Submission for Machine Learning Assignment 4

Team: 47
Students: Juan Alvarez, Prabhav Gupta, Praffful Ravuri

November 12, 2025

1 Support Vector Machines

1. The kernel trick provides a solution to the problem of high computation costs when trying to **make data linearly separable in a higher dimensional space** when applying a **nonlinear mapping** to the data by using the so called kernel functions.

   Kernel functions are functions that compute the dot product in a higher dimensional space, without explicitly computing the mapping.

2. Attached below

ML Assignment 04
group number 47

1. Support Vector Machines.

2)

$$\frac{\partial}{\partial w}\left(\frac{1}{n}\left(\sum_{i=1}^{n} \max(0, 1-y_i \cdot (w^T \cdot x_i))\right) + \frac{\lambda}{2}\left(\sum_{j=1}^{m} w_j^2\right)\right)$$

$$= \underbrace{\frac{\partial}{\partial w}\left(\frac{1}{n}\left(\sum_{i=1}^{n} \max(0, 1-y_i (w^T \cdot x_i))\right)\right)}_{1} + \underbrace{\frac{\partial}{\partial w}\left(\frac{\lambda}{2}\left(\sum_{j=1}^{m} w_j^2\right)\right)}_{2} \quad *$$

1 $\frac{\partial}{\partial w}\left(\frac{1}{n}\left(\sum_{i=1}^{n} \max(0, 1-y_i (w^T \cdot x_i))\right)\right)$

$$= \frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial w} \max(0, 1-y_i (w^T \cdot x_i))$$

$$= \frac{1}{n} \sum_{i=1}^{n} \max\left(0, \frac{\partial}{\partial w}(1-y_i (w^T \cdot x_i))\right)$$

$$\boxed{= \frac{1}{n} \sum_{i=1}^{n} \max(0, -y_i x_i)}$$

2 $\frac{\partial}{\partial w}\left(\frac{\lambda}{2}\left(\sum_{j=1}^{m} w_j^2\right)\right) = \frac{\lambda}{2} \sum_{j=1}^{m} \frac{\partial}{\partial w} w_j^2 = \boxed{\frac{\lambda}{2} \cdot \sum_{j=1}^{m} 2 \cdot w_j}$

*

$$\boxed{\frac{1}{n} \sum_{i=1}^{n} \max(0, -y_i x_i) + \frac{\lambda}{\cancel{2}} \cdot \sum_{j=1}^{m} \cancel{2} w_j}$$

ML Assignment 04

group number 47

2)

a) Pseudocode for Gradient descent.

for 1...i do:

$$w \leftarrow w - \eta \left(\frac{1}{n} \cdot \sum_{i=1}^{n} \max(0, -y_i x_i) + \frac{\lambda}{2} \sum_{j=1}^{m} \cdot w_j \right)$$

return w

b)

