# Lecture 1. Introduction

Mathematical optimization problem

$$\text{minimize} \quad f(x)$$

$$\text{subject to} \quad x \in \Omega$$

where $f: \mathbb{R}^n \to \mathbb{R}$ objective function 目标函数

$x = (x_1, \ldots, x_n)^T \in \mathbb{R}^n$. optimization variables.

$\Omega \subseteq \mathbb{R}^n$ feasible set / constraint set

- $x \in \Omega$ feasible 可行 infeasible o.w.
- $\Omega$ specified by constraint functions $g_1, \ldots g_m$.

$$\min_{x} \quad f(x)$$

$$\text{s.t.} \quad g_i(x) \leq 0, \quad i = 1, \ldots, m.$$

$x^*$: optimal solution. $f(x)$ achieves optimal.

$$x^* = \arg\min f(x).$$

Remark: maximizing $f(x)$ is equivalent to minimizing $-f(x)$

why linear and convex?

In general optimization problems are very difficult to solve.

Knapsack problem. 背包问题

$n$ types of knapsacks.

$i^{th}$ type carry $a_i$ pencils and $b_i$ books. costs. $c_i$

$A$ pencils and $B$ books in total.

Goal: spend least money.

$$\min \quad \sum_i c_i x_i$$

$$\text{s.t.} \quad \sum_i a_i x_i \geq A$$

$$\sum_i b_i x_i \geq B.$$

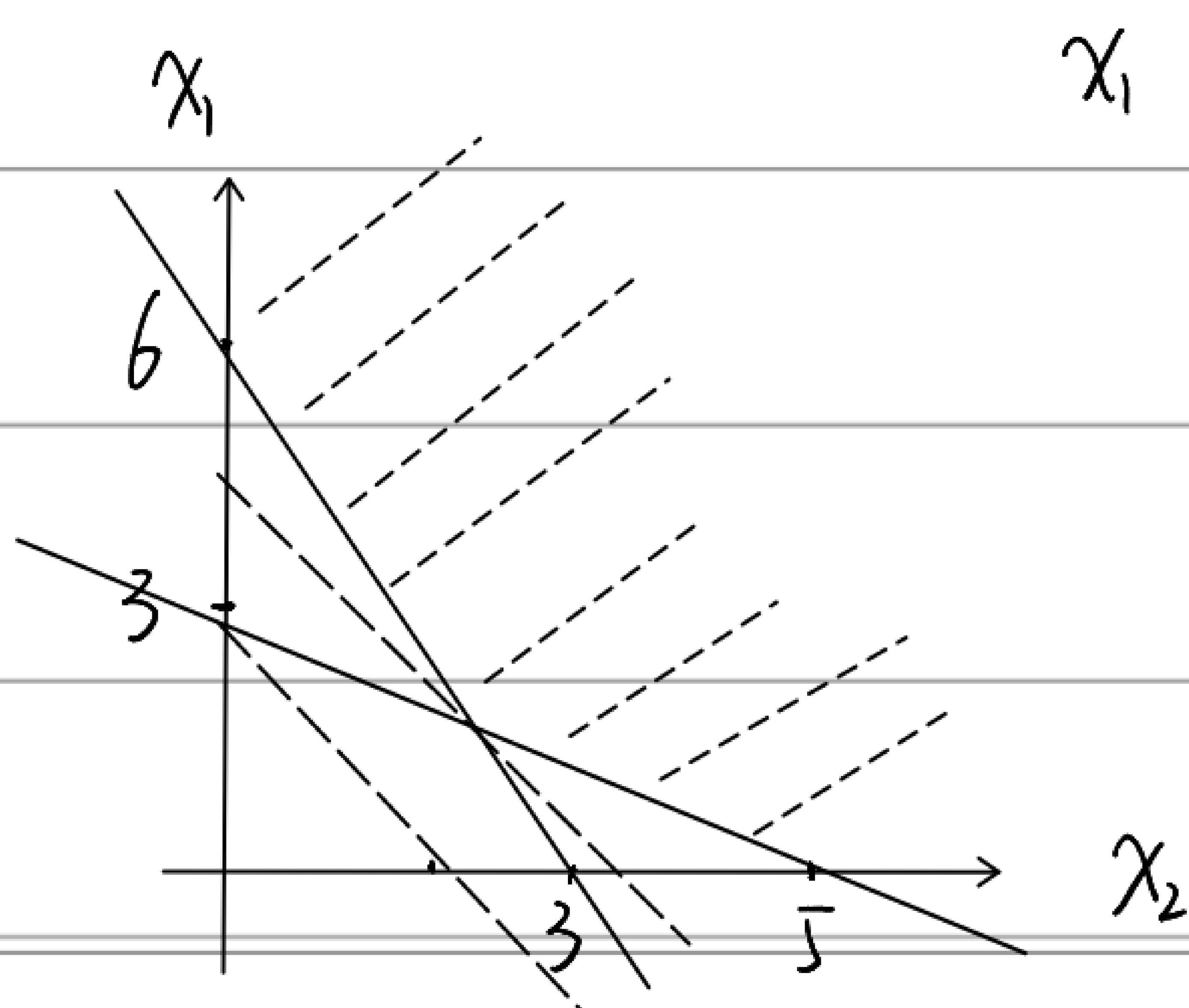
How to solve it?

