Supervise Learning →
classification Technique.

Consider data points in $\vec{p}$ dimensional space. we want to separate them by a (p-1) dimensional hyperplane.
Choose the hyperplane such that the distance from it to the nearest data point on each side is maximized.
If such a hyperplane exists → maximum margin hyperplane and linear classifier → max. margin classifier.

Linear SVM

Consider training set of 'n' points.
of the form. $(\vec{x_1}, y_1), \cdots (\vec{x_n}, y_n)$ where $y_i \in (-1, 1)$

Problem stmt: Find max. margin hyperplane that divides the group of points $\vec{x_i}$ for which $y_i = 1$ from $y_i = -1$, which is so defined such that the distance between the hyperplane and the nearest point $\vec{x_i}$ from either group is maximized.

Some References

→

- Cortes C, Vapnik V. SVN.
- Bennett KP, Campbell C. Support vector machines.

SVM → Supervised learning Algorithm

Linear Model learning.

e.g. Finding a line which separates the data in 2D.

→ Data to be classified can be separated by a line.

→ $y = wx + b$.

→ infinite possibilities by varying 'w' and 'b'.

→ Use algorithm to determine which are the values of 'w' and 'b' giving the "best" line separating the data.

SVM and classification:

→ The original. Maximal Margin Classifier

→ Kernel Trick version

→ Soft Margin version

Soft Margin Kernelized version.

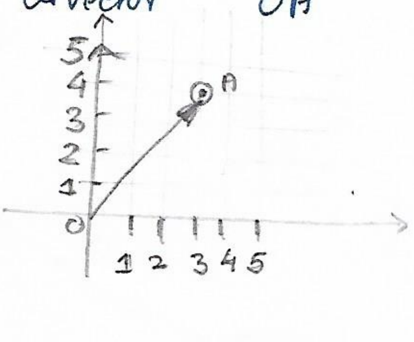
Regression: → using Support Vector Regression.

Part 1] → Margin.

Goal:] Goal of SVM is to find the optimal separating hyperplane which maximizes the margin of training data.

Revisiting the Basics.

A vector $\vec{OA}$ $A(3,4)$



Any point $x = (x_1, x_2)$, $x \neq 0$, in R^2

Specifies a vector in the plane

$\|OA\|$ is the norm. $\|\vec{OA}\|$ = length of segment $\vec{OA}$.

Direction $\vec{u}(u_1, u_2)$ = is $\vec{w}\left(\frac{u_1}{\ldots}, \frac{u_2}{\ldots}\right) \Rightarrow$ norm = 1

Understanding the equation of the hyperplane.

Equation of St. Line.

$y = ax + b$

Eqn. of hyperplane.

$y - ax - b = 0$

$w^T x = 0$

Now consider

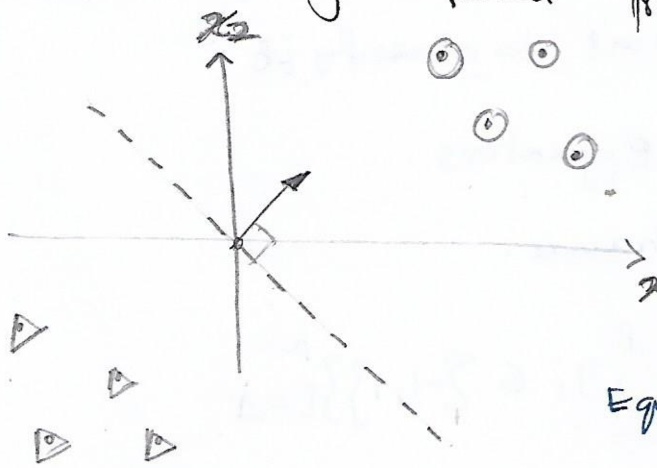
$y - ax - b = 0$

Given two vectors. $w \begin{pmatrix} -b \\ -a \\ 1 \end{pmatrix}$ $x \begin{pmatrix} 1 \\ x \\ y \end{pmatrix}$

$w^T x = -b(1) + (-a)x + 1 \times y$

$\therefore w^T x = y - ax - b$

Distance of a point from the hyperplane.

