Lecture 1: Introduction, Convex Sets and Convex Functions

COL8385/870: Special Topics in ML

Course: Optimization for Machine Learning

Instructor: Adarsh Barik

Days/time: Tuesday/Friday, 3:30-5:00pm

Office Hours: TBD

Classroom: LH 620

Contact: adarshbarik1@cse.iitd.ac.in

Bharti 422

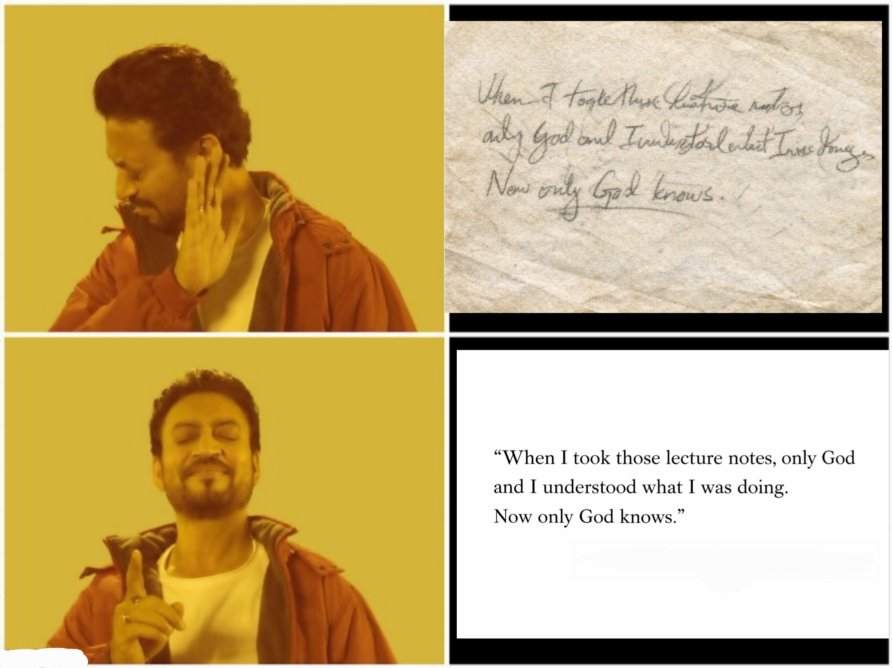
Website: https://adarsh-barik.github.io/courses/col870/col870_web.html

Reference textbook

1. Convex Optimization. S. Boyd and L. Vandenberghe. Cambridge University Press, Cambridge, 2003
2. Introductory Lectures on Convex Optimization: A Basic Course. Y. Nesterov. Kluwer, 2004.
3. Numerical Optimization. J. Nocedal and S. J. Wright, Springer Series in Operations Research, Springer-Verlag, New York, 2006 (2nd edition).
4. A Modern Introduction to Online Learning. Francesco Orabona. arXiv preprint arXiv:1912.13213 (2019).

Tentative Grading Scheme

Midsemester Exam	30%
Final Exam	30%
Assignments	25%
Scribe	10%
In-class participation	5%



Scribe

- Each student will be assigned a lecture to scribe.
- Scribing must be done in LaTeX, and I will provide the required style files and template.
- Scribe will be due within one week of the lecture date.

- Need a volunteer for today's lecture (Poojan Shah)
- (And for the next few lectures - until the roll list is finalized)

AI Everywhere

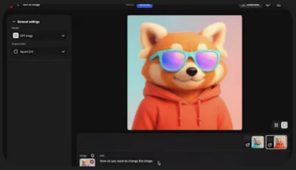
what does it mean to have AI-enabled app?

[redacted]

Training is the Key!

OpenAI

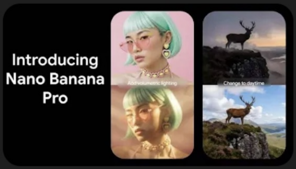
Introducing our latest image generation model in the API



23 Apr

Times of India

Google releases Gemini 3-powered Nano Banana Pro image model: Key features, how to use and how it differs