if only 2 types.

$$\min \quad 10x_1 + 15x_2.$$

$$\text{s.t.} \quad 5x_1 + 3x_2 \geq 15$$

$$x_1 + 2x_2 \geq 6$$




more types?

simplex algorithm.

primal dual

# Data fitting.

| T | V |
|---|---|
| 30 | 1.011 |
| 40 | 1.019 |
| 50 | 1.032 |
| 60 | 1.041 |
| ... | ... |

$$\begin{array}{ccccc} V & & T & & \\ \downarrow & & \downarrow & & \\ y & = & k x & + & b. \end{array}$$

what are the two coefficient $k$ and $b$ ?

## Least squares method.

given $n$ measurements. $(x_1, y_1)$. $\cdots$ $(x_n, y_n)$.

assume $i^{th}$ error denoted by $\varepsilon_i$.

least squares criterion minimize $\Sigma \varepsilon_i = \Sigma (y_i - kx_i - b)^2$

## Geometric explanation : projection

$$\begin{cases} 30k + b = 1.011 \\ 40k + b = 1.019 \\ 50k + b = 1.032 \end{cases} \qquad k \begin{bmatrix} 30 \\ 40 \\ 50 \end{bmatrix} + b \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = ?$$

projection of $\begin{bmatrix} 1.011 \\ 1.019 \\ 1.032 \end{bmatrix}$ onto the subspace spanned by $\begin{bmatrix} 30 \\ 40 \\ 50 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

minimize $\left\| \begin{bmatrix} 30 & 1 \\ 40 & 1 \\ 50 & 1 \end{bmatrix} \begin{bmatrix} k \\ b \end{bmatrix} - \begin{bmatrix} 1.011 \\ 1.019 \\ 1.032 \end{bmatrix} \right\|_2^2$

Norm. inner product. distance to hyperplane

not necessary $\mathbb{R}^n$ Fourier

IP: an inner product $\langle \cdot, \cdot \rangle$ is a function $S \times S \to R$. s.t.

1. nonnegative $\langle x, x \rangle \geq 0$. $= 0$ iff $x = 0$

2. symmetric $\langle x, y \rangle = \langle y, x \rangle$

3. linearity $\langle sx, y \rangle = s \langle x, y \rangle$ (homogeneity)

$\langle x+y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$ (additivity)

if $\langle x. y \rangle = 0$ then x and y are called orthogonal. 正交.

Euclidean inner product space: $\langle x. y \rangle = x^T y = \sum x_i y_i$

Norm: a norm $\|\cdot\|$ is a function $\mathbb{R}^n \to R$. s.t.

1. nonnegative $\|x\| \geq 0$ $= 0$ iff $x = 0$.

2. positive homogeneity $\|ax\| = |a| \, \|x\|$

3. triangle inequality $\|x+y\| \leq \|x\| + \|y\|$.

$L^P$-norm. or p-norm for real $p \geq 1$.

$\|x\|_p = (|x_1|^p + |x_2|^p + \cdots + |x_n|^p)^{1/p}$.