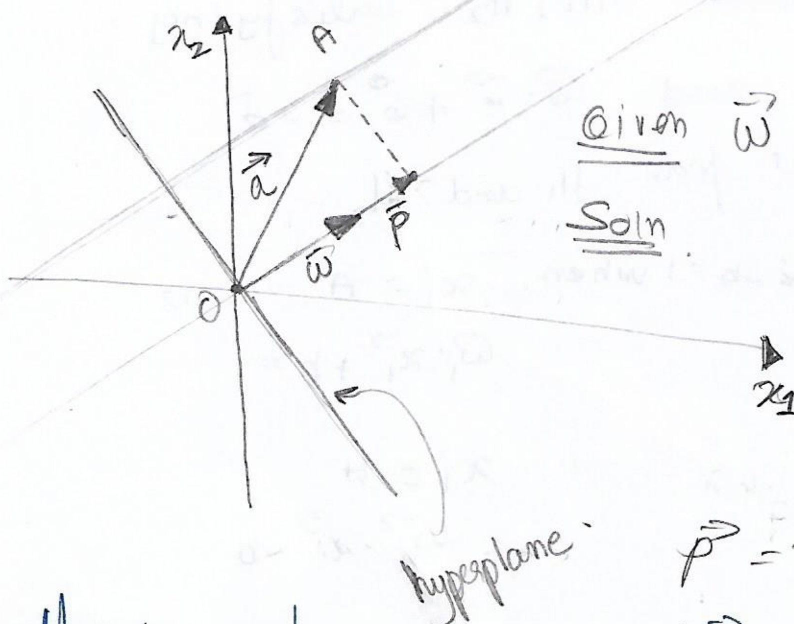


Eqn of hyperplane $x_2 = -2x_1$

$w^T x = 0$

with $w = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ and $x \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$

Say compute distance between (3,4) to the hyperplane



Given $\vec{w} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ $\vec{OA} = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$

Soln

$\|\vec{w}\| = \sqrt{2^2 + 1^2} = \sqrt{5}$

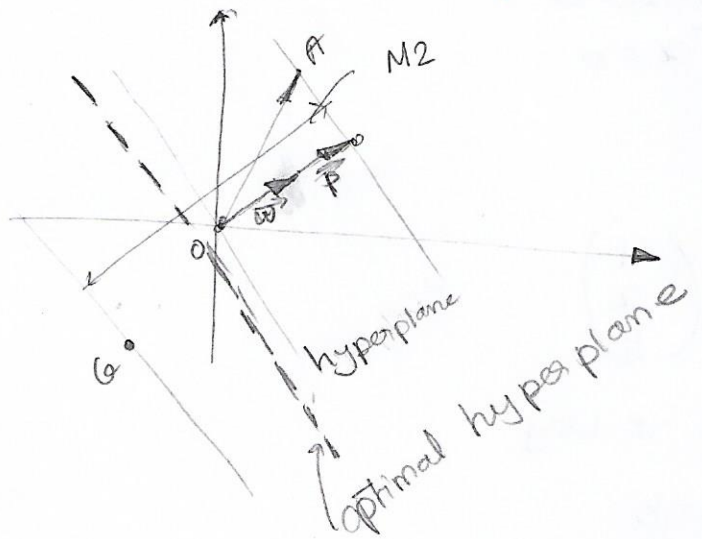
Let $\vec{u} = \frac{\vec{w}}{\|\vec{w}\|} = \left(\frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}}\right)$

$\vec{p} = (\vec{a} \cdot \vec{u})\vec{u}$

Margin of hyperplane $= 2\|\vec{p}\|$

Selecting the optimal hyperplane

→ we found distn: between point $\vec{A}$ and a plane as $\vec{p}$
margin is then $2\|\vec{p}\|$



Steps to Find biggest margin.

- Take dataset
- Select two hyperplanes which separate the data with no points between them
- Maximise the margin.

① You have data and want to classify it.

(x_i, y_i) → n rows, p factors.

initial Data set thus have.

