Last time:

~~ont~~

learning from experts.

PAC Model. → binary class.

Adaboost → classification.

PAC:

X : instance space.

(feature vec. of imgs)

Y : label space

$\{+1, -1\}$

D : $X \times Y$

① all data $\overset{iid}{\sim} D$.

D_{train} & D_{test}.

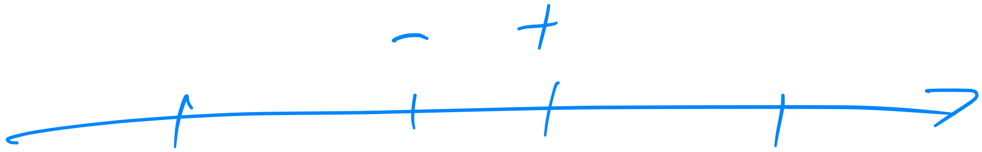
~~can~~ construct S_{train}.

such that. $\exists\, h$.

$$err(h; S_{train}) = 0$$

but $err(h; S_{test}) = 1$

②



$$n = \frac{1}{\epsilon^2} \log \frac{1}{\delta}$$

w.p. $1-\delta$. $|\hat{\alpha} - \alpha| \leq \epsilon$

Empirical risk minimization

(ERM)

- $S = \{(x_i, y_i)\}_{i=1}^{n}$

- $H = \{h : x \to y\}$.

$\ell(h; x, y)$ - loss func.

$$\ell(h; S) \triangleq \frac{1}{n} \sum_{i=1}^{n} \ell(h; x_i, y_i)$$

$$\underline{\min_{h \in H} \ell(h; S)} \quad \text{ERM}.$$

Linear Regression

$\cancel{h : x}$

(x, y)

$$y \approx w \cdot x$$

$$h_w : x \to w \cdot x$$

$$L(h; x, y) = (\underbrace{w \cdot x}_{= h(x)} - y)^2$$

Binary Classification:

(x, y)

$$y \approx \operatorname{sign}(w \cdot x)$$

$$h_w : x \to \operatorname{sign}(w \cdot x)$$

0/1 - loss:

$$\ell(h; x, y) = \mathbb{1}\{h(x) \neq y\}$$

$h(x) = \text{sign}(w \cdot x)$

$$\mathbb{1}\{E\} = \begin{cases} 1, & E \text{ happens} \\ 0 & \text{else.} \end{cases}$$

hinge loss

$$\ell(h; x, y) = \max\{0, 1 - y \cdot h(x)\}.$$

ReLU:

$$y \approx \max\{0, w \cdot x\}$$

$$h : x \to \max\{0, w \cdot x\}.$$

$$\ell(h; x, y) = (h(x) - y)^2$$

$$\min_{h \in H} L(h; S)$$

comp. opt.

Q1: How to find global min. $\hat{h}$?

Q2: When ~~is~~ $\hat{h}$ performs well?

stat.

w.h.p.

$n = \square$

$$p\ \mathrm{err}(\hat{h}) \le \epsilon$$

$$\Pr_{(X,Y)\sim D}(\hat{h}(X) \neq Y) \le \epsilon$$

~~min~~

$$\min_{\boxed{w \in C}} F(w) \rightarrow L(h;S), \quad h \leftarrow w$$

$$C = \{w : \|w\| \le 1\}$$

$$= \{w : \|\ \ \|_1 \le 1\}$$

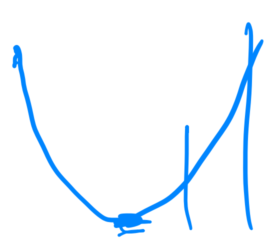
$$= \mathbb{R}^d \ldots$$

$$\nabla F(w) = 0 \rightarrow \hat{w}$$

$$F(w) = e^w - w^2$$

① $\nabla F(w) = \boxed{e^w - 2w = 0}$

②



Gradient Descent.

w^0

$t = 1, \dots T$

$$\underbrace{w^{t+1}}_{\mathbb{R}^d} \leftarrow \underbrace{w^t}_{} - \eta \cdot \underbrace{\nabla F(w^t)}_{\mathbb{R}^d}$$