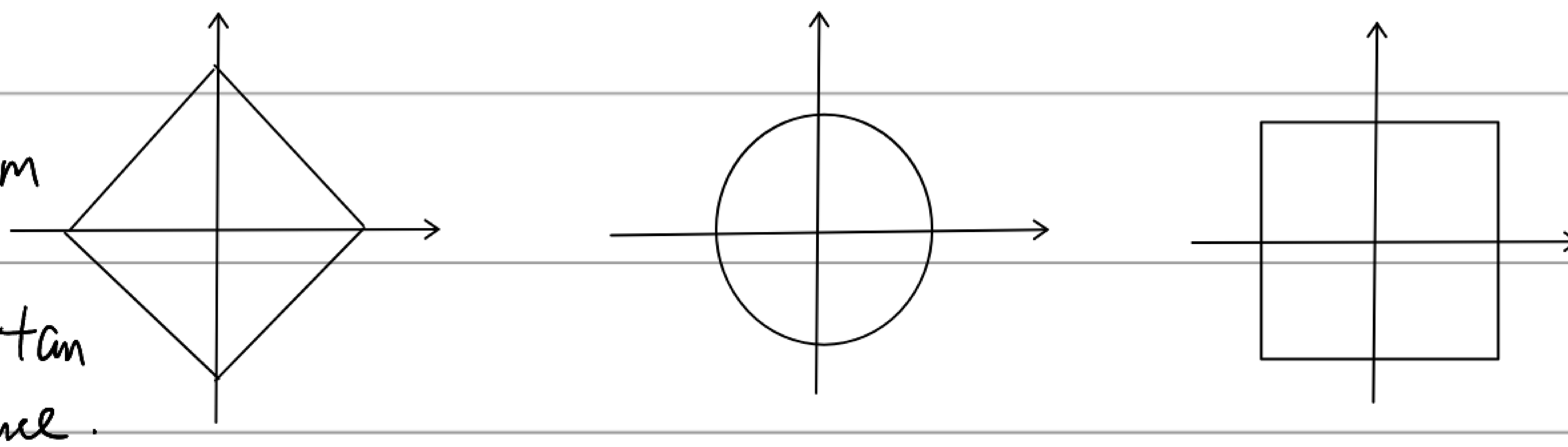
in particular. 1-norm : $\|x\|_1 = \sum_i |x_i|$

(Euclidean norm) 2 : $\underset{\text{default}}{\|x\|} \triangleq \|x\|_2 = \sqrt{\langle x, x \rangle} = \sqrt{\sum_i x_i^2}$

$\infty$ - norm : $\|x\|_\infty = \max\{|x_1|, |x_2|, \cdots, |x_n|\}$.

1 norm

Manhattan distance.



Cauchy – Schwarz inequality

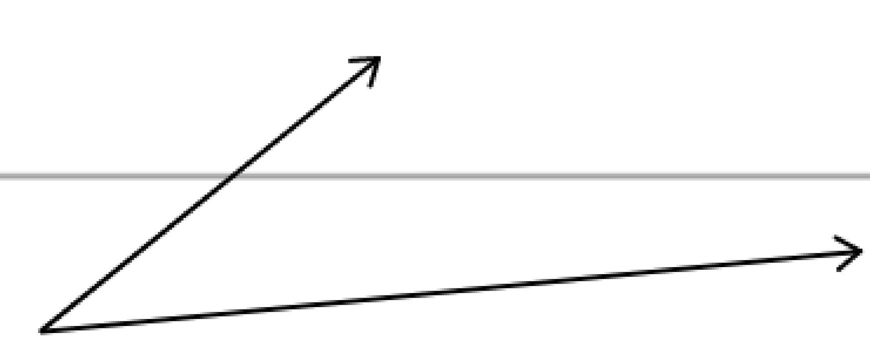
$$|\langle u, v\rangle|^2 \le \langle u, u\rangle \cdot \langle v, v\rangle \quad \text{or.}$$

$$|\langle u, v\rangle| \le \|u\| \, \|v\|.$$

n-dimensional Euclidean space. ($\mathbb{R}^n$)

$$\left(\sum_i u_i v_i\right)^2 \le \left(\sum_i u_i^2\right)\left(\sum_i v_i^2\right).$$

Geometric explanation: projection.