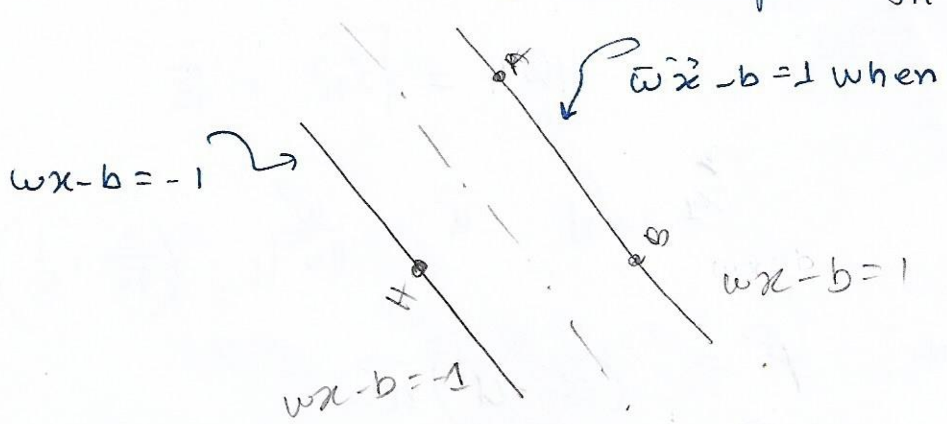
$$D = \{(x_i, y_i) \mid x_i \in R^p, y_i \in \{-1, 1\}\}_{i=1}^{n}$$

② Given a hyperplane H_0 → separating the dataset and satisfying $\vec{w} \cdot \vec{x} + b = 0$

Select 2 other hyperplanes H_1, H_2 satisfying.

$\vec{W} \cdot \vec{x} + b = \delta$ and $\vec{w} \cdot \vec{x} + b = -\delta$

So that H_0 is equidistant from H_1 and H_2.



$x_i = A$

$\vec{w_i} \cdot \vec{x_i} + b = 1$

$x_i = H$

$\vec{w_i} \cdot \vec{x_i} - b$

Getting the optimal hyperplane.

How to Find the biggest margin?

- Take the dataset.
- Select 2 hyperplanes with No data between them
- Maximize the distance.

a hyperplane → $w \cdot x + b = 0$

or $w(b, -a, 1)$ $x(1, x, y)$

$\underline{w \cdot x} = y - ax + b$] with 3D vectors.

with 2D] $w'(-a, 1) \cdot x'(x, y)$

- Consider the hyperplane H_0. Separating the dataset and satisfying $\vec{w} \cdot \vec{x} + b = 0$
- We can select 2 other hyperplanes H_1 and H_2. which have the following form.

(H_1) $\vec{w} \cdot \vec{x} + b = \delta$

and (H_2) $\vec{w} \cdot \vec{x} + b = -\delta$] Such that H_0 is equidistant from H_1 and H_2.

we can set $\delta = 1$

Thus we get $w \cdot x + b = 1$

and $w \cdot x + b = -1$

We select only those hyperplanes ... which.

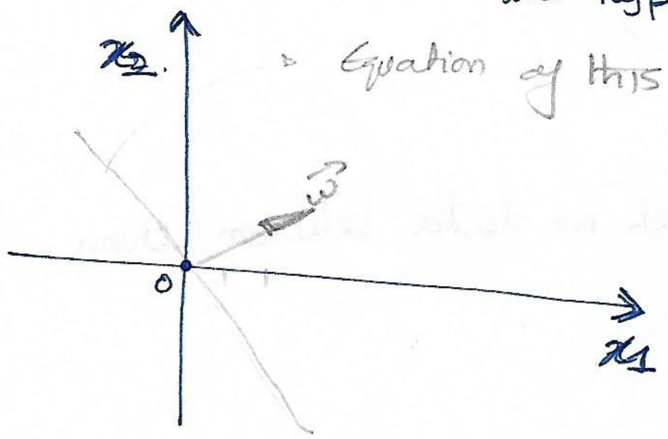
$w \cdot x_i + b \geq 1$ For class 1

and $w \cdot x_i + b \leq -1$ For class -1

writing together $y_i(w \cdot x_i + b) \geq 1$ For all $1 \leq i \leq n$.

