Optimization : Constrained VS Unconstrained

Unconstrained Minimization / Maximization :-

Assume :-

Let $f: \Omega \to R$ be continuously differentiable function

Necessary and Sufficient Conditions for a local minimum / Maximum :-

x^* is local minimum / maximum of $f(x)$ iff

① f has zero gradient at x^* :

$$\nabla_x f(x^*) = 0$$

② And the hessian of f at x^* is +ve / -ve semi-definite :

$$v^t \left(\nabla^2 f(x^*)\right) v \underset{\leq}{\geq} 0 \;, \quad \forall v \in R^n$$

where

$$\nabla^2 f(x) := \begin{pmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{pmatrix}$$

Constrained optimization :

Problem :-

This is the constrained optimization we want to solve

$$\min_{x \in R^2} f(x) \quad \text{subjected to} \quad h(x) = 0$$

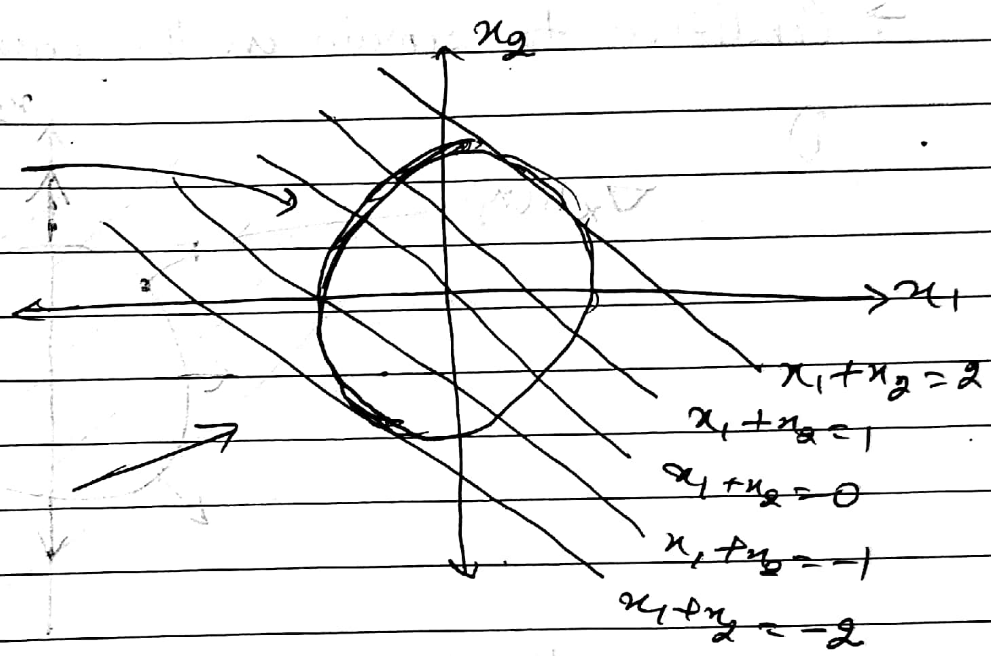
where

$$f(x) = x_1 + x_2 \quad \text{and} \quad h(x) = x_1^2 + x_2^2 - 2$$

→ **Cost function :-**

feasible region$;$

$h(x) = 0$

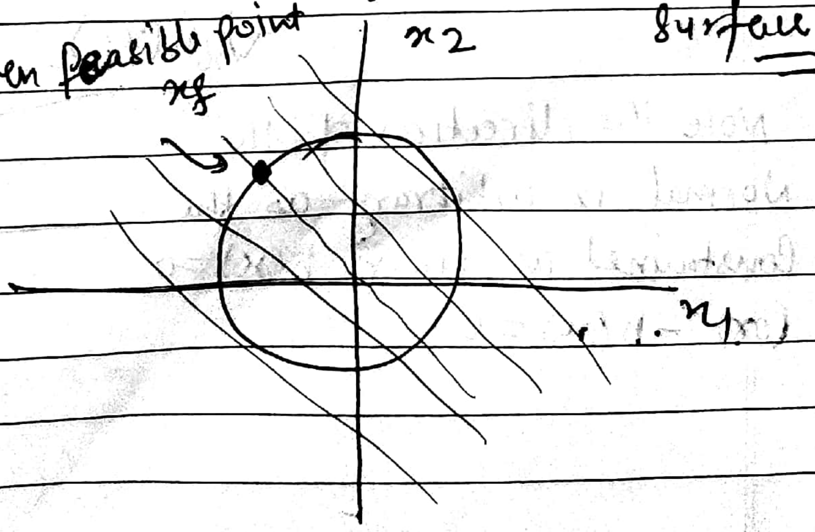


Iso-Contours of $f(x)$

$f(x) = x_1 + x_2$

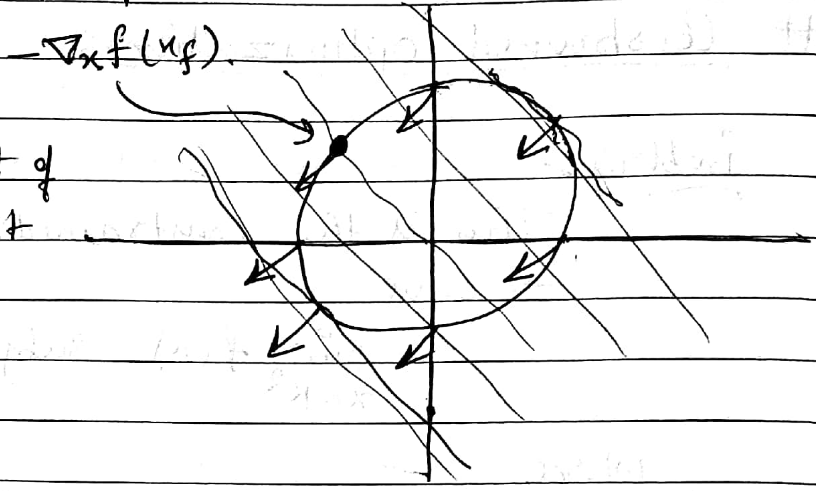
$h(x) = x_1^2 + x_2^2 - 2.$

→ Given feasible point x_f on the constraint surface



Gradient descent (Steepest descent

→ Condition to dec. cost function

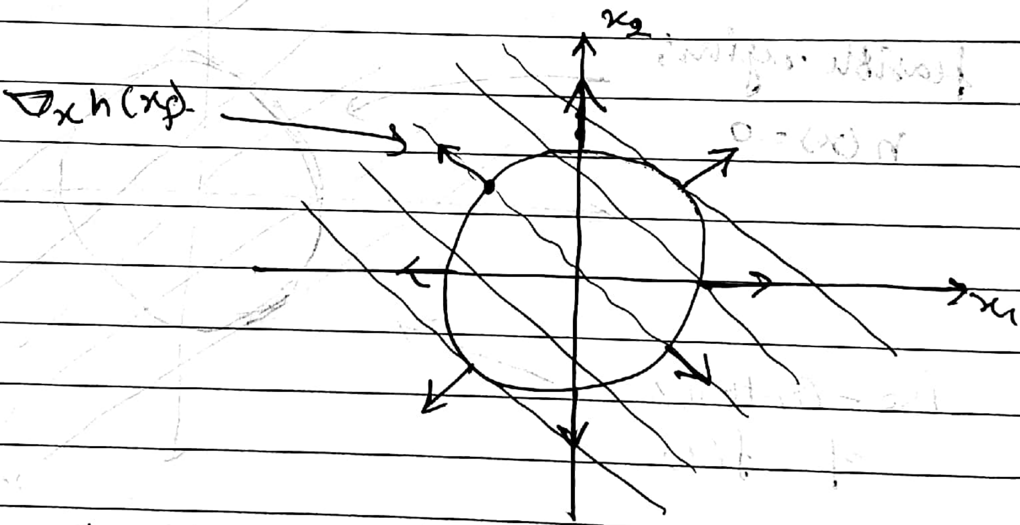


At any point $\tilde{x}$, the direct of steepest descent of the cost function $f(x)$ given by