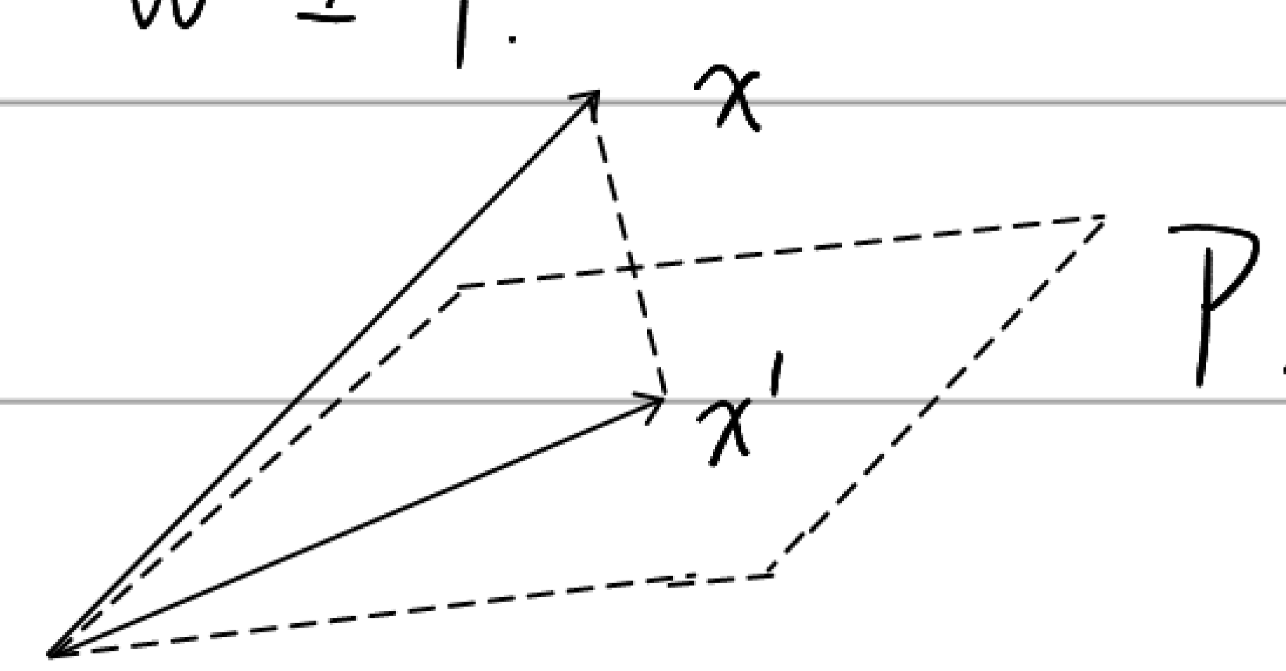


Distance to hyperplane.

hyperplane $P$: $w^T x + b = 0$ $\quad w \perp P$.

orthogonal projection.

$(x - x') \perp P \quad w^T x' + b = 0.$



$x - x' = r \cdot w$ for some $r \in \mathbb{R}$.

$w^T (x - r \cdot w) + b = 0 \Rightarrow r = \frac{w^T x + b}{w^T w}$ $\leftarrow \|w\|^2$

distance from $x$ to $P$ is.

$\min_{y \in P} \| x - y \| = \| x - x' \| = \| r w \| = \frac{|w^T x + b|}{\|w\|}$.

Linear least squares regression.

given $m$ measurements $(x_1, y_1) \dots (x_m, y_m)$.

assume that $y = w^T x + b$.

The least squares regression. is to compute the following OPT.

$\min_{w \in \mathbb{R}^n, b \in \mathbb{R}} \sum_i (Xw + b \cdot \mathbf{1} - y)^2$

$= \min_{w \in \mathbb{R}^n, b \in \mathbb{R}} \| Xw + b \cdot \mathbf{1} - \mathbf{y} \|$

where $X = (x_1, \dots x_n)^T \in \mathbb{R}^{m \times n}$. $y = (y_1, \dots y_m)^T \in \mathbb{R}^m$

$e \perp$ hyperspace spanned by column vector of $X$.

$y - e \in \cancel{}$ $\qquad X^T e = 0$

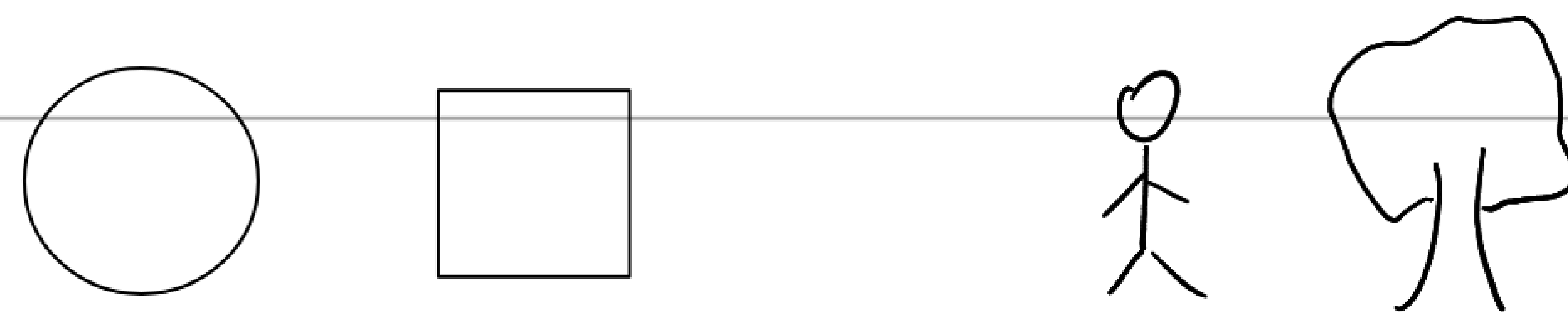
suppose $y - e = X \hat{w}$ thus we have.