22 Nov

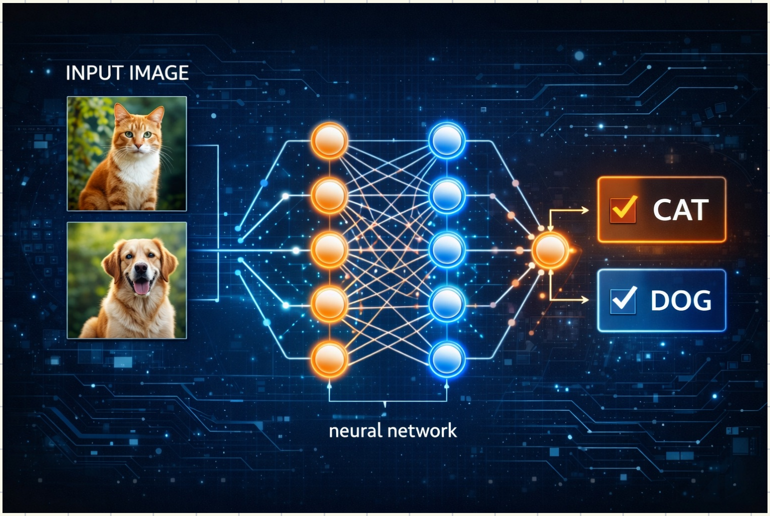
Fortune

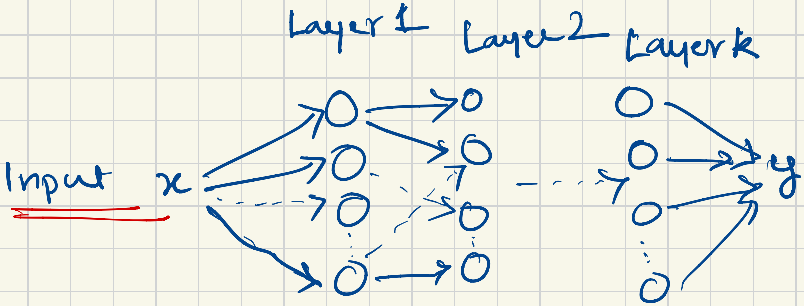
OpenAI releases new image model as it races to outpace Google's Nano Banana amid company code red

15 days ago • By Sharon Goldman

$\boxed{\text{Training data}} \longrightarrow \boxed{\text{ML Algorithm}} \longrightarrow \boxed{\text{Trained ML Model}}$

Training a Neural Network





$$x \longrightarrow \sigma_1(W_1 x + b_1) \longrightarrow \sigma_2\left(W_2\,\sigma(W_1 x + b_1) + b_2\right) \dashrightarrow y$$

Training: Finding the weights (W) that minimize the training loss over the training data

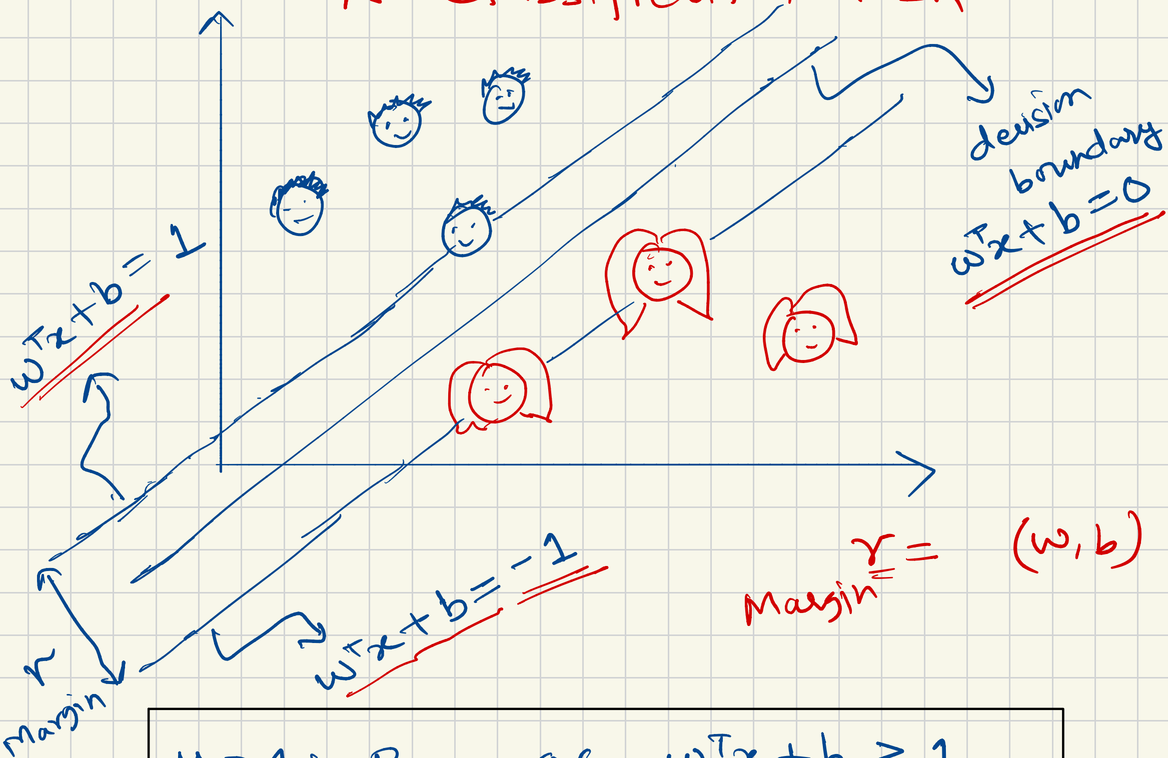
$$\underset{W}{\text{minimize}} \quad \text{Training\_Loss}(W, \text{Data})$$

