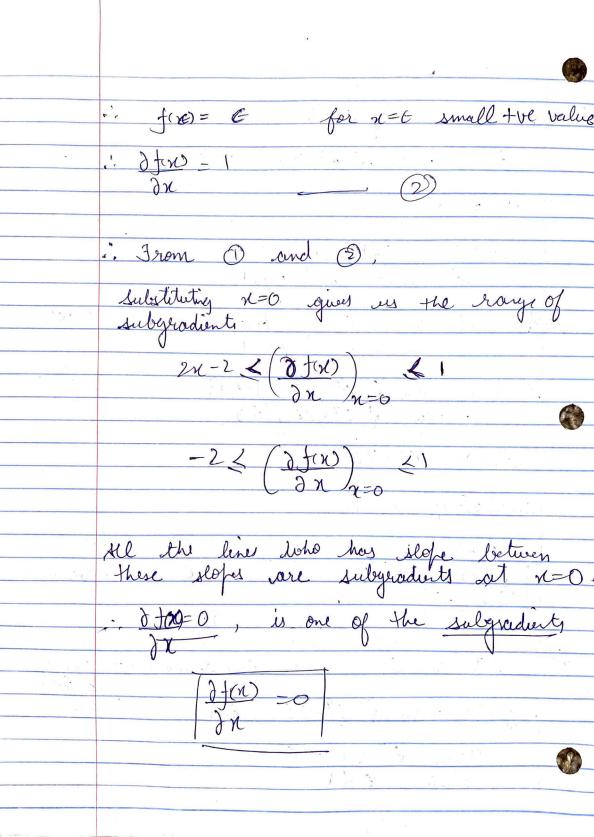
Warm- up 02. compute subgradients (a) $f(x) = man \left\{ x^2 - 2x, |x| \right\}$ (1) x=0 egradient does not exist souse let x= - E where E is a small tol $f(x) = \max_{x \in \mathbb{Z}} \{ \frac{e^2 + 2e}{2t}, \frac{1-e}{2t} \}$ $f(c) = e^2 + 2e$ substituting or, $f(n) = n^2 - 2n$ $\int f(x) = 2n - 2 \qquad \text{for } x = -\epsilon \qquad \text{small ve value}$ let n= E where E is a small tol value. & EXI :- f(x) = man & E2-2t, Eg Hou, E²-ZE X E because E 21



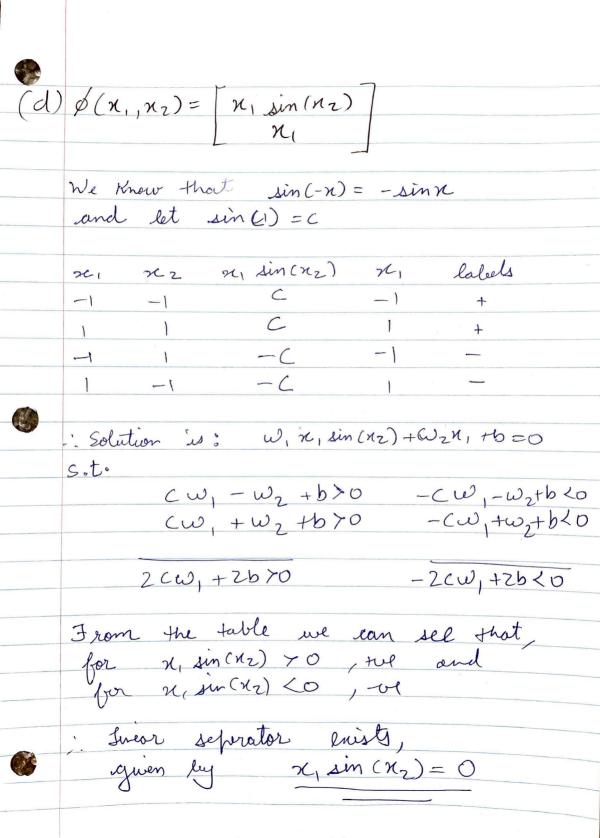
(11) -(11)= man & x2-2xx, 1x12 $\kappa^2 - 2\kappa = +4 + 4 = 8$ (n2-227/21/ at n=2 1° +(x) = n2-2x 2+(n) = 2n-2 for $\kappa = -2$ 0f(x) = 2(-2)-2 = -6(b) $q(x) = man (x-1)^2, (x-2)^2 \frac{1}{2}$ At x=105, $(x-1)^2=(n-2)^2=(0.5)^2=0.25$ The gradient does not exist, Jake E = some small the value.

