Discussion 2:

- Intro. ML
- Intro PMTK
- review on optimization

ML:

1 - predictive / Supervised learning
2 - descriptive / Unsupervised learning

(1) inpt $X \xrightarrow{\text{map}}$ out P

given $\mathcal{D} \triangleq \{(\underline{x_i}, y_i)\}_{i=1}^{N}$

N → # sample

$\mathcal{D}$ → training set

(y_i) → response

$\underline{x_i}$ → D-dimensional

response y_i < categorical (finite set $\{1, 2, \ldots, C\}$) ; C → # classes
numerical → any real-value number

y_i < categorical → pattern recognition / classification
numerical → regression

(2) input X

Goal: is to find a good pattern input

classification prob:

prob. distribution over possible, given test data x, training set D

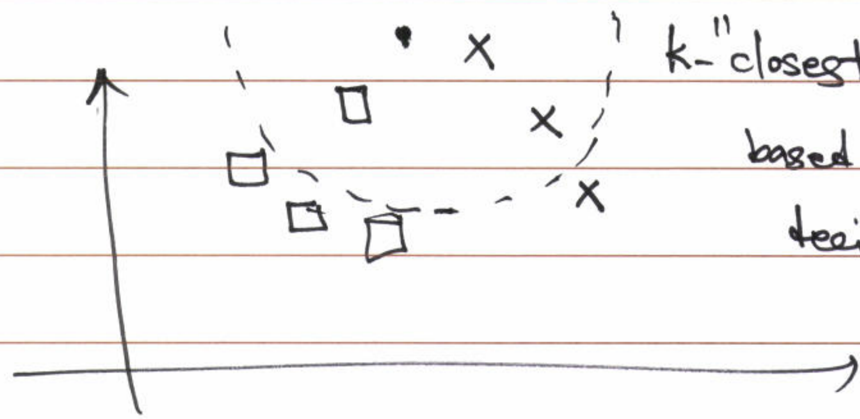
$p(y \mid x, D)$ vector with length c

$\equiv p(y \mid x, D, M)$

decision rule:

$$\underset{d \in C = \{1, 2, \dots, c\}}{\arg\max}\ p(y=d \mid x, D)$$

one of classification method kNN.



k-"closest" one and based majority, it decides

Optimization:

$$\min \ f_0(x)$$
$$\text{s.t.} \quad f_i(x) \leq 0 \qquad i=1,2,\dots,m$$
$$h_i(x) = 0 \qquad i=1,2,\dots,k$$

hard due to:

1. result in local opt.
2. finding a feasible point
3. stopping criteria in general optimization algorithm are arbitrary.
4. opt. algorithm can have poor convergence rate

Convex:

→ results global opt. ✓

→ feasibility of prob. can be unambiguously determined (duality)

→ very precise stopping criteria

convex set:

a function $f: R^n \to R^m$ is affine if it has the form of $f(x) = Ax + b$

extension $f: R^n \to R^{p\times q}$ is affine if

$$F(x) = A_0 + x_1 A_1 + \cdots + x_n A_n$$

Recall subspace $S \subseteq R^n$ if it contains
the plane through any two points & origin

$$x, y \in S \quad \lambda, \mu \in \mathbb{R} \implies \lambda x + \mu y \in S$$

Example: matrix

$$\text{range}(A) = \{ \underset{[a_1, a_2, \cdots, a_q]}{\underset{\downarrow}{A}} w \mid w \in R^q \}$$

$$\text{nullspace}(A) = \{ w \mid Aw = 0 \}$$

Two common representation for affine set

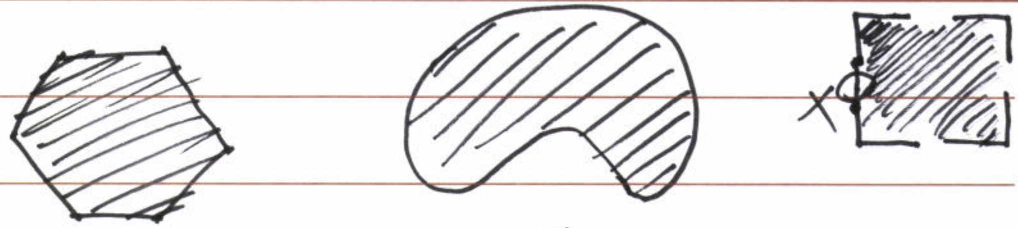
① range of affine $S = \{ Az + b \mid z \in \mathbb{R}^q \}$

② solution a set of linear equations, i.e.

$$S = \{ x \mid b_1^T x = d_1, \cdots, b_n^T x = d_n \}$$

A set is Convex if it contains the line segment joining any of its points, i.e.

$$x, y \in S, \; \underset{\lambda + \mu = 1}{\lambda, \mu \geq 0} \implies \lambda x + \mu y \in S$$



✓ x

Convex Function:

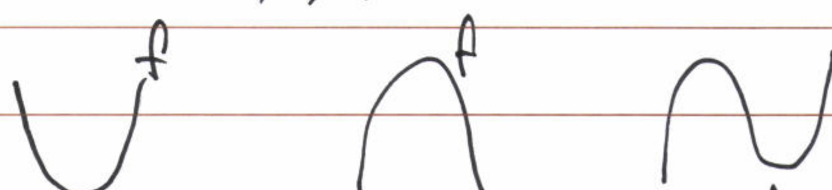
- important convex function
- techinques to verify convexity