$$\text{such that} \quad W \in \mathcal{W}$$

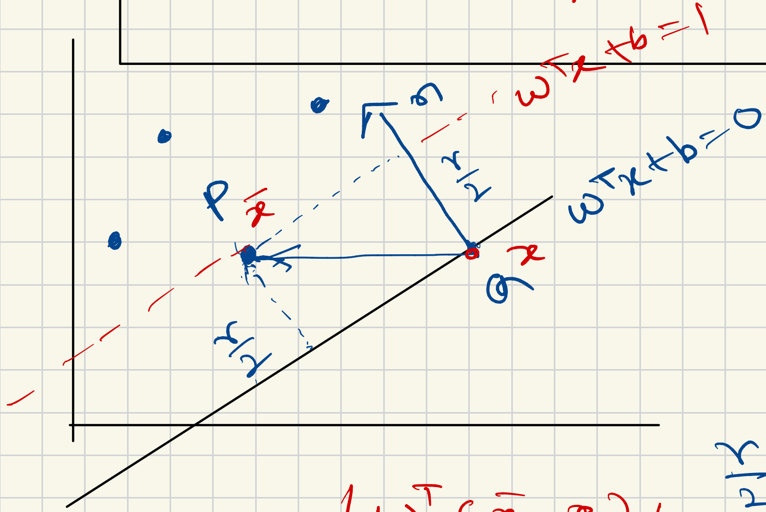
→ some constraint set

How to solve this optimization problem?

A Classification Task



$y=1$: Boy: if	$w^Tx+b \geq 1$
$y=-1$: Girl: if	$w^Tx+b \leq -1$



normal $\vec{n} = w$

$$\frac{\gamma}{2} = \frac{|\overrightarrow{QP} \cdot \vec{n}|}{\|\vec{n}\|}$$

$$\frac{\gamma}{2} = \boxed{\frac{|\overrightarrow{QP} \cdot w|}{\|w\|}}$$

$$|w^T(\bar{x}-x)| = 1$$

$$\frac{r}{2} = \frac{1}{||w||}$$

$$\boxed{\Rightarrow \quad r = \frac{2}{||w||}}$$

$$\text{maximize } r = \text{minimize } ||w||$$

$$= \text{minimize } ||w||^2$$

$$\text{minimize}_{w,b} \quad ||w||^2$$

such that

$$\underbrace{y_i (w^T x_i + b) \geq 1}, \quad \forall (x_i, y_i)$$

$$i = 1, \dots, n$$

$$\text{minimize}_{w} \quad f(w)$$

objective function

decision variable

such that

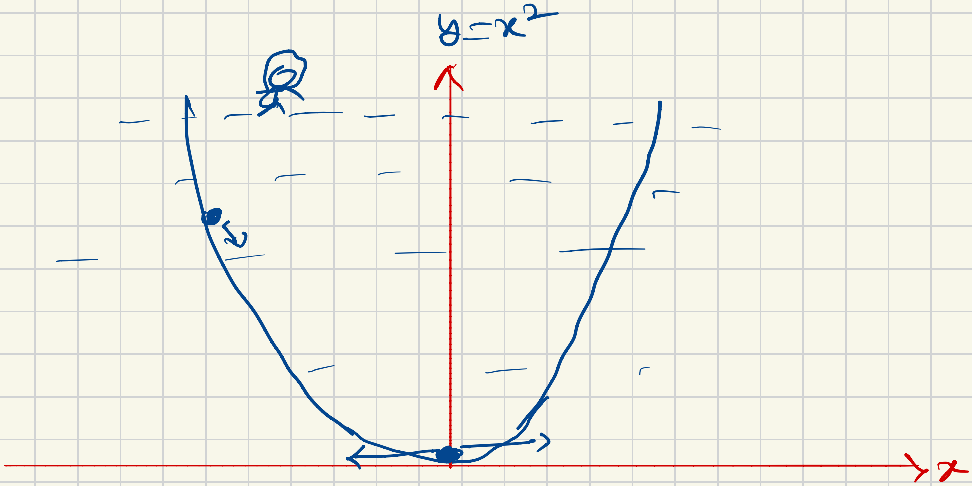
$$g_i(w) \leq 0, \quad i = 1, \dots, n$$