$$-\nabla_x f(\tilde{x}).$$

→ Condition to remain on the constraint surface

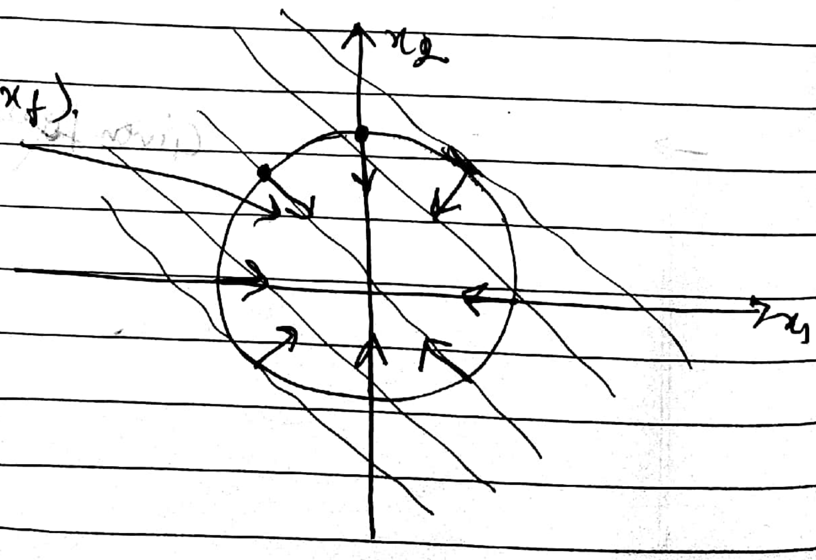
i)



Normal to the constraint surface are given by $\nabla_x h(x)$.

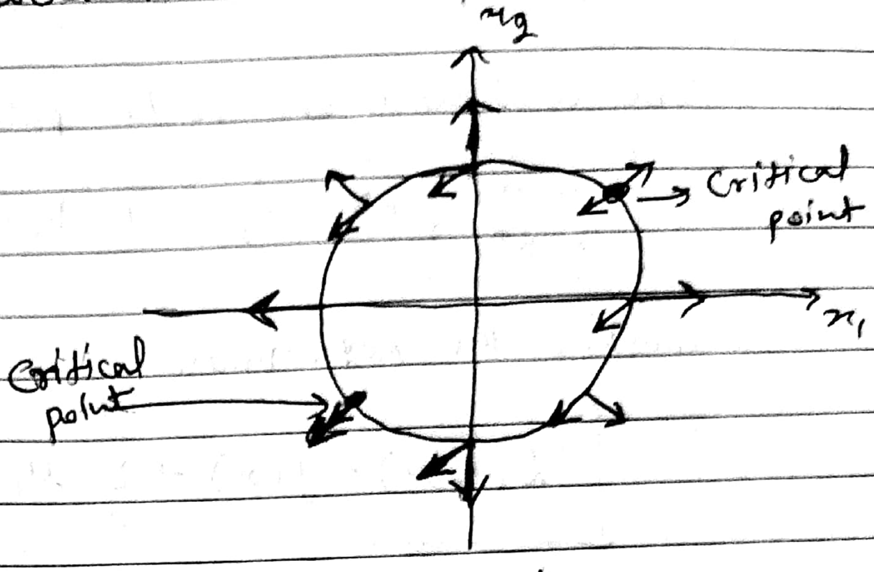
ii)

$\nabla_x h(x_f)$.



Note the direction of the normal is arbitrary as the constrained as either $h(x)=0$ (or) $-h(x)=0$

→ Condition for a local minimum



A Constrained local optimum occurs at x^* when $\nabla_x f(x^*)$ and $\nabla_x h(x^*)$ is parallel i.e

$$\boxed{\nabla_x f(x^*) = \mu \nabla_x h(x^*).}$$

→ From this fact Lagrange multipliers make sense :-

Our constrained optimization problem is

$$\boxed{\min_{x \in \mathbb{R}^2} f(x) \text{ subjected to } h(x) = 0}$$

Define Lagrangian is   note $L(x^*, \mu^*) = f(x^*)$

$$L(x, \mu) = f(x) + \mu h(x)$$

then x^* is a local min $\Longleftrightarrow$ there exists a unique μ^* s.t.

