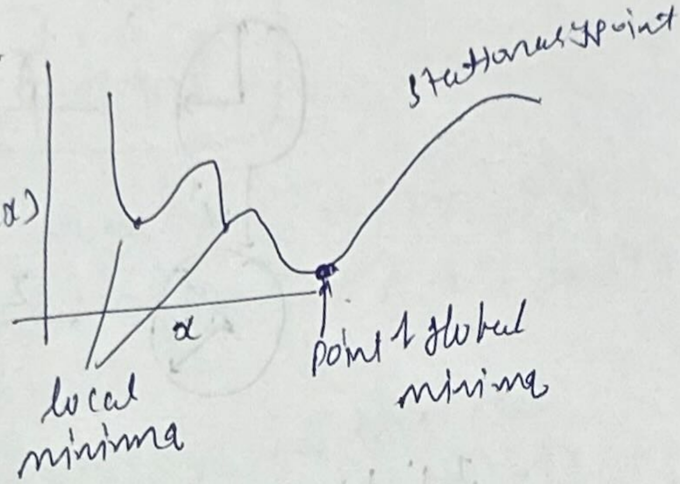
* Why optimization for ML?

* Convex Optimization:



$f: \mathbb{R}^n \to \mathbb{R}$ ← Constrained

$\min f(x)$

s.t. $x \in X \subseteq \mathbb{R}^n$

Unconstrained optimization

$\min f(x)$

s.t. $x \in \mathbb{R}^n$

- SVM - Constrain optimization problem.

- Unconstrained $\min_{x \in \mathbb{R}^d} \|Ax - b\|_2^2 + \lambda \|x\|_1$ (Lasso)

$\lambda > 0$

- Constrained $\min_{x \in \mathbb{R}^d} \|Ax - b\|_2^2$ (Lasso)

s.t. $\underline{\|x\|_1 \leq K}$

* Imp. Results / Definitions

- let $f: \mathbb{R}^n \to \mathbb{R}$ be continuously $\underline{\text{differentiable}}$

x^* is pt. of local optimum $\Rightarrow$ $\nabla f(x^*) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix}$

$\underline{0} =$

this is ~~necess~~ necessary Condition

mean every direction gradients $= 0$.

→ let's F twice differentiable

$$H(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1} & \frac{\partial^2 f(x)}{\partial x_1 x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ & & & \\ & & & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix}$$

~~Hessian~~

Hussian Mattix

$$\boxed{\frac{\partial^2 f(x)}{\partial x_1 x_2} = \frac{\partial}{\partial x_1}\left(\frac{\partial f(x)}{\partial x_2}\right)}$$

then $\nabla f(x^*) = 0$ & $H(x^*)$ is P.D. ⇒ x^* is a P. t. d. (point) local optimum (min)

↑ satisfulutes condition

(*) Positive ~~Mattix~~ / Negative definiteness → scaler value

→ PD → $A^{n \times n}$ is P.D. mattix if $x^T A x > 0$ $\forall x \in \mathbb{R}^n$ $x \neq 0$

(Positive definitness)

all eigenvalues are +ve.

→ if $x^T A x \geq 0$ → P.D

~~N.D~~

ND ⇒ non -ve values

$x^T A x < 0$ → N.D.

* $x^T A x \geq 0$ → Positive Semi definute Mattix

* $x^T A x \leq 0$ → N S D.

$x^T A x = \|x\|_A$ → Muhu~~lanobis~~ molenous norm.

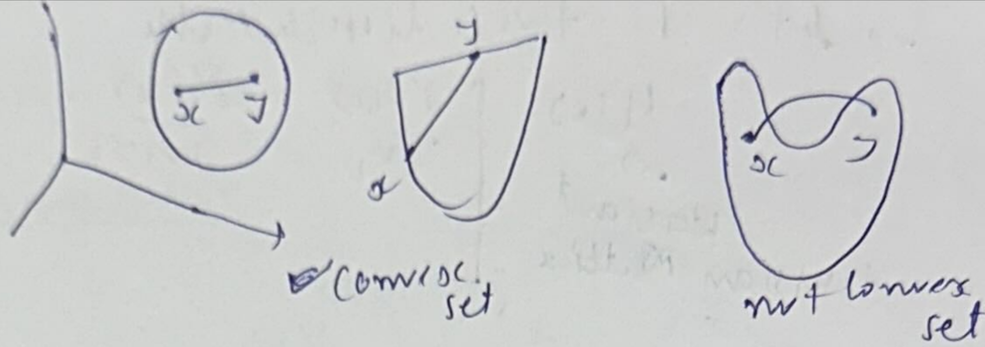
(*) Convex Set: $X \subseteq \mathbb{R}^n$ is said to be convex if $\forall x, y \in X$

$\lambda x + (1-\lambda) y \in X$ $\forall$ ~~λ~~ $0 \leq \lambda \leq 1$

→ this Convex Combinations. ↓ gives line segment of x & y.