inequality constraints

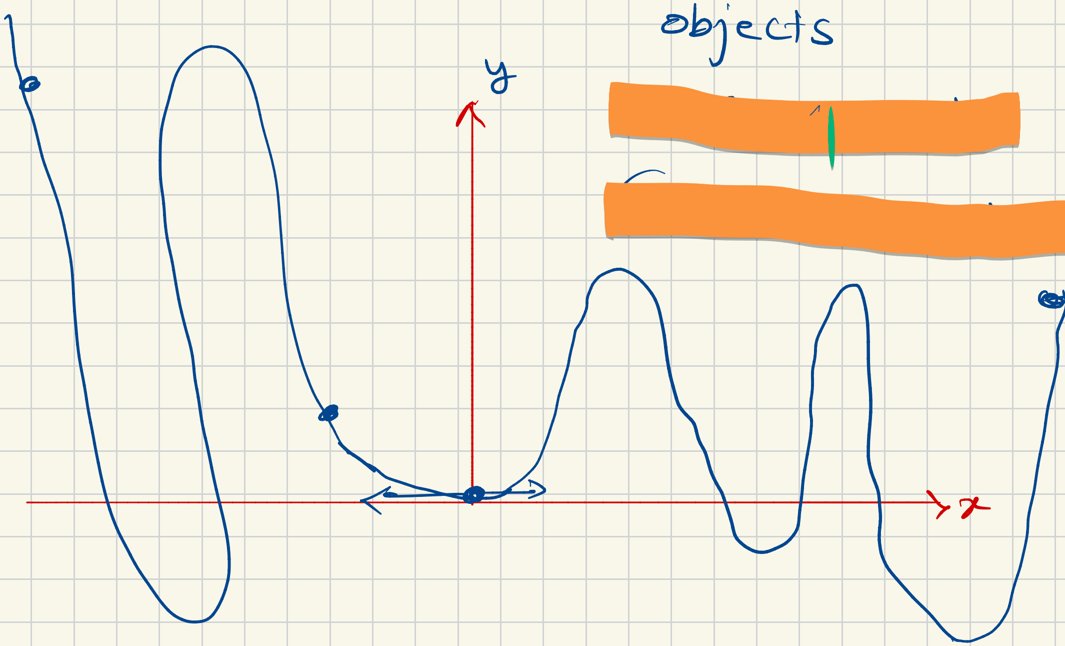
$$h_i(w) = 0, \quad i = 1, \dots, m$$

equality constraints

How to Minimize a function ?



Two important objects



NEXT

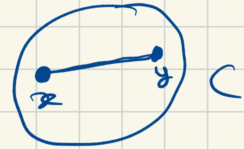
1. Convex sets

- (a) Examples
- (b) Operations that preserve convexity
- (c) Generalized inequality
- (d) separating and supporting hyperplanes

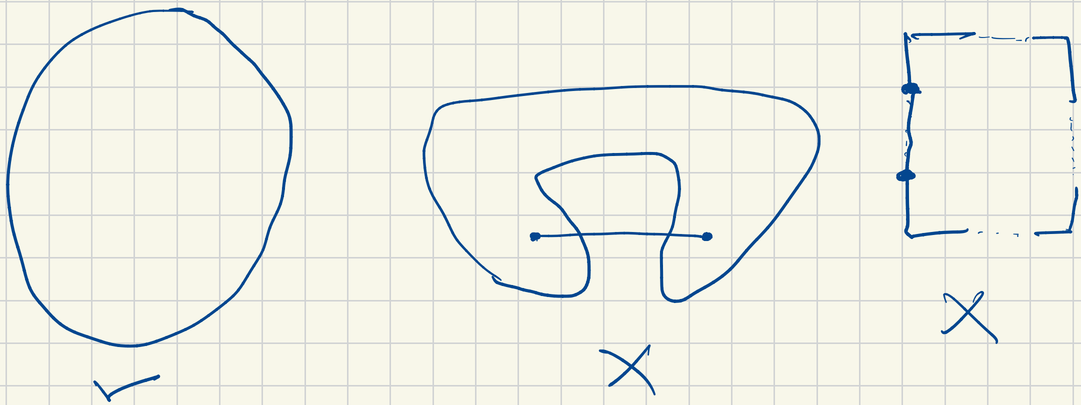
2. Convex functions

- (a) Examples
- (b) Operations that preserve convexity
- (c) Conjugate functions
- (d) Beyond convexity