At n=1.5, substituting x=1.5 in O and O; $2(n-2) < \left(\frac{\partial G(x)}{\partial n}\right) < 2(x-1)$ -1 < 34(n) < 1 All the lines with this slope at x=105 are subgradients 29(n) = 0 is a subgradust at n=1.5 (11) fox g(x)= man f(x-v)2 g $\left(: \left(\chi - 2 \right)^2 \right) \left(\chi - 1 \right)^2$ dy(n) = 2(n-2) $\frac{1}{2} \frac{\partial q(x)}{\partial x} = -4$ is subgradut at x=0

Problem 2 Is the data linearly separable $\phi(x_1, x_2) = \begin{bmatrix} x_1 + x_2 \\ x_1 - 2x_2 \end{bmatrix}$ (a) x_1 x_2 $x_1 + x_2$ $x_1 - \lambda x_2$ label 0 ". We need to find a line $\omega_1 n'_1 + \omega_2 n'_2 + b = 0$ (+) 2b<0 —© From (1) and (2); now, $-2\omega_1 + \omega_2 \neq 0 \Rightarrow \omega_2 \neq 2\omega_1$ 2w, -w, 70 => 2w, > wz

 $\beta W_1 > 2W_2$ W2>2W, connot be satisfied with any value of w, wz $(b) \qquad \phi(x_1, x_2) = \begin{bmatrix} x_1^2 \\ x_2^2 \\ x_1 x_2 \end{bmatrix}$ χ_1 χ_2 χ_1^2 χ_2^2 χ_1^2 -1 | | -1 Jo find a solution, w, x,2 + w2 x22 + w3 x, x2 +b=0 $\omega_{1} + \omega_{2} + \omega_{3} + b \neq 0$ _ (2) The solution to this is this is our linear separator

 $(() \phi(n_1, n_2) = [exp(n_1)]$ $enp(n_2)$: We want wiex + wiex + b = 0 $w,e^{-1}+w_2e^{-1}+b > 0$ _ w, $e^{-1}+w_2e^{-1}+b > 0$ _ w $w_1e^{-1} + w_2e + b < 0 = 0$ $w_1e + w_2e^{-1} + b < 0 = 0$ Add D and D; $w_1(e^{-1}+e) + w_2(e+e^{-1}) + 2b > 0$ Add 3 cend (9), $w_{1}(e^{-1}+e)+w_{2}(e^{-1}+e)+2b (0)$ From 3 and 6, We conclude that a linear separator is not possible.



Problem 2: 02 Polynomial regression for 2-D data points. The data in 2-cl box is of the form (n, y). Thus, for polynomial regression, we med a K-degree polynomal of n for which the output isy. ... Our hypothesis function: WKNK + WK-1 x + + bx = y In vector notation; J(n) = 60 X + b + where $X = \begin{bmatrix} n \\ n^{K-1} \end{bmatrix}$ Flature Vectors

Teature vector X is generated from our input x. The loss function: mean square error $2(f) = \frac{1}{M} \left(\frac{f(n^m - y^m)^2}{m} \right)$ $L(f) = \frac{1}{M} \left((\omega^T \kappa^m + b) - y^m \right)^2$ And the objective function is min $L(f) = \min_{\omega,b} \frac{1}{M} \underbrace{\sum_{m}^{\infty} (\omega_{n}^{T} + b) - y_{m}^{m}^{2}}$ gradient descent to compute the w,b for.

". Gradient descent $\frac{\partial \mathcal{M}}{\partial \omega} = \frac{2}{M} \neq ((\omega^T n^M b) - y^M), n^M$ $\frac{\partial L}{\partial b} = \frac{2}{M} \left\{ \left((\omega^T n^m + b) - y^m \right) \right\}$ Iteratively compute; Wen: = Wt - X DL bt1 := bt - 2)b untill, 2L 40 and 2L 40 Time complenity (K degree polynoid, m doite points) For a R degree polynomical, the Predictions will take O(Km) time. And the gradient computing using the predictions olso takes O(Km) time:

Time Complainty = O(Km) + O(Km) = O(Km)

Problem 3 Enponential Regression

Input (n,y) & R²

M input adata points fin) = emplantb) $L(f) = 1 \leq \left(\exp(ax^m + b) - y^m \right)^2$ objective min $L(f) = \min_{a,b} L \left(exp(ax^m + b) - y^m \right)^2$ nle minimize this function using spradient obescent.

(2) Gradut Descent Main L(f)= min I & (exp(axm+b)-ym)²
a,b M m. We have to compute a, b to mining loss- $\frac{1}{3a} = \frac{2}{M} \leq \frac{(enp(ax^m+b)-y^m)enp(ax^m+b)}{2} \cdot x$ The momentum of the second of the modern of the momentum of the momentum of the modern Iteratively: (similtaneously) an at = at - x dl $b_{tH} = b_t - \propto \frac{\partial L}{\partial b}$ untill, de so and de so

(3) Is the optimization froblem conven?. Given the loss fur, if we can get the convenity for one input, we can say it remains conven by sum of conven fur theorem i. f(x) = (enp(an +b) -y2)2 fin) = (enp Can (+b)) -1 enplan(+b) y +yc2 For the term orporat: $(enp(an'+b))^{2} = enp(z(an'+b))$ $= enp^{2an'} \cdot enp^{2b}$ (i) $e^{2ax^{\ell}} = 2x^{\ell} \left(1 + a + a^{2} + \frac{a^{3}}{3!} + \frac{a^{3$ ab, ab², a²b²,... Thus, the eyn is not comen. Imus, our loss for enformential regression is not conven.