① $\nabla_x L(x^*, \mu^*) = 0$. ← encodes $\nabla_x f(x^*) = \mu^* \nabla_x h(x^*)$

② $\nabla_\mu L(x^*, \mu^*) = 0$ ← encodes the equality constraint $h(x^*) = 0$

③ $y^t (\nabla^2_{xx} L(x^*, \mu^*)) . y \geq 0 \quad \forall y$ s.t $\nabla_x h(x^*)^t y = 0$

+ve definite Hessian tells us we have a local minimum.

→ The Case of multiple equality Constraints :-

The Constrained optimization problem is

$$\boxed{\min_{x \in R^2} f(x) \text{ subject to } h_i(x) = 0 \text{ for } i = 1, \dots, l}$$

Construct the Lagrangian

$$\boxed{L(x, \mu) = f(x) + \sum_{i=1}^{l} \mu_i h_i(x) = f(x) + \mu^t h(x).}$$

Then x^* is local minimum $\Leftrightarrow$ there exists a unique μ^* s.t.

① $\nabla_x L(x^*, \mu^*) = 0$

② $\nabla_\mu L(x^*, \mu^*) = 0$

③ $y^t (\nabla^2_{xx} L(x^*, \mu^*)) y \geq 0 \; \forall y$ s.t. $\nabla_x h(x^*)^t y = 0$.

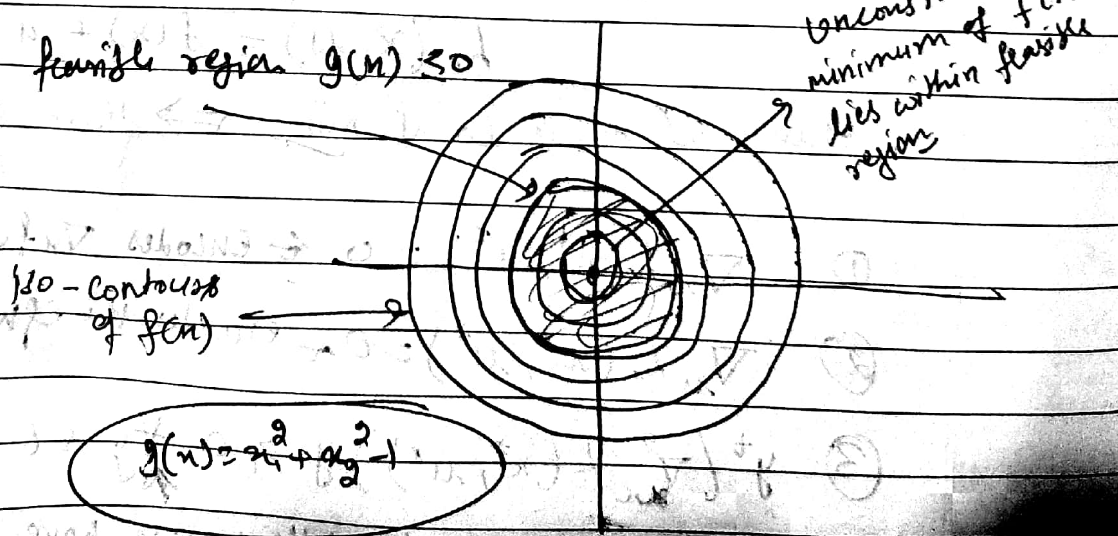
Example - ①

Problem :- Consider this Constrained optimization problem

$$\min_{x \in R^2} f(x) \text{ subjected to } g(x) \leq 0$$

where $f(x) = x_1^2 + x_2^2$ and $g(x) = x_1^2 + x_2^2 - 1$

Ans:



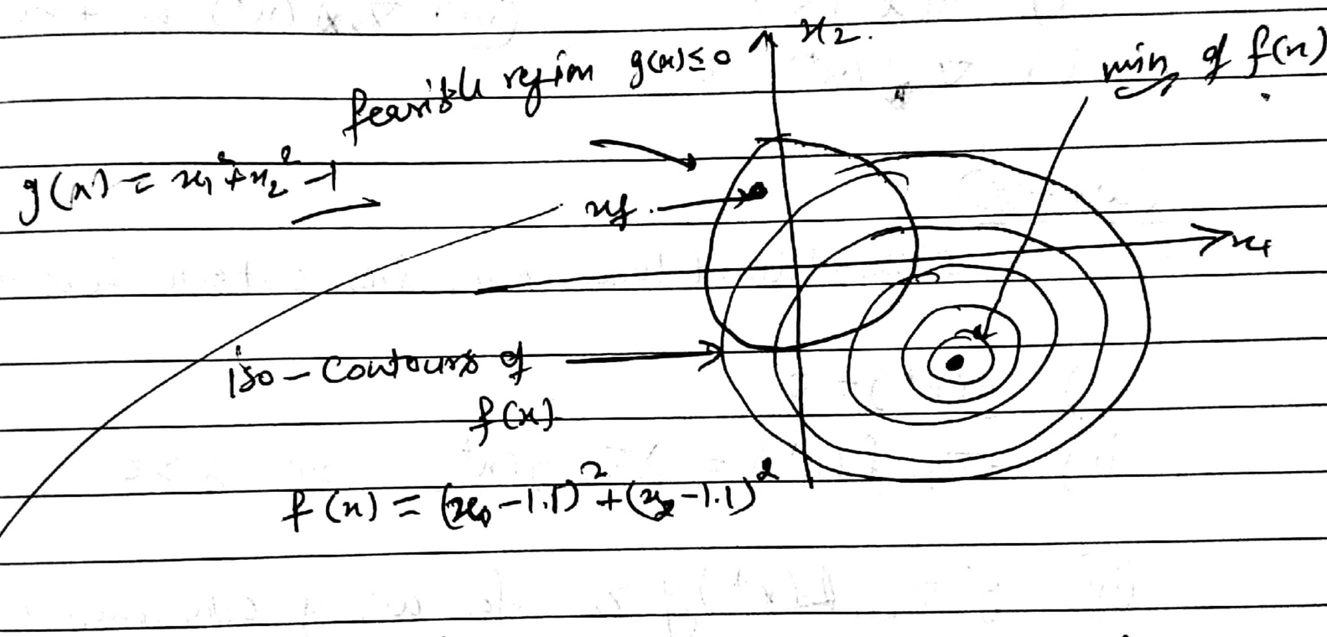
Necessary & sufficient condition for Constrained local minimum same as constrained local minimum

$\nabla_x f(x_f) = 0$ and $\nabla_{xx} f(x_f)$ is +ve definite.

Example - 2

where
$$f(x) = (x_1 - 1.1)^2 + (x_2 - 1.1)^2$$
and $g(x) = x_1^2 + x_2^2 - 1$

Sol^n



$\therefore$ the Constrained local minimum occurs on the surface of the Constraint surface.

Date
Page

Summary of optimization with one inequality constraint :-

Given $\boxed{\min_{x \in \mathbb{R}^2} f(x) \text{ subject to } g(x) \leq 0}$

If x^* corresponds to a constrained local minimum then

Case - 1

Unconstrained local min occurs in a feasible ~~set~~ region.

i) $g(x^*) < 0$

ii) $\nabla_x f(x^*) = 0$

iii) $\nabla_{xx} f(x^*)$ is +ve semi-definite matrix

Case - 2

Unconstrained local min lies outside feasible region:

i) $g(x^*) = 0$

ii) $-\nabla_x f(x^*) = \lambda \nabla_x g(x^*)$ with $\lambda > 0$

iii) $y^t \nabla_{xx} L(x^*) y \geq 0$ for all y orthogonal to $\nabla_x g(x^*)$

⇒ Karush-Kuhn-Tucker Conditions encode these Conditions

Scanned by CamScanner

KKT Conditions :-

Given optimization problem

$$\boxed{\min_{x \in \mathbb{R}^2} f(x) \text{ subject to } g(x) \leq 0}$$

Define the Lagrangian as

$$\boxed{L(x, \lambda) = f(x) + \lambda\, g(x)}$$

Thus x^* is local min $\Leftrightarrow$ there exists a unique λ^* s.t.