Convex sets



Def: A set $C \subseteq \mathbb{R}^d$ is called convex, if for any $x, y \in C$, the line segment $\theta x + (1-\theta) y$ $(0 \le \theta \le 1)$ also lies in C.



combination of vectors

① Linear : $\theta_1 x + \theta_2 y$, $x, y \in C$

② Conic : if $\theta_1, \theta_2 \geq 0$

③ Convex : $\theta_1, \theta_2 \geq 0$, $\theta_1 + \theta_2 = 1$

Examples

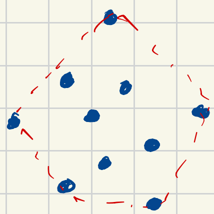
① Convex hull: $S = \{x_1, \ldots, x_k\} \subseteq \mathbb{R}^d$

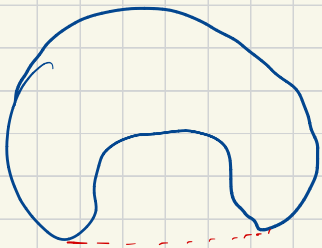
Their convex hull is

$$\text{conv}(x_1, \ldots, x_k) \triangleq \left\{ \sum_{i=1}^{k} \theta_i x_i \;\middle|\; \theta_i \geq 0, \sum_{i=1}^{k} \theta_i = 1 \right\}$$

or

$\text{conv}(S)$





② Convex cone: set that contains all conic combinations of points in the set

Example:

Lorentz cone $\{(x, t) \in \mathbb{R}^n \times \mathbb{R}_{++} \mid \|x\|_2 \leq t\}$

(norm cone)

$$C = \{(x,t) \in \mathbb{R}^n \times \mathbb{R}_{++} \mid \|x\|_2 \le t\}$$

$(x,t), (y,\bar{t}) \in C$ $\lambda \in [0,1]$

To show:

$$(\lambda x + (1-\lambda) y, \ \lambda t + (1-\lambda)\bar{t}) \in C$$

$$\underline{\|\lambda x + (1-\lambda) y\|_2}$$

$$\le \|\lambda x\|_2 + \|(1-\lambda) y\|_2$$

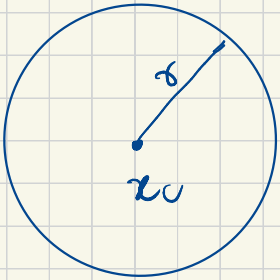
$$\le \underline{\lambda \cdot t} + \underline{(1-\lambda)} \cdot \bar{t}$$

③ Hyperplane : $\{x \mid a^T x = b\}$, $a \neq 0$

④ Halfspace : $\{x \mid a^T x \leq b\}$, $a \neq 0$

⑤ Euclidean ball

$$B(x_c, r) = \{x \mid \|x - x_c\| \leq r\}$$



⑥ Ellipsoid

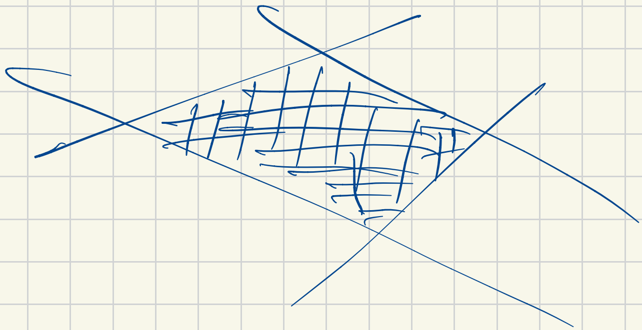
$$\{x \mid (x - x_c)^T A (x - x_c) \leq 1\}$$

$A \in S^n_{++}$ symmetric positive definite

⑦ Polyhedron:

component-wise

$$\{x \mid Ax \leq b,\ Cx = d\}$$



⑧ Positive semidefinite cone

ⓐ S^n : set of symmetric $n \times n$ matrix

ⓑ S^n_+ : $\{X \in S^n \mid \underbrace{X \succeq 0}\}$

$z^T X z \geq 0, \forall z$

a convex cone (positive semidefinite cone)