$$X^T (y - X \hat{w}) = 0.$$

$$X^T X \hat{w} = X^T \mathbf{y}$$

Classification and support vector machine.

classify

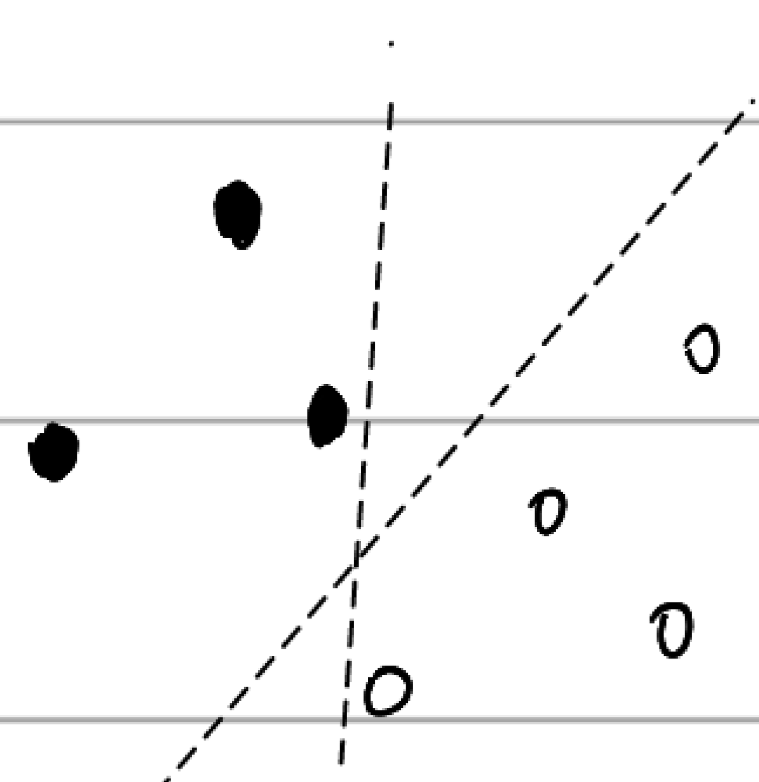


given $m$ data points. $(x_1, y_1) \cdots (x_m, y_m)$.

classfier is a function s.t.

$$\begin{cases} f(x_i) > 0 & \text{iff } y_i = +1 \\ f(x_i) < 0 & \text{iff } y_i = -1 \end{cases} \iff y_i f(x_i) > 0.$$

linear classifier : $f(x) = w^T x + b$.



which one is better?

maximize the minimum distance to the hyper plane. against noise

Support vector machine : linear classifer with max margin

$$\max_{w,b} \min_{1 \le i \le m} \frac{|w^T x_i + b|}{\|w\|}.$$

$$\text{s.t.} \quad y_i (w^T x_i + b) > 0.$$

since $y_i = \text{sgn}(w^T x_i + b)$. $|w^T x_i + b| = y_i (w^T x_i + b)$.

$\forall \alpha > 0$. $\tilde{w} = \alpha w$. $\tilde{b} = \alpha b$ also feasible.

choosing $\alpha$ properly. s.t. $\min_{1 \le i \le m} y_i(\tilde{w}^T x_i + \tilde{b}) = 1$.

$$\max \frac{1}{\|w\|}$$

$$\text{s.t. } y_i(\tilde{w}^T x_i + \tilde{b}) \geq 1.$$

which is equivalent to.

$$\min \frac{1}{2}\|\tilde{w}\|^2$$

$$\text{s.t. } y_i(\tilde{w}^T x_i + \tilde{b}) \geq 1.$$

Global optima and local optima.

$\min_{x \in X} f(x)$. let $x^* \triangleq \arg\min_{x \in X} f(x)$.

$x^*$ is a global minimum iff $f(x^*) \le f(x)$

global maximum.

global optima may not exist.

- $f(x) = x$. $X = \mathbb{R}$. $\inf f(x) = -\infty$.
- $f(x) = \frac{1}{x}$ $X = \mathbb{R}_{>0}$ $\inf f(x) = 0$.

When will global optima exist?

Continuous functions on compact sets have global optima.