KKT Condition

1. $\nabla_x L(x^*, \lambda^*) = 0$
2. $\lambda^* \geq 0$
3. $\lambda^* g(x^*) = 0$
4. $g(x^*) \leq 0$
5. Plus +ve definite Constraint on $\nabla_{xx} L(x^*, \lambda^*)$

→ for multiple inequality Constraints

Given optimization problem

$$\boxed{\min_{x \in \mathbb{R}^2} f(x) \text{ subject to } g_j(x) \leq 0 \text{ for } j = 1, \dots, m}$$

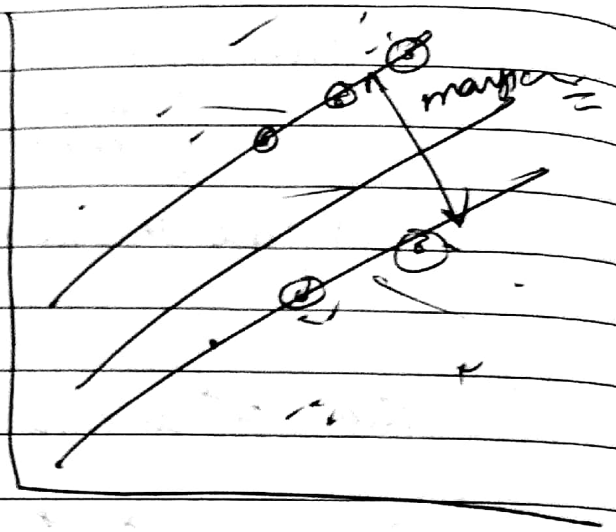
Define Lagrangian as

$$\boxed{L(x, \lambda) = f(x) + \sum_{j=1}^{m} \lambda_j g_j(x) = f(x) + \lambda^t g(x)}$$

Then x^* is local min $\Leftrightarrow$ there exists a unique λ^* s.t.

1. $\nabla_x L(x^*, \lambda^*) = 0$
2. $\lambda_j^* \geq 0$ for $j = 1, \dots, m$
3. $\lambda_j^* g_j(x^*) = 0$ for $j = 1, \dots, m$
4. $g_j(x^*) \leq 0$ for $j = 1, \dots, m$
5. Plus +ve definite Constraints on $\nabla_{xx} L(x^*, \lambda^*)$

SVM I :-



Suppose 2 lines

$w^T x + w_0 = -1$

$w^T x + w_0 = 0$

$w^T \cancel{x_0} + w_0 = 1$

Explain Concept

"Maximum margin intuition" width $= \frac{2}{||w||}$ Maximi

$\delta(w)$

mi mini $\frac{1}{2} w^2 = \frac{1}{2} w^T w$

$y_i(w^T x_i + w_0) \geq 1$

⇒ SVM : Lagrangian & KKT Conditions :-

$$\frac{\partial}{\partial w} L(w, w_0, \lambda) = 0$$

$$\frac{\partial}{\partial w_0} L(w, w_0, \lambda) = 0$$

$\lambda_i \geq 0$, $i = 1, 2 \ldots N$

$$\lambda_i \left[y_i (w^T x_i + w_0) - 1 \right] = 0, \quad i = 1, 2 \ldots N$$

$$\boxed{L(w, w_0, \lambda) = \frac{1}{2} w^T w - \sum_{i=1}^{N} \lambda_i \left[y_i (w^T x_i + w_0) - 1 \right]}$$

$$w = \sum_{i=1}^{N} \lambda_i y_i x_i$$

$$\sum_{i=1}^{N} \lambda_i w_i = 0$$

Wolfe Dual Representation :-

A Convex programming problem is equivalent to

$$\max_{\lambda \geq 0} L(\theta, \lambda).$$

$$\text{Subject } \frac{\partial}{\partial \theta} L(\theta, \lambda) = 0$$

The last eqⁿ guarantees that θ is a minimum of lagrangian

Q:- Consider quadratic problem

$$\text{minimize } \frac{1}{2}\theta^T\theta$$

$$\text{subject to } A\theta \geq b.$$

This is Convex Programming problem, hence wolf dual representation is valid

$$\text{maximize } \frac{1}{2}\theta^T\theta - \lambda^T(A\theta - b)$$

$$\text{subject to } \theta - A^T\lambda = 0$$

$$\Downarrow$$

$$\max_{\lambda} \left\{ -\frac{1}{2}\lambda^T A A^T \lambda + \lambda^T b \right\}$$

$$\text{subject to } \lambda \geq 0.$$

Quadratic problem but set of constraints is now simpler.