Maximize distance between the two hyperplanes.

Consider in 2D. the hyperplane below (line)



▸ Equation of this Line | $x_2 = -2x_1$

| Thus $x_2 + 2x_1 = 0$

Say $\vec{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$

$\vec{w} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$

Thus $\vec{w} \cdot \vec{x}$ or $w^T x = 2x + x_2 = 0$

Thus Equation of hyperplane is

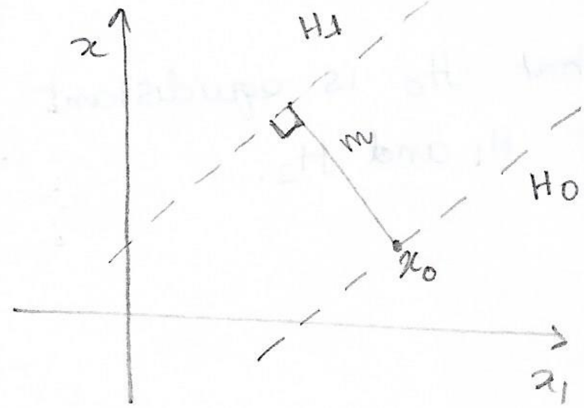
$\vec{w} \cdot \vec{x} = 0$. in 2 Dimensions.

$\vec{w}$ is a vector $\perp$ to the hyperplane

<u>Margin</u> : Consider the below Scenario.

$H_0 \to$ hyperplane $w \cdot x + b = -1$

$H_1 \to$ hyper plane $w \cdot x + b = 1$.



x_0 is a point on H_0

'm' is the distance between the hyperplanes.

Consider $H_1 = \vec{w} \cdot \vec{x} + b = 1$ Let $\vec{u} = \frac{\vec{w}}{\|w\|}$

Thus $\vec{k} = m\vec{u}$ $\|\vec{k}\| = m$

$\vec{z_0} = \vec{x_0} + \vec{k}$.

again $\vec{w} \cdot \vec{z_0} + b = 1$ or $\vec{w} \cdot (\vec{x_0} + \vec{k}) + b = 1$

or $\vec{w} \cdot \left(\vec{x_0} + \frac{m\vec{w}}{\|w\|}\right) + b = 1$

or $\vec{w} \cdot \vec{x_0} + m\frac{\vec{w} \cdot \vec{w}}{\|w\|} + b = 1$ $\Rightarrow$ $\boxed{\vec{w} \cdot \vec{x_0} + m\|w\| + b = 1}$

Now as x_0 is in H_0

$$\therefore \quad \vec{w} \cdot \bar{x}_0 + b = -1$$

$\therefore$ From. $\underbrace{\vec{w} \cdot \vec{x}_0} + m\,\|w\| \underbrace{+\, b} = 1$

we get $m\,\|w\| = 2$

or $m = \dfrac{2}{\|w\|}$

Thus maximizing the margin 'm' is the same as minimizing the norm of $\vec{w}$.

Thus we have the following optimization problem.

$$\left.\begin{array}{ll} \text{minimize in } (w, b) & \|w\| \\ \text{Subject to} & y_i\,(w \cdot x_1 + b) \geq 1 \end{array}\right]$$

A Constrained optimization problem.

Revisiting Math basics of optimization.

x^* being a Local minimum

Theorem : Let $f : \Omega \rightarrow R$ be continuously twice differentiable function at x^*.

If x^* Satisfies $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*)$ is +ve definite then x^* is a local minimum.

in other words

if 1) $\nabla f(x^*) = 0$

and the hessian of f at x^* is +ve definite

$$z^T\left(\nabla^2 f(x^*)\right) z > 0$$

where $\nabla^2 f(x) = \begin{pmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \cdots & \dfrac{\partial^2 f}{\partial x_1 \partial x_n} \\ \dfrac{\partial^2 f}{\cdots} & \cdots & \dfrac{\partial^2 f}{\partial \cdots^2} \end{pmatrix}$ then x^* is a Local minimum

$\nabla f(x^*) = 0$ means. $\frac{\partial f}{\partial x_1}(x^*) = 0$

$$\frac{\partial f}{\partial x_2}(x^*) = 0$$

$$\frac{\partial f}{\partial x_3}(x^*) = 0$$

Theorem: The following statements are equivalent

> Symmetric Matrix A is +ve def.
> All Eigen values of A +ve
> All Leading principal minors are +ve.
> There exists Non Singular sq. Matrix B such that $A = B^T B$

Computing the leading principal minors.