$\mathbb{R}^d$

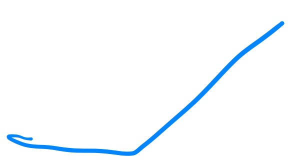
$w \in \mathbb{R}$ ✓ $\quad \boxed{w^{t+1} = w^t}$

$\Longleftrightarrow \nabla F(w^t) = 0$

$w \in \mathbb{R}^d \quad \cancel{\nabla F(w)}$

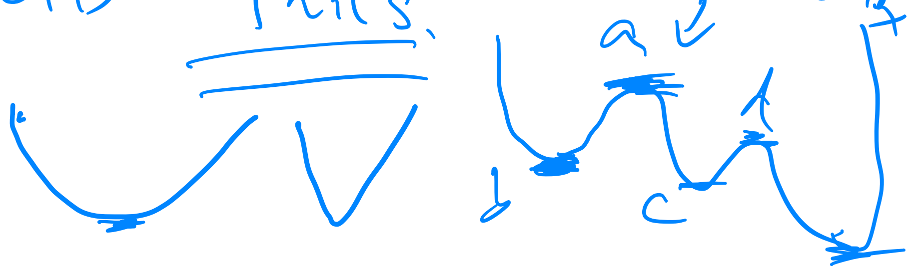
$$\nabla F(w) \triangleq \left(\frac{\partial F}{\partial w_1}, \frac{\partial F}{\partial w_2}, \dots \frac{\partial F}{\partial w_d}\right)$$

① When GD fails.



~~How~~

non-convex



EX1:

$$w = (w_1, w_2, w_3)$$

$$F(w) = w_1^2 + w_2 + w_3$$

$$\boxed{\nabla F(w) = (2w_1, 1, 1)}$$

EX2:

$$F(w) = \underline{w_1^2} \cdot \underline{w_2} + \underline{w_3}$$

$$\frac{\partial F}{\partial w_1} = 2w_1 \cdot w_2 + 0 = 2w_1 w_2$$

$$\frac{\partial F}{\partial w_2} = w_1^2 + 0 = w_1^2$$

$$= 1$$

$$\boxed{\nabla F(w) = (2w_1 w_2, w_1^2, 1)}$$

$$\nabla F(w) = (2w_1, 1, 1)$$

$$\frac{\partial}{\partial w_1} = 2, \quad \frac{\partial}{\partial w_2} = 0, \quad \frac{\partial}{\partial w_3} = 0$$

$$\begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$= \begin{pmatrix} 2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \checkmark$$

$\downarrow$

F convex

$$\nabla F(w) = (2w_1w_2,\ \underline{w_1}^2,\ 1)$$

$$\begin{pmatrix} 2w_2 & 2w_1 & 0 \\ 2w_1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$\cancel{w =}$

$$w = (0,\ -1,\ 0)$$

$$\begin{pmatrix} -2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \not\succeq 0$$

$\Rightarrow F$ non-convex

$n^{\sqrt{1+M}} \cdot d$

$\cancel{2^n \cdot d}$

poly(n, d).