Lagrangian Duality method :-

maximize $L(w, w_0, \lambda)$

subject to $w = \sum_{i=1}^{N} \lambda_i y_i x_i$

$$\sum_{i=1}^{N} \lambda_i y_i = 0 \implies \max\left(\sum_{i=1}^{N} \lambda_i - \frac{1}{2} \lambda_i \lambda_j y_i y_j x_i^T x_j\right)$$

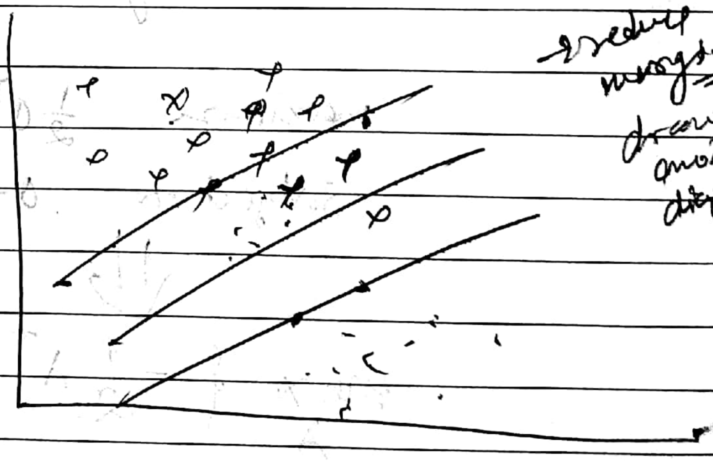
$\lambda \geq 0$

subject $\sum_{i=1}^{N} \lambda_i y_i = 0$

$\lambda \geq 0.$

SVM II :-

Linearly non-separable data : Search for line with errors



reduce margin

draw another diagram

Introduce slack variables:

→ ~~Vector fall~~ $\xi_i = 0$

Vector that fall outside the band & are correctly classified.

→ $0 < \xi_i \leq 1$

Vector falling inside the band & are correctly classified

$$0 \leq y_i(w^T x_i + w_0) < 1$$

- $\xi_i > 1$
vector that are missclassified.

$$y_i(w^T x + w_0) < 0$$

$$y_i(w^T x + w_0) \geq 1 - \xi_i \Rightarrow \text{slack variable}$$

(2) SVM II Cost function : Lagrangian & KKT :-

minimize $J(w, w_0, \xi_i) = \frac{1}{2}\|w\|^2 + C \sum_{i=1}^{N} \xi_i$

Subject to $y_i(w^T x_i + w_0) \geq 1 - \xi_i, \quad i = 1, 2, \ldots N$

$$\xi_i \geq 0, \quad i = 1, 2, \ldots N$$

$$L(w, w_0, \xi, \lambda, \mu) = \frac{1}{2}\|w\|^2 + C \sum_{i=1}^{N} \xi_i - \sum_{i=1}^{N} \mu_i \xi_i$$

$$- \sum_{i=1}^{N} \lambda_i [y_i(w^T x_i + w_0) - 1 + \xi_i]$$

$\frac{\partial L}{\partial w} = 0$ or $w = \sum_{i=1}^{N} \lambda_i y_i x_i$

$\frac{\partial L}{\partial w_0} = 0$ or $\sum_{i=1}^{N} \lambda_i y_i = 0$

$\frac{\partial L}{\partial \xi_i} = 0$ or $C - \mu_i - \lambda_i = 0$

$\lambda_i [y_i(w^T x_i + w_0) - 1 + \xi_i] = 0, \quad i = 1, 2, \ldots N$

$$\mu_i \xi_i = 0, \quad i = 1, 2, \ldots N$$

$\mu_i \geq 0, \lambda_i \geq 0. \quad i = 1, 2, \ldots N$

Scanned by CamScanner

classmate

Date

Page

⑤ Dual form

maximize $L(w, w_0, \lambda, \varepsilon, \mu)$

Subject to $w = \sum_{i=1}^{N} \lambda_i y_i x_i$

$\sum_{i=1}^{N} \lambda_i y_i = 0$, $C - \mu_i - \lambda_i = 0$, $i = 1, 2, \dots N$