Def.: a function $f: \mathbb{R}^n \to \mathbb{R}$ is convex if its domain f is convex AND for all x, y in domain f, $\theta \in [0,1]$

$$f(\theta \underline{x} + (1-\theta)\underline{y}) \le \theta f(\underline{x}) + (1-\theta) f(\underline{y})$$

$\boxed{<}$ (~~strict~~ strict convex)

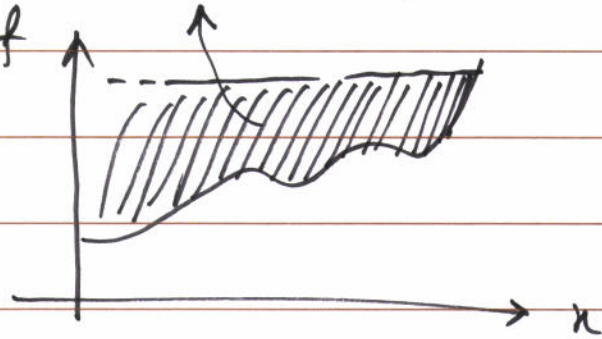
If f is convex, then $-f$ is concave.



conv. concave neither

- x^2 with $\mathrm{dom}\, f = \mathbb{R}$ convex
- $\log x$ with $\mathrm{dom}\, f = \mathbb{R}_{++}$ concave
- $\frac{1}{x}$ with $\mathrm{dom}\, f = \mathbb{R}_{++}$ convex

Def. : epigraph $f = \{(x,t) \mid x \in \mathrm{dom}\, f,\ f(x) \leq t\}$



Def. : the α-sublevel set of a function f is

$$S_\alpha = \{x \in \mathrm{dom}\, f \mid f(x) \leq \alpha\}$$

re-write def. of convex function:

f is convex $\underset{\Leftarrow}{\text{iff}}$ its epigraph is a convex set.

f is convex $\Rightarrow$ its sublevel sets S_α are convex

The convexity of differentiable $f: \mathbb{R}^n \to \mathbb{R}$

* can be characterized by condition on its gradient ∇f & Hessian $\nabla^2 f$

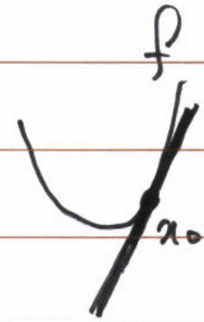
$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} \qquad \nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots \\ & \ddots & \\ & & \ddots \end{bmatrix}$$

– first order of Taylor expansion

$$f(x) \simeq f(x_0) + \nabla f(x_0)^T (x - x_0)$$

Condition for convexity:

f is convex <u>iff</u> $f(x) \geq f(x_0) + \nabla f(x_0)^T (x - x_0)$ $\underline{\forall x, x_0}$



~~Second~~

– Second order approximation

$$f(x) \simeq f(x_0) + \nabla f(x_0)^T (x - x_0) + \frac{1}{2} (x - x_0)^T \nabla^2 f(x_0)(x - x_0)$$

Nece. & Sufficient condition for differentiable f :

f is convex <u>iff</u> $\nabla^2 f \succeq 0 \quad \forall x \in \mathrm{dom}\, f$

Def: A is $A \succeq 0$ if

$x^T A x \geq 0 \quad \forall x$

- x^α is convex R_{++} $\alpha \geq 1$
- $x \log x$ is convex for R_{++}

- log-sum-exp $f(\underline{x}) = \log \sum e^{x_i}$

- quadratic: $f = x^T P x + 2q^T x + r \quad (P = P^T)$

f is convex $\Leftrightarrow P \succeq 0$

Fundemental Rules on Convex function:

- $f: R^n \longrightarrow R$ is convex iff it is convex on all lines.

$\tilde{f}(t) \triangleq f(x_0 + th)$ is convex in $t \in R$ for all $h, x_0 \in R^n$

- non-neg. sum of conv. function is conv.

$\alpha_1, \alpha_2 \geq 0$ & f_1, f_2 convex

$\Downarrow$

$\alpha_1 f_1 + \alpha_2 f_2$ is convex

- non-neg. infinite sum or integral

$p(y) \geq 0$

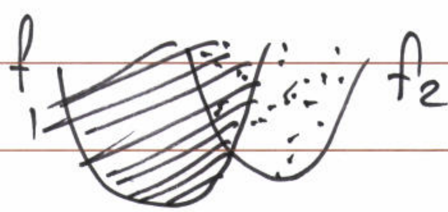
$g(x, y)$ is convex in x $\Longrightarrow$ $\int p(y) g(x, y) \, dy$ is Convex

- point_wise supremum

$f_\alpha(x)$ is convex

$\Downarrow$

- $\sup\limits_{\alpha \in A} f_\alpha(x)$ is convex (corresponds the intersection of epigraphs)



- affine transformation of domain

f is convex $\Rightarrow f(Ax+b)$ is convex.

Example:

$$f(x) = \max_i \{ a_i^T \underline{x} + b_i \}$$