⑨ probability simplex

$$\{x \in \mathbb{R}^d \mid x_i \geq 0, \sum_i x_i = 1\}$$

Operations that preserve Convexity

① Intersection

Intersection of convex sets is convex

> Show that if C_1 and C_2 are convex sets. Then, $C_1 \cap C_2$ is also convex

Proof: [redacted]

[redacted]

[redacted]

[redacted]

② Affine mapping:

ⓐ $f : \mathbb{R}^n \to \mathbb{R}^m$ is affine $\triangleq$ $f(x) = Ax + b$

$A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$

ⓑ image of a convex set under f is convex

$S \subseteq \mathbb{R}^n$ is convex $\Rightarrow$ $f(S) = \{ f(x) \mid x \in S \}$ is convex

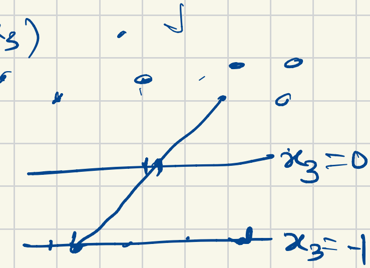
ⓒ inverse image of f

$$\underline{f^{-1}(S)} \triangleq \{ x \mid f(x) \in S \}$$

S is convex $\Rightarrow$ $f^{-1}(S)$ is convex ✓

(x_1, x_2, x_3)

③ Perspective function



$$\underline{P} : \mathbb{R}^{n+1} \longrightarrow \mathbb{R}^n$$

$P(x, t) = \dfrac{x}{t}$, $\operatorname{dom} P = \{ (x, t) \mid t > 0 \}$

images and inverse images of

convex sets under perspective are convex.

(4) Linear- fractional function

$$f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$$

$$f(x) = \frac{Ax+b}{c^T x + d}, \quad \operatorname{dom} f = \{x \mid c^T x + d > 0\}$$

images and inverse images of convex sets under linear-fractional function are convex.

Generalized Inequality

Proper cone: A convex cone $K \subseteq \mathbb{R}^n$ is called a proper cone if

① K is closed

② K is solid (non-empty interior)

③ K is pointed (contains no lines)

Examples:

ⓐ Non-negative orthant

$$K = \mathbb{R}^n_+ = \{ x \in \mathbb{R}^n \mid x_i \geq 0, i=1,\dots,n \}$$

ⓑ positive semidefinite cone

$$K = S^n_+$$

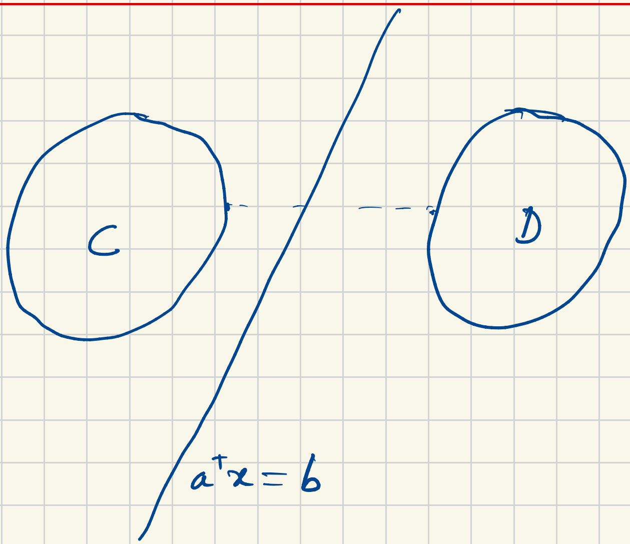
Generalized inequality defined by a proper cone K:

$$x \preceq_K y \iff y - x \in K, \qquad x \prec_K y \iff y - x \in \operatorname{int}(K)$$

Separating Hyperplane Theorem

Thm: If C and D are non-empty disjoint convex sets, then there exists $a \neq 0, b$ s.t.

$$a^T x \leq b \quad \forall x \in C, \quad a^T x \geq b, \quad \forall x \in D$$



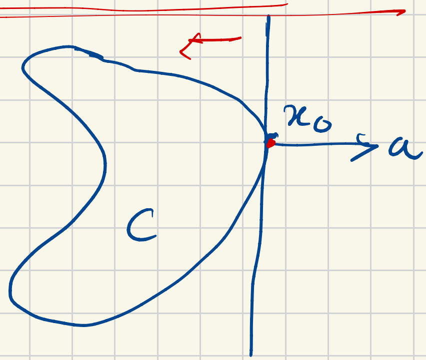
Supporting Hyperplanes

Suppose $C \subseteq \mathbb{R}^n$ and $x_0 \in \mathrm{bd}(C)$

if $a \neq 0$ satisfies:

$$a^T x \leq a^T x_0 \;, \quad \forall x \in C$$

then $\{x \mid a^T x = a^T x_0\}$ is called a supporting hyperplane.



