DSA5202 $\boxed{\text{Lec1}}$.

1. Recap of DSA5105

① Linear Basis Model
② SVM
③ Kernel Trick + Regularization
④ Neural Network
⑤ Reinforcement Learning
⑥ PCA, Auto-encoder
⑦ K-means, GMM

hyperparameters

⑧ Cross-Validation to determine learning rate
⑨ package like scikit
⑩ Testing set to validate model performance

2. This course:

① Optimization Theory → training method { gradient descent / Newton method

② Trainability Issue for Deep Learning

(confidence)
③ quantify the uncertainty of model

↓

$\boxed{\text{risk of model}}$

3. Today's lecture $\longrightarrow$ $\boxed{\text{optimization}}$

① Framework:

$\boxed{h_w : \text{our model}}$

$\boxed{\text{ideally}}$:

$\hookrightarrow w^* = \arg\min_w f(w)$

$f(w) := \mathbb{E}_{(X,Y)\sim D}[\ell(Y, h_w(X))]$ $\Uparrow$ $\boxed{\text{population loss}}$

$\ell(\cdot,\cdot) \longrightarrow$ distance measure $\uparrow$ LLN

$\boxed{\text{practically}}$:

$\searrow \hat{w} = \arg\min_w \hat{f}(w)$

$\hat{f}(w) := \frac{1}{N}\sum_{i=1}^{N} \ell(y_i, h_w(x_i))$

$\Downarrow$ $\boxed{\text{empirical loss}}$

$\boxed{(x_i, y_i) \xrightarrow{\text{sample}} D}$

(finite set of training point)

$\boxed{\text{idea}}$

To achieve this, one necessary condition is : $\boxed{\nabla_w \hat{f}(\hat{w}) = 0}$

(it is difficult to find the exact minimizer $\hat{w}$)

② Definition: (our interested w)

$\rightarrow$ w is stationary point $\iff \nabla f(w) = 0$

$\rightarrow$ w is ε-stationary point $\iff \|\nabla f(w)\| \le \varepsilon$

③ Convexity & optimality $\longrightarrow$ guarantee stationary point $\longleftrightarrow$ minima

Definition:

$\rightarrow$ D is a convex set $\iff \forall x, y \in D$, for all $t \in [0,1]$

$tx + (1-t)y \in D$

$\rightarrow$ f is a convex function defined on a convex set D

$\iff \forall x, y \in D$, $t \in [0,1]$, $f(tx + (1-t)y) \le t f(x) + (1-t) f(y)$

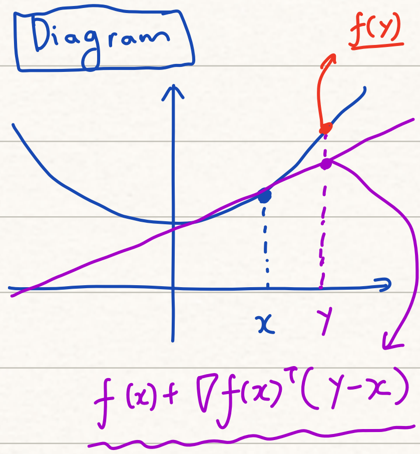
$\rightarrow$ if $f \in C^1$, then we have

$$f \text{ is convex} \iff f(y) \geq f(x) + \nabla f(x)^T (y - x) \quad \forall x, y \in D$$

Pf Sketch : " $\Rightarrow$ " f is convex $+$ $f \in C^1$ $\Rightarrow$ $\partial f(x) = \{\nabla f(x)\}$ $\forall x$

$\Rightarrow \forall x, y \in D.\ f(y) \geq f(x) + \nabla f(x)^T (y - x)$

Diagram



" $\Leftarrow$ " Pf by construction :

f is convex $\iff f(\bar{x}) \leq t f(x) + (1-t) f(y)$

$\boxed{\bar{x} = t x + (1-t) y}$

① $f(x) \geq f(\bar{x}) + \nabla f(\bar{x})^T (x - \bar{x})$

② $f(y) \geq f(\bar{x}) + \nabla f(\bar{x})^T (y - \bar{x})$

$t \times ① + (1-t) \times ② \Rightarrow t f(x) + (1-t) f(y)$

$\geq f(\bar{x}) + 0$

$\Rightarrow$ f is convex !