Minor. eg Compute Minor of M_{12} Pos.

$$\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} \rightarrow \text{Remove Row 1, Col 2.}$$

$\begin{pmatrix} d & f \\ g & i \end{pmatrix} \rightarrow$ compute its det $=$ ~~di~~ $di - fg$.

$\therefore M_{12} = di - fg$

Principal minor $i = j$

$\therefore M_{11} = ei - fh$

$M_{22} = ai - cg$

$M_{33} = ae - bd$.

leading principal minor

The leading principal minor of order 'k' is the minor of order 'k' obtained by deleting the last n-k rows and cols

$\therefore D_1 = a$ (Last 2 lines, cols deleted)

$D_2 = ae - bd$ (last line, col deleted)

$D_3 = a(ei - fh) - b(di - fg) + c(dh - eg)$

eg] Find minimum of function.

$$f(x,y) = (2-x)^2 + 100(y-x^2)^2$$

↳ Rosenbrock's banana function.

Compute $\nabla f(x,y)$

$$\nabla f(x,y) = \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{pmatrix}$$

$$\frac{\partial f}{\partial x} = 2(2-x)(-1) + 200(y-x^2)(-2x)$$

$$= -4 + 2x - 400xy + 400x^3$$

$$= 2(200x^3 - 200xy + x - 2)$$

$$\frac{\partial f}{\partial y} = 200(y-x^2)$$

For $\nabla f(x,y) = 0$

We need to solve

$$2(200x^3 - 200xy + x - 2) = 0$$

$$200(y-x^2) = 0$$

or

$$400x^3 - 400xy + 2x - 4 = 0 \quad \text{— (1)}$$

$$200y - 200x^2 = 0 \quad \text{(2)}$$

$2x$ (2) + (1) $= \quad 2x - 4 = 0$

$\Rightarrow x = 2$

and from $200(y-x^2) = 0$

$\Rightarrow y = 4$

Thus $(x,y) = (2,4)$.

Check if the point is a minima.

The Hessian is. $$\nabla^2 f(x,y) = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{pmatrix}$$

$$\frac{\partial^2 f}{\partial x \partial y} = -400x$$

$$\frac{\partial^2 f}{\partial y \partial x} = -400x$$

$$\frac{\partial^2 f}{\partial y^2} = 200$$

we have the following

$$\frac{\partial f}{\partial x} = 2(2-x)(-1) + 2 \times 100(y-x^2)(-2x) = 400(-xy+x^3)$$

$$\frac{\partial^2 f}{\partial x^2} = -2(-1) + 400(-y+3x^2)$$

$$= 2 - 400y + 1200x^2 = 1200x^2 - 400y + 2$$

Thus we have the Hessian Matrix as follows.

$$\nabla^2 f(x,y) = \begin{pmatrix} 3202 & -800 \\ -800 & 200 \end{pmatrix} \quad \text{at } x = 2,\ y = 4$$

Minors of Rank 1.

$M_{11} = 3202$

Minors of Rank 2 $= 3202 \times 200 - (-800) \times (-800)$

$= 400$

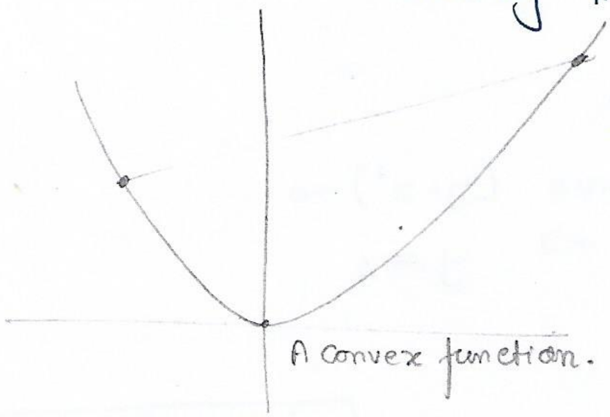
All Leading principal minors are +ve.