Thm: if C is convex, then there exists a supporting hyperplane at every boundary point of C.

Convex Functions

(a) Examples

(b) Operations that preserve convexity

(c) Conjugate functions

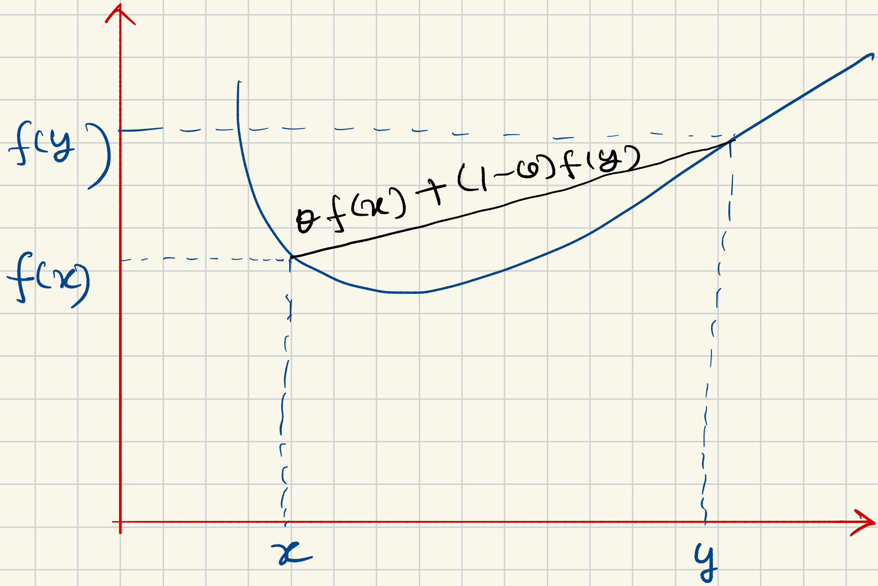
(d) Beyond convexity

Defn: $f: \mathbb{R}^n \to \mathbb{R}$ is convex if

(a) $\text{dom} f$ is a convex set and

(b) $\forall\ x, y \in \text{dom} f$, $0 \leq \theta \leq 1$

$$f(\theta x + (1-\theta) y) \leq \theta f(x) + (1-\theta) f(y)$$



$$\boxed{f(\theta x+(1-\theta)y) \leq \theta f(x)+(1-\theta)f(y)}$$

first order condition:

(a) Suppose f is differentiable then

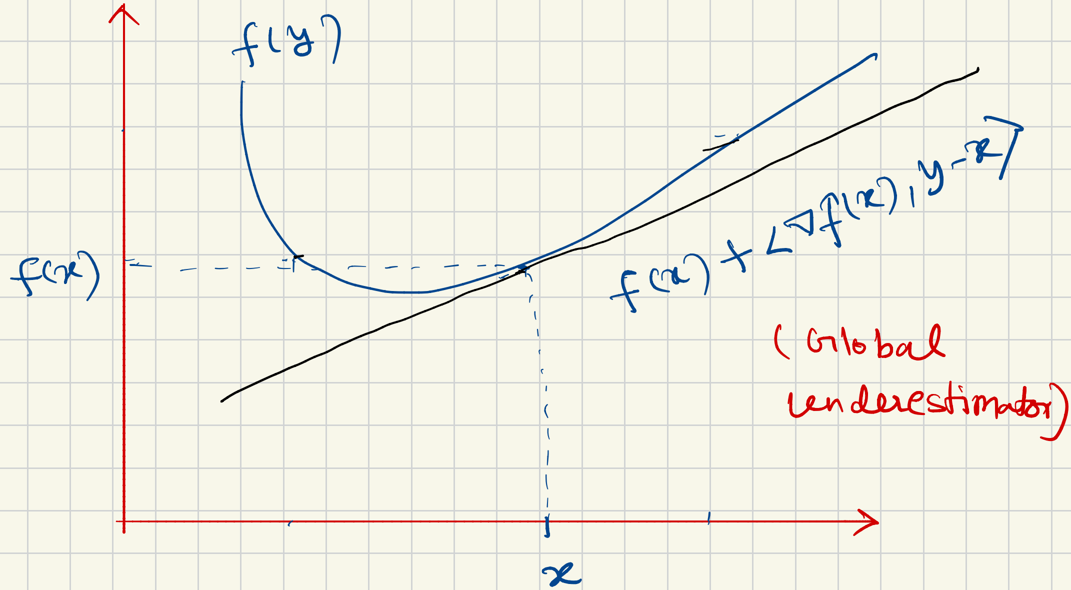
$$\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \ldots, \frac{\partial f(x)}{\partial x_n}\right) \in \mathbb{R}^n$$

exists

(b) f is convex iff

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle$$

$x, y \in \mathrm{dom}\, f$



Second-order condition

(a) Suppose f is twice differentiable

$$\nabla^2 f(x) \in S^n \quad \text{(Hessian)}$$

$$\nabla^2 f(x)_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}, \quad i,j = 1, \dots, n$$