~~λx~~ $\alpha x + \beta y$

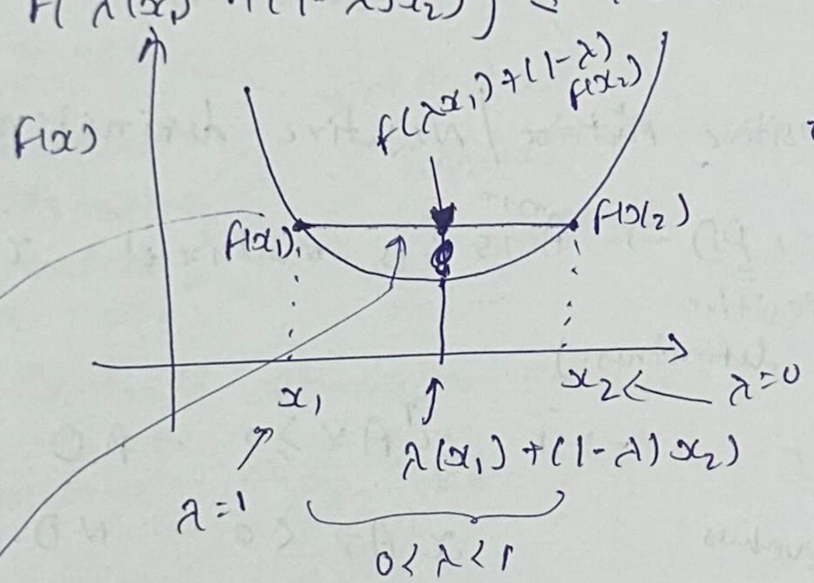
→ x, y ⇒ linear combination of two vectors



* Convex ~~set~~ function : $f: R^n \to R$ is define over a Convex set in following way

$\forall x_1, x_2$

$$f(\lambda(x_1) + (1-\lambda)x_2) \le \lambda f(x_1) + (1-\lambda) f(x_2)$$



$f(\cdot)$ is Convex $\iff H(x)$ is P.S.D. $\forall x \in X$

-> for convex function it is always above the Convex set of values

-> but at $\lambda = 0, \lambda = 1 \Rightarrow f(x)$ value same as Curve value

* why Convex fn important → b'coz local minima = global minima

* gradient descent ⇒ unconstrained function.

* ~~Convex~~ Continued Convex Optimization

let $f_0 : R^n \to R$ be some f^n

$$\min f_0(x)$$
$$\text{s.t. } f_i(x) \leq 0 \quad \forall i = 1 \ldots m$$
$$f_j(x) = 0 \quad \forall j \ 1 \ldots l$$

→ if $f_i(x) \geq 0$ → contain we need to convert in ≤ 0 by multiplying '−'.

$$2x + 3y \geq 0$$
$$-2x - 3y \leq 0$$

⇒ let here $\lambda_1, \lambda_2 \ldots \lambda_m$ & $u_1, u_2 \ldots u_l$

define (for $\lambda_i \geq 0$)