$n d$

$n^{100} \cdot d$

non-convex

2010

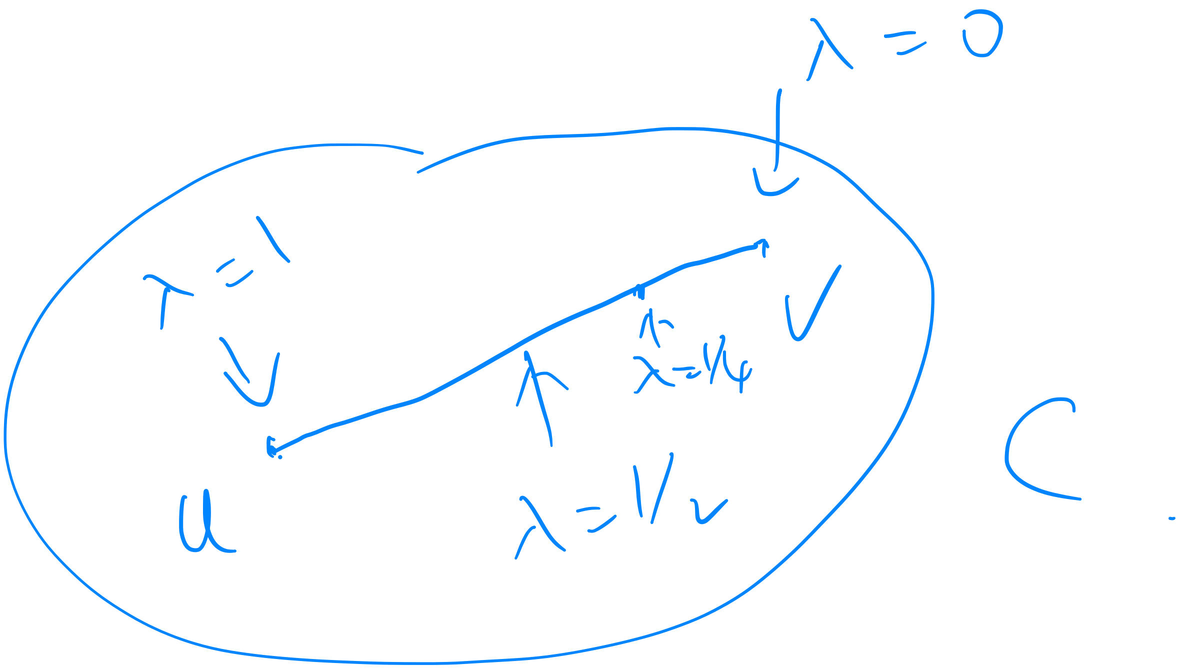
10 years

convex

1950's.

70 years

convex set



C is convex iff

$\forall\, u, v \in C$.

$\underbrace{\lambda u + (1-\lambda) v \in C}$.