(b) f is convex iff

$$\nabla^2 f(x) \succeq 0, \quad \forall x \in \operatorname{dom} f$$

$(\succ)$

$\rightarrow$ strict

Examples

① Affine: $a^T x + b$, $a \in \mathbb{R}^n$, $b \in \mathbb{R}$

② Exponential: $\exp(ax)$, $a \in \mathbb{R}$

③ $|x|^p$ on $\mathbb{R}$, for $p \geq 1$

④ ReLU: $\max\{0, x\}$

5 Log-sum-exp: (soft max)

$$f(x) = \log\left(e^{x_1} + \cdots + e^{x_n}\right)$$

Proof:

[redacted]

$i \neq j$ [redacted]

$i = j$: [redacted]

$+$ [redacted]

$=$ [redacted]

[redacted]

$(1^T z)^2$

[redacted]

$(1^T z)$ $\sum_{i=1}^{n}$ [redacted] 2

[redacted]

[redacted]

[redacted]

6 Norms:

l_p norm :

$$\|x\|_p = \left(|x_1|^p + \cdots + |x_n|^p\right)^{1/p} \quad \text{for } p \geq 1$$

Defn: Let $f : \mathbb{R}^d \to \mathbb{R}$ be a function that satisfies:

1. $f(x) \geq 0$, and $f(x) = 0$ iff $x = 0$
   (definiteness)
2. $f(\lambda x) = |\lambda| f(x)$ for any $\lambda \in \mathbb{R}$ (positive homogeneity)
3. $f(x+y) \leq f(x) + f(y)$
   (subadditivity)

Thm: Norms are convex

Proof:

Partial minimization :

Thm: Let $\mathcal{Y}$ be a non-empty convex set. Suppose $L(x, y)$ is convex in (x, y), then,

$$f(x) \triangleq \inf_{y \in \mathcal{Y}} L(x, y)$$

is a convex function of x provided $f(x) > -\infty$.

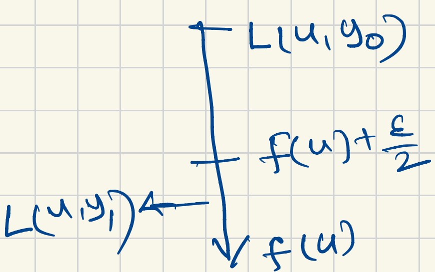
Proof: Let $u, v \in \operatorname{dom} f$. Since

$$f(u) = \inf_y L(u, y), \text{ for each } \varepsilon > 0,$$

$\exists\, y_1 \in \mathcal{Y}$, s.t. $L(u, y_1) \le f(u) + \frac{\varepsilon}{2}$.

Similarly $\exists\, y_2 \in \mathcal{Y}$, s.t.

$$L(v, y_2) \le f(v) + \frac{\varepsilon}{2}.$$



$$f(\lambda u+(1-\lambda)v) = \inf_{y \in \mathcal{Y}} L(\lambda u+(1-\lambda)v, y)$$

$$\leq L(\lambda u+(1-\lambda)v, \lambda y_1+(1-\lambda)y_2)$$

$$\leq \lambda L(u, y_1)+(1-\lambda) L(v, y_2)$$

$$\leq \lambda f(u)+(1-\lambda) f(v) + \varepsilon.$$

$\downarrow$ arbitrarily small

Application: Schur Complements

Let A, B, C be matrices such that $C \succ 0$, and let

$$Z \triangleq \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \succeq 0,$$

then the Schur complement

$$A - BC^{-1}B^T \succeq 0.$$

Proof: Let

$$L(x,y) = [x,y]^T Z [x,y].$$

$L(x,y)$ is convex. Thus,

$$f(x) = \inf_y L(x,y) = x^T (A - BC^{-1}B^T) x$$

convex

$$\therefore \quad \nabla^2 f(x) = A - BC^{-1}B^T \succeq 0.$$