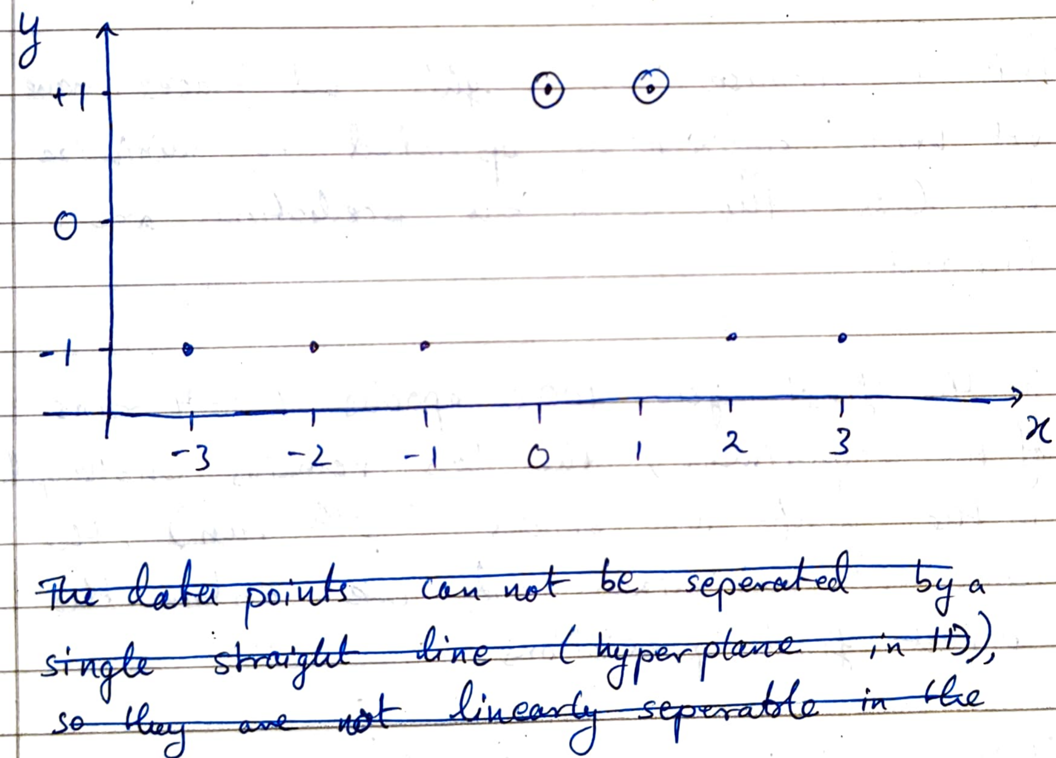
$w = \begin{bmatrix} 1.0 \\ 1.0 \end{bmatrix}$ $\quad x = \begin{bmatrix} 1.0 \\ 0.0 \end{bmatrix}$ $\quad \lambda = 0{,}5$ $\quad \eta = 0.01$ $\quad y = 1.$

$$w' = \begin{bmatrix} 1.0 \\ 1.0 \end{bmatrix} - 0{,}01 \cdot \left(0 + \frac{0{,}5}{2} \cdot \begin{bmatrix} 1.0 \\ 1.0 \end{bmatrix} \right) = \begin{bmatrix} 1.0 \\ 1.0 \end{bmatrix} - \begin{bmatrix} 0.005 \\ 0.005 \end{bmatrix} = \boxed{\begin{bmatrix} 0{,}995 \\ 0{,}995 \end{bmatrix}}$$

Q2 Given Dataset:

x	y
-3	-1
-2	-1
-1	-1
0	1
1	1
2	-1
3	-1

1)



~~The data points can not be seperated by a single straight line (hyperplane in 1D), so they are not linearly seperable in the original space.~~