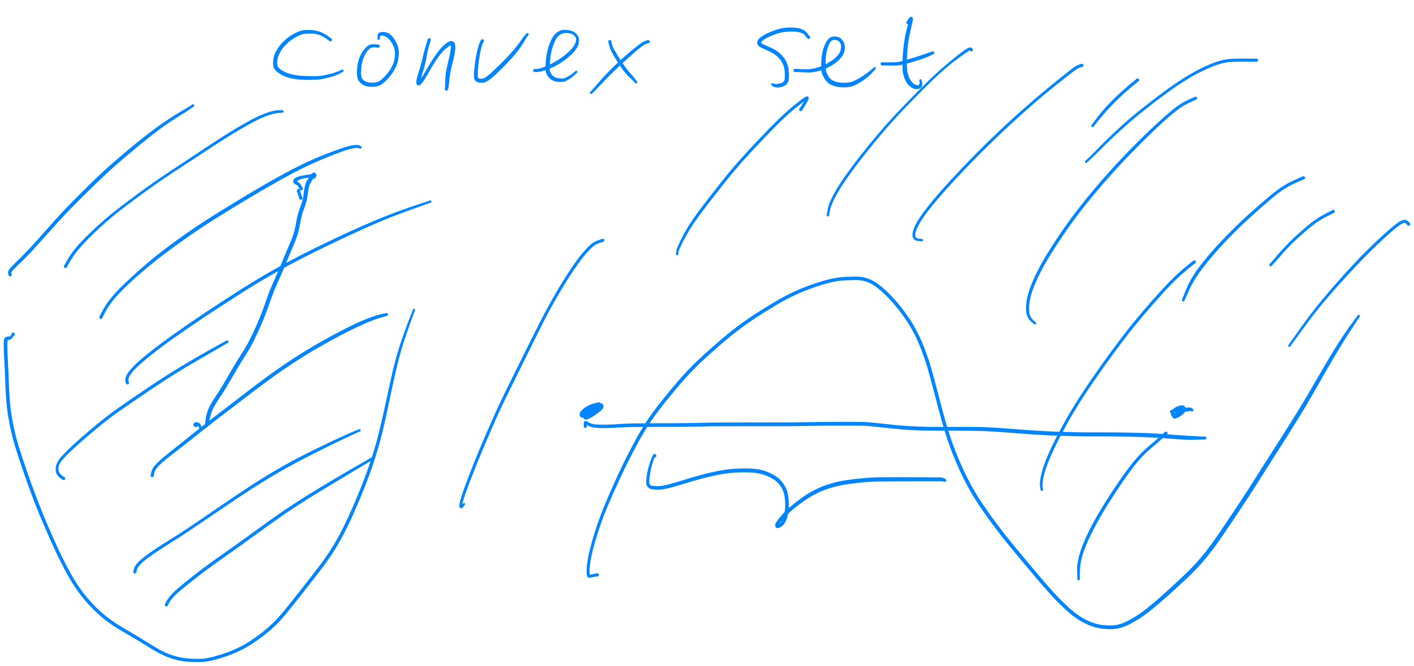
hold for all $\lambda \in [0,1]$

$\lambda = 0$. v

$\lambda = 1$. u

F is convex

$\Leftrightarrow$ epi-graph of F is convex set

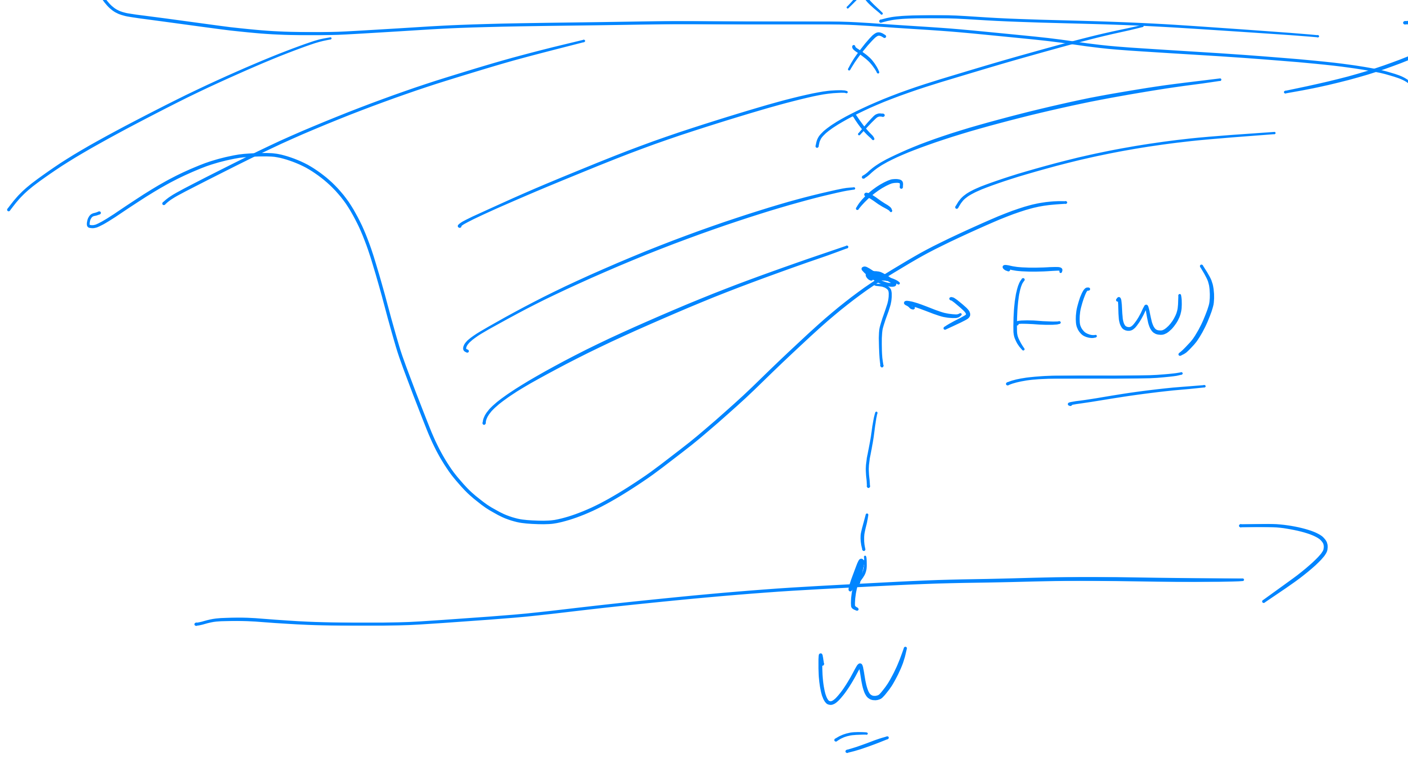