Thus the point $(2,4)$ is a local minimum.

Convex Functions

Q] What is a convex function?

-> If you can ~~trace~~ a line between two of its points without crossing the function line.

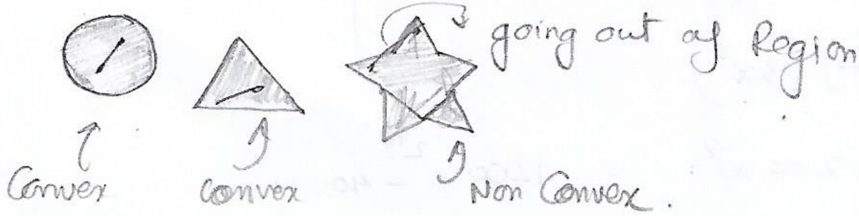


A convex function.

1. A function is convex if its epigraph (the set of points on or above the graph of the function) is a convex set.

Convex Set

in Euclidean space, a convex set is the region such that, for every pair of points within the region every point on the straight line segment that joins the pair of points is also within the region.



Understanding duality

Lower bound Consider subset of $\mathbb{R}$

$$S = \{2, 4, 8, 12\}$$

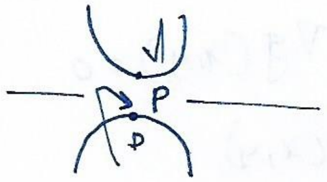
Because 1 is less than or equal 2, 4, 8, 12

1 is a Lower bound of S.

Given '2' is a (less than =) Lower bound.

'2' is called infimum. or greatest Lower bound.

Duality



$(P-D) > 0$

Typical way to write optimisation problem.

minimize $f(x)$

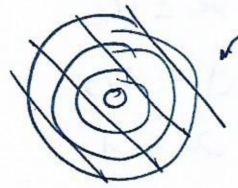
Subject to $g_i(x) = 0$, $i = 1, \ldots, p$

$h_i(x) \leq 0$, $i = 1, \ldots, m$.

Consider the below problem

Minimize $f(x, y) = x^2 + y^2$

Subject to $g_i(x, y) = x + y - 1 = 0$



$y = 1 - x$] This Line is the Feasible set.

Min of $f(x, y)$ under the constraint $g(x, y) = 0$ is obtained

$\Rightarrow$ When their gradients point in the same direction.

Mathematically → How does it translate?

$\nabla f(x,y) = \lambda \nabla g(x,y)$

↳ Lagrange Multiplies.

↳ Even if 2 vecs point in same direction, they may not have the same length. Some factor 'λ' allowing to transform one in the other.

How do we find points for which $\nabla f(x,y) = \lambda \nabla g(x,y)$?

Note we can write $\nabla f(x,y) - \lambda \nabla g(x,y) = 0$

define $L(x,y,\lambda) = f(x,y) - \lambda g(x,y)$

then its gradient $\nabla L(x,y,\lambda) = \nabla f(x,y) - \lambda \nabla g(x,y)$

eg Solve.] minimize $f(x,y) = x^2 + y^2$

subject to $g(x,y) = x + y - 1 = 0$

$L(x,y,\lambda) = f(x,y) - \lambda g(x,y)$

$\nabla L(x,y,\lambda) = \nabla f(x,y) - \lambda \nabla g(x,y)$

We solve $\nabla L(x,y,\lambda) = 0$

∴ Solve.

$$\left.\begin{array}{l} \frac{\partial L}{\partial x} = 0 \\ \frac{\partial L}{\partial y} = 0 \\ \frac{\partial L}{\partial \lambda} = 0 \end{array}\right] \quad \left.\begin{array}{l} 2x - \lambda = 0 \\ 2y - \lambda = 0 \\ -x - y + 1 = 0 \end{array}\right] > \begin{array}{l} x = 1/2 \\ y = 1/2 \\ \lambda = 1. \end{array}$$