Direct Pf " $\Rightarrow$ " f is convex

$\Rightarrow f(tx + (1-t)y) \leq t f(x) + (1-t) f(y)$

$\Rightarrow f(x + (1-t)(y-x)) \leq t f(x) + (1-t) f(y)$

$\Rightarrow f(x) + (1-t) \nabla f(x)^T (y - x) + O((1-t)^2) \leq t f(x) + (1-t) f(y)$

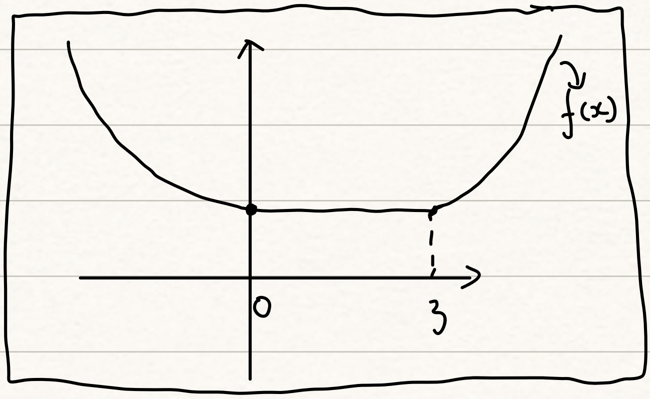
$\Rightarrow f(x) + \nabla f(x)^T (y-x) + O(1-t) \leq f(y)$ $\boxed{t \to 1^-}$

$\Rightarrow f(x) + \nabla f(x)^T (y-x) \leq f(y)$

and $\boxed{f \text{ is convex}}$

[Proposition] if $\nabla f(x) = 0$. then x is ONE minimizer

(might be multiple minimizers)



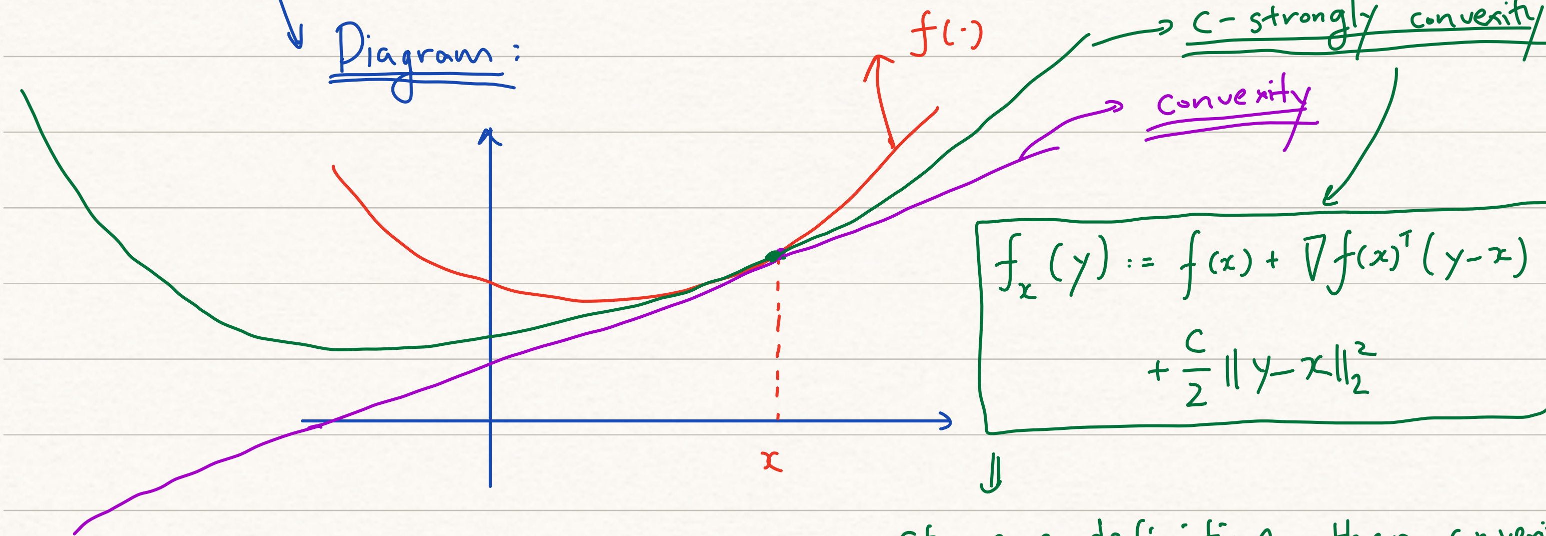
→ **Definition**: $f \in C^1$ is a **C-strongly convex function**

$$\Leftrightarrow f(y) \geq f(x) + \nabla f(x)^T (y-x) + \frac{C}{2}\|y-x\|_2^2$$

$$\boxed{\forall x, y \in D}$$

Recap: $f \in C^1$ is convex $\Leftrightarrow f(y) \geq f(x) + \nabla f(x)^T (y-x)$