convex

7:40

Given $F(w)$

$$\boxed{\mathrm{epi}(F) = \{(w, y) : w \in C,\ y \geq F(w)\}}$$



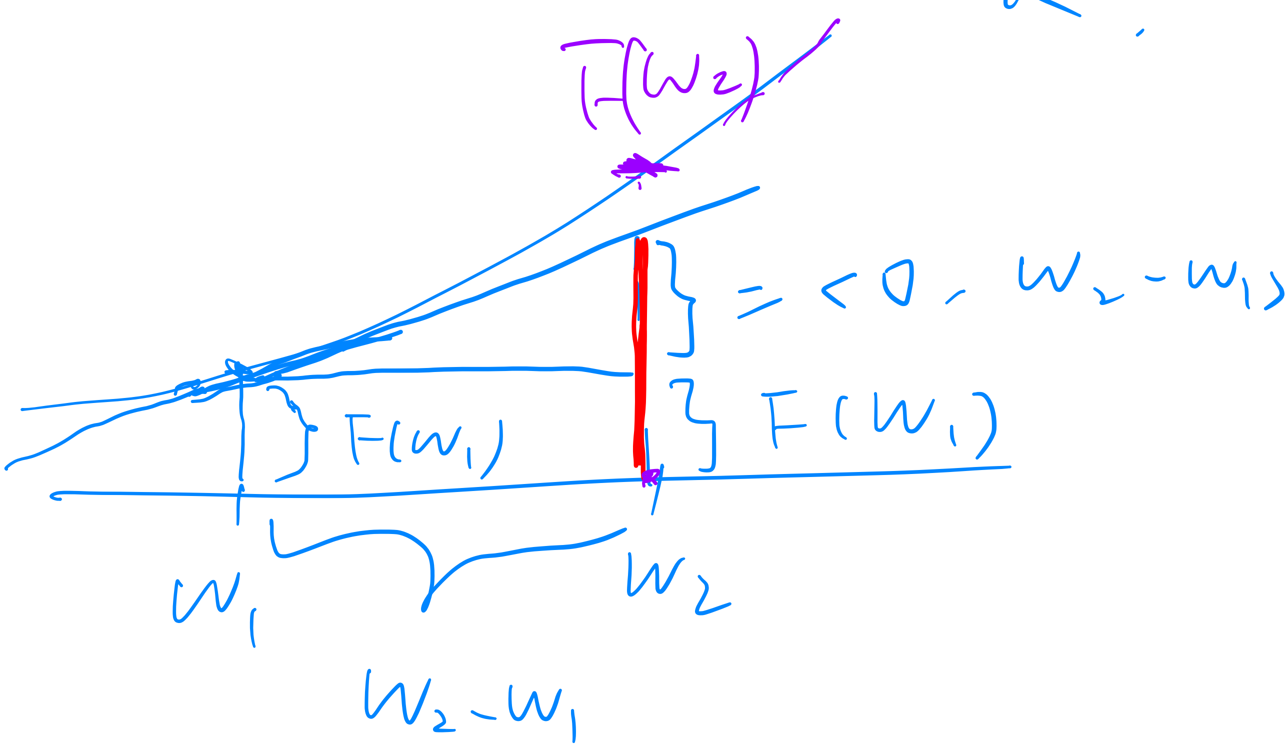
$(w, \underline{y}) \rightarrow$

$F \iff \mathrm{epi}(F)$ is convex set.