$\lambda_i \geq 0, \mu_i \geq 0$

$\Downarrow$

$$\max_{\lambda} \left(\sum_{i=1}^{N} \lambda_i - \frac{1}{2} \sum_{i,j} \lambda_i \lambda_j y_i y_j x_i^T x_j \right)$$

Subject to $0 \leq \lambda_i \leq C$, $i = 1, 2, \dots N$

$\sum_{i=1}^{N} \lambda_i y_i = 0$

SVM III :- Kernel trick

$\vec{x} = (x_1, x_2)$

$\vec{z} = (z_1, z_2)$

$K(\vec{x}, \vec{z}) = (\vec{x}^T \vec{z})^2$

$= (x_1 z_1 + x_2 z_2)^2$

$= (x_1^2 z_1^2 + x_2^2 z_2^2 + 2 x_1 z_1 x_2 z_2)$

$= (x_1^2, \sqrt{2} x_1 x_2, x_2^2)(z_1^2, \sqrt{2} z_1 z_2, z_2^2)$

$= \phi(\vec{x})^T \phi(\vec{z})$

Recall SVM in dual form :-

$$g(x) = w^T x + w_0$$

$$= \sum_{i=1}^{N_s} \alpha_i y_i x_i^T x + w_0$$

assign x in $\omega_1 (\omega_2)$ if $g(x) = \sum_{i=1}^{N} \alpha_i y_i K(x_i, x_j) < (>0)$

$$\max_{\lambda} \left(\sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j x_i^T x_j \right)$$

subject to $0 \le \alpha_i \le C$, $i = 1, 2, \dots N$

$$\sum_{i=1}^{N} \alpha_i y_i = 0$$

$$\Downarrow$$

$$\max_{\lambda} \left(\sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \right)$$

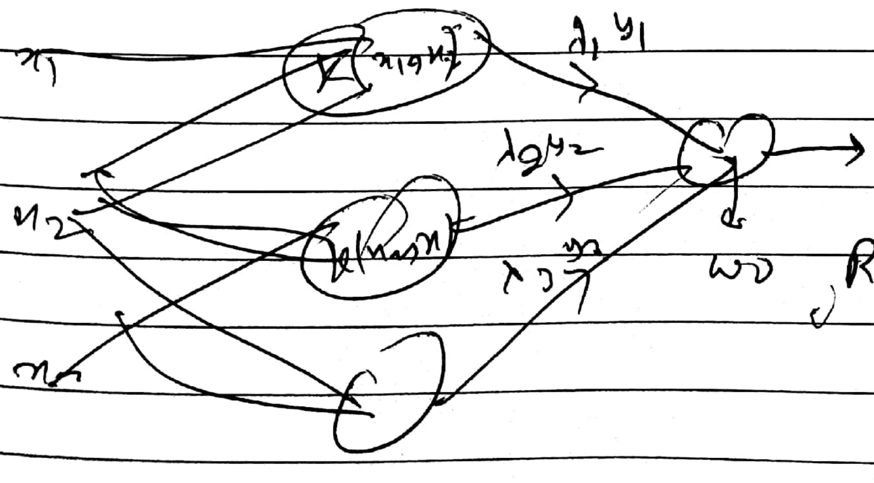
subject to $0 \le \alpha_i \le C$, $i = 1, 2, \dots N$

$$\sum_i \alpha_i y_i = 0$$

Types of kernal

$$\cancel{g(x) = w^T x + w_0}$$

$$= \sum_{i=1}^{N} \alpha_i y_i x_i^T x + w_0$$



$$RBF(x, x') = \exp\left(-\frac{\|x - x'\|^2}{2\sigma^2}\right)$$