Diagram:



$$\boxed{\begin{aligned} f_x(y) := f(x) &+ \nabla f(x)^T (y-x) \\ &+ \frac{C}{2}\|y-x\|_2^2 \end{aligned}}$$

$\Downarrow$

Stronger definition than convexity

Remark: [C-strongly convexity] requires the function $f(\cdot)$ cannot contain a line !!!

$$f(y) \geq f(x) + \nabla f(x)^T (y-x) + \frac{C}{2}\|y-x\|_2^2$$

→ $f \in C^1$ is C-strongly convex

$\Leftrightarrow \forall x, y \in D,\ \langle \nabla f(x) - \nabla f(y), x-y \rangle \geq C\|y-x\|_2^2$

Pf: "$\Rightarrow$" $\forall x, y \in D$, we have

$$\begin{cases} f(y) \geq f(x) + \nabla f(x)^T (y-x) + \frac{C}{2}\|y-x\|_2^2 \\ f(x) \geq f(y) + \nabla f(y)^T (x-y) + \frac{C}{2}\|y-x\|_2^2 \end{cases}$$

$$\Rightarrow \langle \nabla f(x) - \nabla f(y), x-y \rangle \geq C\|y-x\|_2^2$$

"$\Leftarrow$" $g(t) := f(x+t(y-x)) - \frac{C}{2}\|x+t(y-x)\|_2^2$

$g'(t) = \langle \nabla f(x+t(y-x)), y-x \rangle - C\langle x+t(y-x), y-x \rangle$

$$g'(0) = \langle \nabla f(x), y-x \rangle - c \langle x, y-x \rangle$$

$$g'(t) - g'(0) = \langle \nabla f(x+t(y-x)) - \nabla f(x), y-x \rangle$$
$$- ct \langle y-x, y-x \rangle$$

$$= \frac{1}{t} \langle \nabla f(x+t(y-x)) - \nabla f(x), t(y-x) \rangle$$
$$- ct \langle y-x, y-x \rangle$$
$$\geq \frac{1}{t} \cdot c \| t(y-x) \|_2^2 - ct \| y-x \|_2^2$$

$$= 0 \qquad \underline{\underline{\forall t \in [0,1]}}$$

$$g(1) = f(y) - \frac{c}{2} \|y\|_2^2 = g(0) + \int_0^1 g'(t)\, dt$$

$$\geq g(0) + 1 \cdot g'(0)$$

$$= f(x) + \frac{c}{2} \|x\|_2^2 + \nabla f(x)^T (y-x)$$
$$- c x^T y$$

$$\Rightarrow f(y) \geq f(x) + \nabla f(x)^T (y-x) + \frac{c}{2} \| y-x \|_2^2$$

#

Moreover, we can show: $f \in C^1$ is convex

$$\Leftrightarrow f(y) \geq f(x) + \nabla f(x)^T (y-x)$$

$$\Leftrightarrow \underline{\langle \nabla f(y) - \nabla f(x), y-x \rangle \geq 0}$$

$\downarrow$

$\underline{\underline{\nabla f \text{ is a monotone increasing operator}}}$

Proposition: if $f \in C^1$ is c-strongly convex function,

then $\nabla f(x) = 0 \Rightarrow x$ is the UNIQUE minimizer

Pf: $f \in C^1$ is c-strongly convex $\forall x, y$.

$$\Rightarrow f(y) \geq f(x) + \nabla f(x)^T (y-x) + \frac{c}{2} \|y-x\|_2^2$$

if $\nabla f(x) = 0$, then $\forall y$, $f(y) \geq f(x) + \frac{c}{2}\|y-x\|_2^2$

$$\Rightarrow \forall y \neq x, \quad f(y) \geq f(x) + \frac{c}{2}\|y-x\|_2^2 > f(x)$$

$\Rightarrow$ we show that, $\forall y \neq x$, $f(y) > f(x)$

$\Rightarrow$ x is the unique minimizer

Remark: we can also prove via $\begin{cases} \text{contradiction} \\ \langle \nabla f(x) - \nabla f(y), x-y \rangle \geq 0 \end{cases}$

Remark: Up to now, we find a SPECIAL class of function $\begin{cases} \text{convex} \\ \text{strongly convex (unique)} \end{cases}$ such that stationary point $\Longleftrightarrow$ $\boxed{\text{global minimizer}}$