$\nabla F(w)$ exists

$\forall\, w_1, w_2 \in C.$

$$\underbrace{F(w_2)}_{\mathbb{R}} \geq \underbrace{F(w_1)}_{\mathbb{R}} + \underbrace{\boxed{\langle \nabla F(w_1), w_2 - w_1 \rangle}}_{\mathbb{R}}.$$



Hessian matrix

$$\nabla F(w) \in \mathbb{R}^d.$$

$$\cancel{\nabla(\nabla F(w))} \quad \nabla^2 F(w)$$

$$\nabla F(w) = (\underset{\substack{\uparrow\\ \mathbb{R}}}{g_1(w)} \quad \cdots \quad \underset{\substack{\uparrow\\ \mathbb{R}}}{g_d(w)})$$

$\underline{\nabla g_1(w)} \in \mathbb{R}^d$ (column)

$$\nabla g_2(w) \quad \cdots \quad \nabla g_d(w)$$

$$\underset{d\times d}{\nabla^2 F(w)} = (\underset{\substack{\downarrow\\ \mathbb{R}^d}}{\nabla g_1(w)} \quad \cdots \quad \underset{\substack{\downarrow\\ \mathbb{R}^d}}{\nabla g_d(w)})$$

F is convex

iff $\nabla^2 F(w) \succeq 0$

$\forall w \in C.$

positive semi-definite

$v^T M v \geq 0$

$\forall v \in \mathbb{R}^d.$

$$w^{t+1} = w^t - \eta \cdot \nabla F(w^t)$$

F is convex

$$F(w) = \frac{1}{2} \cdot w^2 \qquad w \in \mathbb{R}.$$

$\boxed{\eta = 100}$ $\qquad \nabla = w$

$$w^{t+1} = w^t - 100 \cdot w^t$$

$$= -99\, w^t$$

$w^0 = 1, \to -99 \to 99^2 \to -99^3$

diverges.