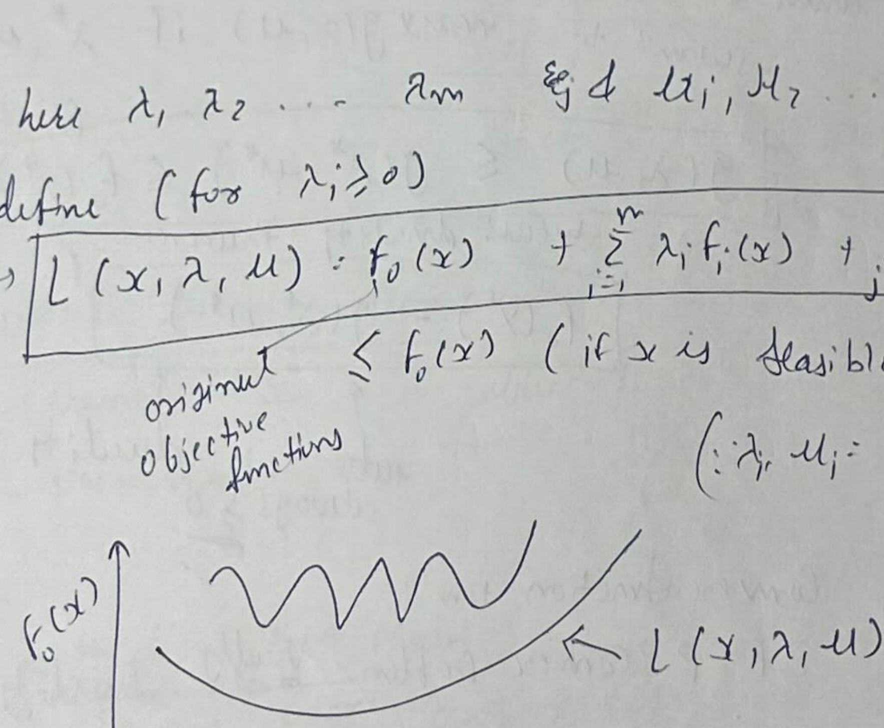
langrangian function →

$$\boxed{L(x, \lambda, u) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{l} u_j f_j(x)}$$

original objective functions

$\leq f_0(x)$ (if x is feasible)

($\because \lambda_i, u_j$ = vectors)

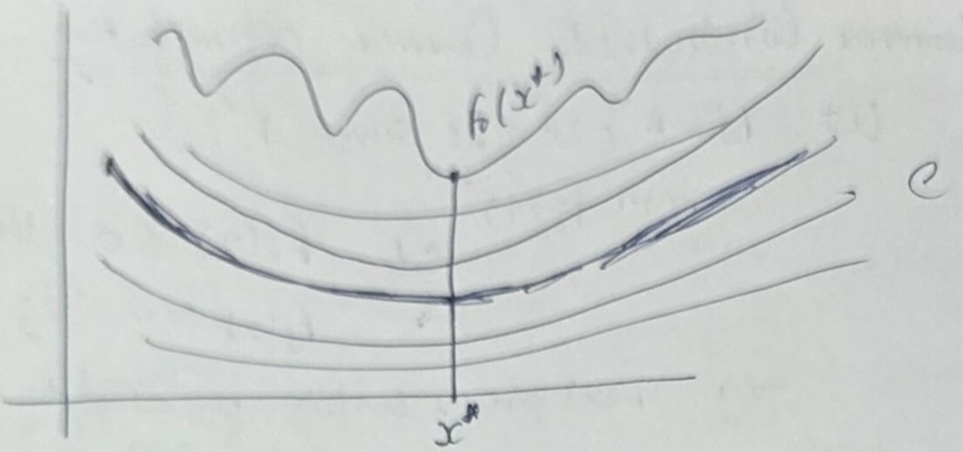


$$\boxed{L(x, \lambda, u) \leq f_0(x)}$$

$$\min_x L(x, \lambda, u) \leq f_0(x^*)$$
$$\text{s.t. } \lambda \geq 0$$

$$\boxed{\max_{(\lambda, u)} \left(\min_x L(x, \lambda, u) \atop \lambda \geq 0 \right) \leq f_0(x^*)}$$

Langrangian dual



$g(\lambda, \mu) = \min_{x} L(x, \lambda, \mu)$

s.t. $\lambda \succeq 0$

dual function

→ want to $\max g(\lambda, \mu)$ if λ^*, μ^* are point 1 maxima

$$\boxed{g(\lambda, \mu) \leq g(\lambda^*, \mu^*) \leq f_0(x^*)}$$

weak duality theorm

$$\boxed{f_0(x^*) - g(\lambda^*, \mu^*)}$$

duality gap

~~zero~~ always ≥ 0

* why Convex function imp(?)
  If P convex problem usually duality gap zero

* KKT Conditions :-

let x^* & (λ^*, μ^*) be any primal & dual optimal point with zero duality ~~bat~~ gap. then we must have

1. $f_i(x^*) \leq 0$ & $f_j(x^*) = 0$ $\forall i, j$ (Primal ~~feas~~ fesibity) (PF)
2. $\lambda^* \geq 0$ (DF) (dual fesibility)

~~(3) $\lambda_i^* f_i(x^*) = 0$ $\forall i$~~

3. $\lambda_i^* f_i(x^*) = 0$ ((CS) complimentry slaeness)
4. $\nabla_x L(x, \underbrace{\lambda^*, \mu^*}_{\text{fix the this value \& x will change}}) = 0$ (LO) (Langrange optimus)