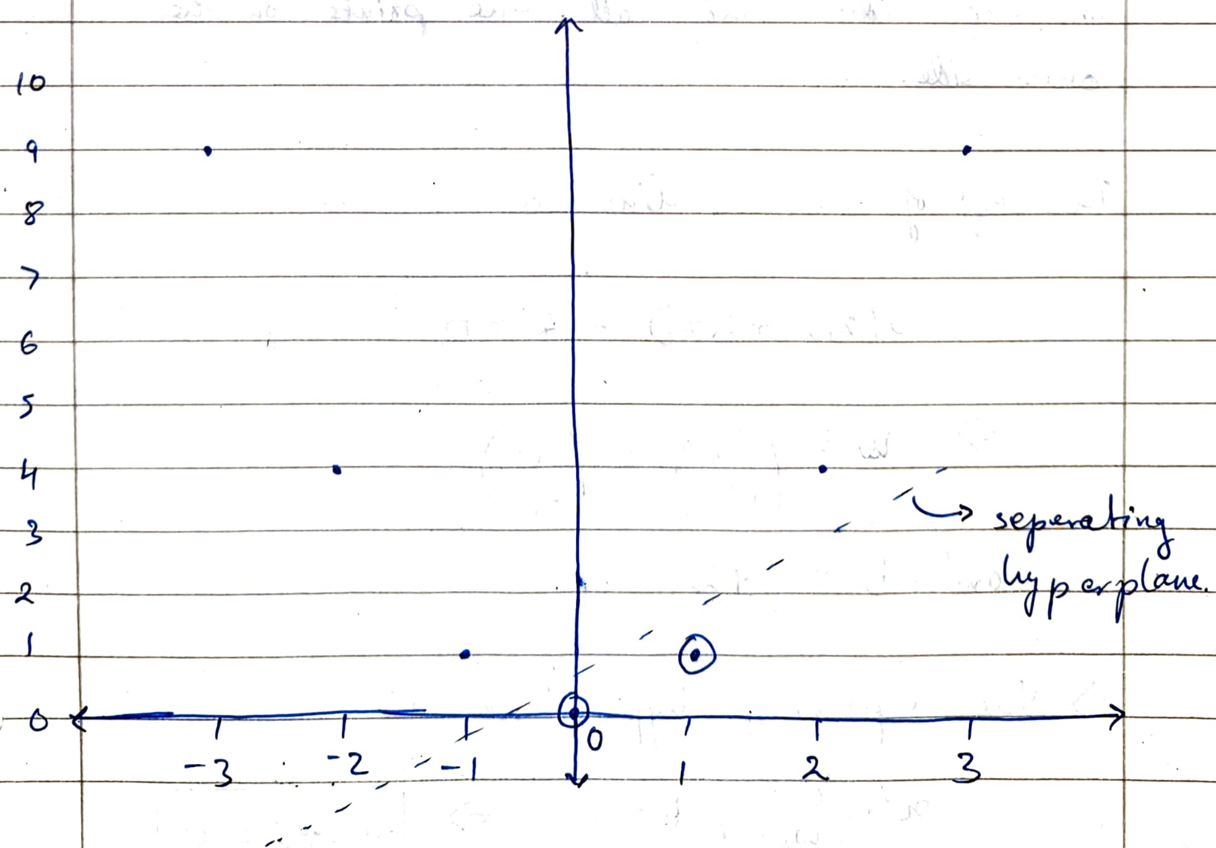
A seperating hyperplane in R is a single point, $x=c$, which must seperate all points of the positive class from all points of the negetive class.

No single point $x=c$ can satisfy this seperation condition. hence it is not linearly seperable.

2) Applying the mapping $g: R \to R^2$ defined by

$$g(x) = \begin{pmatrix} x \\ x^2 \end{pmatrix}$$, we get :

x	x^2	y
-3	9	-1
-2	4	-1
-1	1	-1
0	0	1
1	1	1
2	4	-1
3	9	-1



The transformed data is linearly seperable in R^2. The positive points lie below or on the curve $x_2 = |x_1|$ (since $x_2 = x_1^2 \geq 0$), and generally near the origin, while the -ve points are further from the origin (larger x_2 values).

In R^2, a seperating hyperplane is a line defined by the eq^n :

$$\langle w, x' \rangle + b = 0$$

$$\Rightarrow \quad w_1 x_1 + w_2 x_2 + b = 0$$

We need a line that places all +ve points on one side and all -ve points on the other side.

The eq^n of such a line is

$$0(x_1) + 1(x_2) - 2 = 0$$

$$\Rightarrow \quad w = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$$

and $b = +2$

$\Rightarrow$ the seperating hyperplane is

$$x' \cdot \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} + b = 0 \quad \Rightarrow \quad \boxed{-x_2 + 2 = 0}$$

$\therefore$ seperating hyperplane $\Rightarrow$ $\boxed{x_2 = 2}$

and parapmeters : $w = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$, $b = 2$

where $y = 1$ if $f(x') > 0$
and $y = -1$ if $f(x') < 0$

$$f(x') = -x_2 + 2$$

3) for $x = \dfrac{1+\sqrt{5}}{2}$

$$x_2 = \left(\frac{1+\sqrt{5}}{2}\right)^2 = \frac{1+2\sqrt{5}+5}{4} = \frac{6+2\sqrt{5}}{4} = \frac{3+\sqrt{5}}{2}$$

applying the ~~fn~~ hyperplane fn :

$$f(x_2') = -x_2 + 2 = -\left(\frac{3+\sqrt{5}}{2}\right) + 2$$

$$= \frac{1-\sqrt{5}}{2}$$

$$\approx \underline{\underline{-0.618}}$$

Thus x ~~below~~ < 0
$\Rightarrow$ x belongs to -ve class $(y = -1)$