$0 < \eta \le \square\, \lambda\, f(\theta)^{H}$

$\eta = 1/2$

$w^{t+1} = \frac{1}{2} w^{t}$

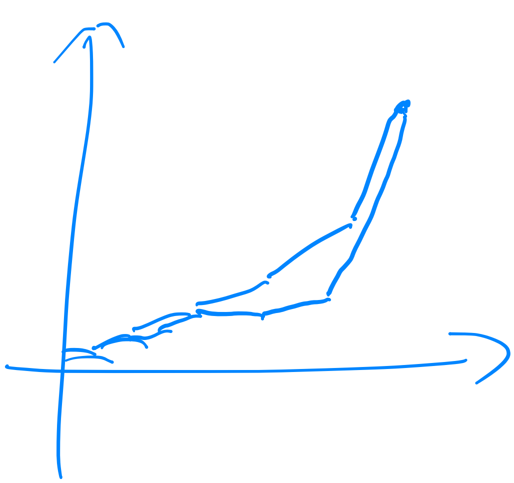
$\Rightarrow 1 \to 1/2 \to 1/4 \to 1/8$

$\dots \to 0$

$1 \to 0.5 \to 0.25 \to 0.125$

$\to 0.0625$

$\to \dots$

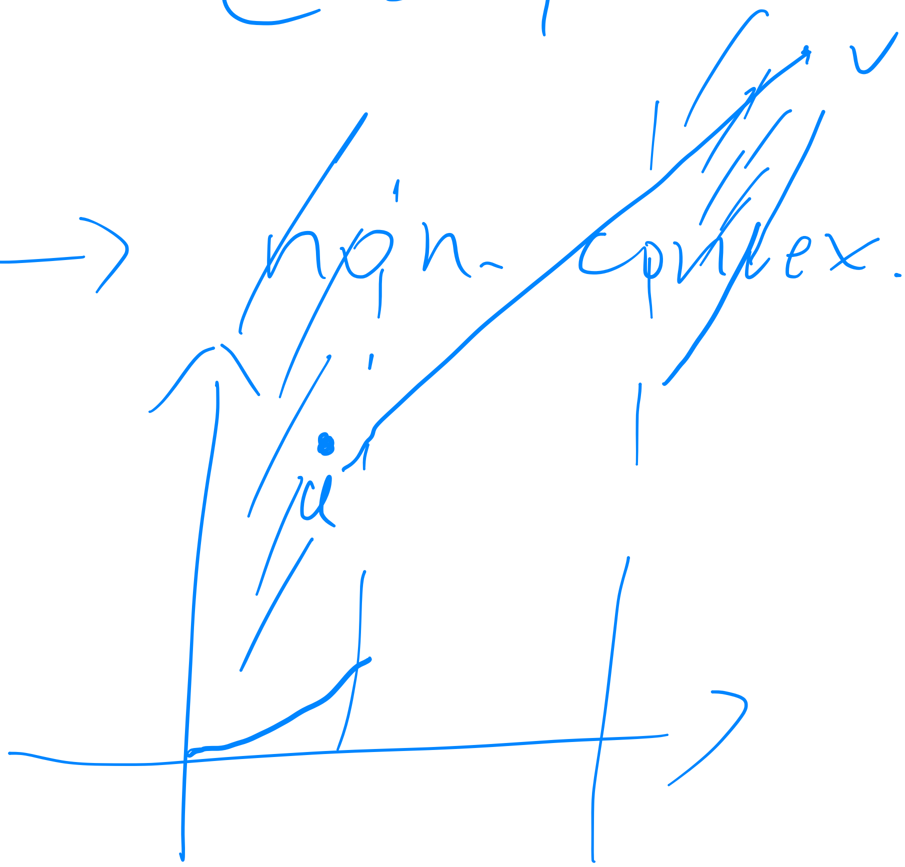


$\frac{1}{2} w^2$, w, 1

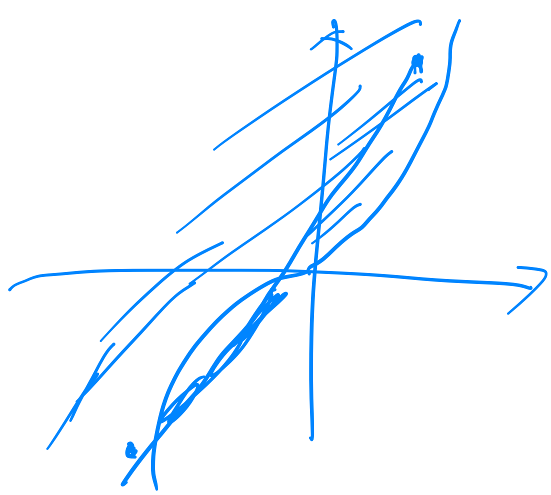
$$F(w) = \frac{1}{2} w^2$$

$$w \in C = \{(0,1) \cup (10,11)\}$$

$\rightarrow$ non-convex.



$$F(w) = w^3, \quad w \geq 0$$



$$F(w) = w^3, \quad w \geq 0$$

$$w^0 = 1 \qquad 3 \cdot w^2 \qquad 6w \geq 0$$

$$\nabla F = 3 \cdot w^2 . \qquad \eta = 1/4$$

$$w^{t+1} = w^t - \eta \cdot 3 \cdot (w^t)^2$$

$$= w^t - \frac{3}{4} \cdot (w^t)^2$$

$$1 \to 1/4 \to \frac{13}{64} \to 0.172$$

$$1 \to 0.25 \to 0.203 \to$$

$$\frac{1}{4} - \frac{3}{4} \cdot \frac{1}{16} = \frac{16}{64} -$$

$$\frac{13}{64} - \frac{3}{4} \cdot \left(\frac{13}{64}\right)^2 \approx 0.172$$

$F(w) = |w|$

non-smooth

$\eta \approx 0$.

w^0

$\eta_t \to 0$

as $t \to \infty$

$w^0 \to w^1 \to w^0 \to w^1$

$\to w^0 \ldots$

HW