Discussion #3: Sep. 5

- Briefly have example on how to apply the prop. that we had before
- Examples related to ML

Example: ① $f(x) = \max_{i \in \mathcal{A}} \{ a_i^T x + b_i \}$ s.t. $\{\underline{a_i}\}$ & $\{b_i\}$ are fixed & $\mathcal{A}$ is given

Convex ?

✓ $\Rightarrow$ convex based on $\sup_{\alpha \in \mathcal{A}} \underbrace{f_\alpha(x)}_{\text{Convex}}$

last property on page 7 of Discussion 2

② expected value :

$f(x, u)$ is convex in x & u

$\Rightarrow g(x) = E_u(f(x,u))$ is convex

(Based on "non-negative infinite integral property on page 7 of Discussion 2)

③ $f(\underline{x}) = x_{[1]} + x_{[2]} + x_{[3]}$

where $x_{[i]}$ represents the ith largest x_j's.

$\begin{pmatrix} 2 \\ 3 \\ 4 \\ 1 \end{pmatrix}$ $x_{[2]} = 3$

$f(x) = \max_i c_i^T x$ where c_i

are the set of ALL vectors with components zero & three 1's.

$$\begin{bmatrix} 0 & 0 & 1 & 1 & 1 & 0 & \cdots \end{bmatrix}$$
$$\hookrightarrow \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & \cdots \end{bmatrix}$$

($f(x)$ is convex based on last property of page 7 of Discussion 2)

(4) max distance to any set.

$$\sup_{s \in S} \| \underbrace{x - s}_{\text{affine}} \|$$

affine, $\|\cdot\|$ is a convex function

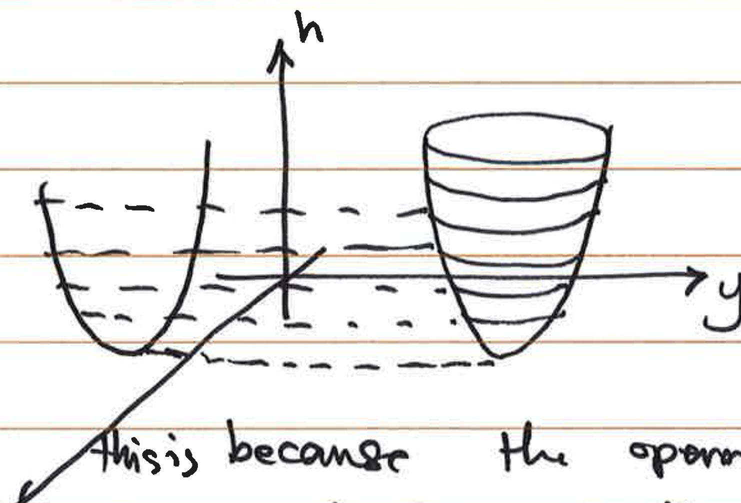
(Page 7 of Discussion 2) $\longrightarrow$ and due to supremum prop. of convex function

$\Downarrow$

Convex.

* Another convexity preserving operation is that minimizing over some variables.

$h(x,y)$ is convex in x & y $\Longrightarrow$ $f(x) = \inf_y h(x,y)$ is convex in x

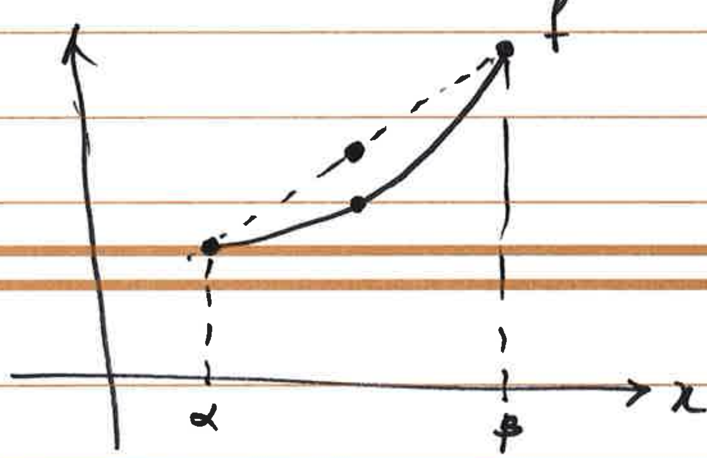


This is because the operation above corresponds to projection of the epigraph $(x,y,t) \longrightarrow (x,t)$

Jensen's Inequality (restatement of convexity)
true for convex function f

$$f\left(\int x\,p(x)\,dx\right) \leqslant \int f(x)\,p(x)\,dx$$

$$f(E(x)) \leqslant E(f(x))$$



$$\frac{f(\alpha)+f(\beta)}{2} \geqslant f\left(\frac{\alpha+\beta}{2}\right)$$

$$\min \quad f(x)$$
$$\text{s.t.} \quad g_i(x) \leq 0$$

To be convex $f(\cdot)$ & $g_i(\cdot)$'s should be convex.

ML-Related Examples:

1. Linear regression problem:

$$D = \{(\underline{x}_i, y_i)\}_{i=1}^{N}$$

Goal: find a linear model to minimize

$$\mathcal{E} \triangleq \frac{1}{N} \sum_{i=1}^{N} (y_i - w^T \underline{x}_i)^2$$

$\Rightarrow$ opt. prob. $\quad \min_{w \in \mathbb{R}^d} \mathcal{E} \qquad (1)$

is (1) convex?

$y_i - w^T x_i$ is affine

$(\cdot)^2$ is convex

postive-sum ~~remains~~ keeps the convexity

$\Rightarrow$ $\mathcal{E}$ is a convex in w

other proof:

$$\mathcal{E} = \frac{1}{N} \| \underline{y} - X\omega \|^2 =$$

where $\underline{y} \triangleq \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$ $\quad X \triangleq \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_n^T \end{bmatrix}$

$$= \frac{1}{N} (y - X\omega)^T (y - X\omega)$$

$$= \frac{1}{N} y^T y - \frac{2}{N} y^T X \omega + \frac{1}{N} \omega^T X^T X \omega$$

we know that if $\nabla^2 \mathcal{E} \succeq 0$

$$\Downarrow$$

(1) is convex.

$$\nabla f = \begin{pmatrix} \frac{\partial f}{\partial w_1} \\ \vdots \\ \frac{\partial f}{\partial w_n} \end{pmatrix}$$

$$\nabla_\omega (a^T \omega) = \nabla_\omega (a_1 w_1 + \cdots + a_n w_n) = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = a$$

$$\nabla_\omega (\omega^T a) = a$$

$$\nabla_\omega (\omega^T Q \omega) = \nabla_\omega \left(\sum_{i=1}^{d} \sum_{j=1}^{d} Q_{ij} w_i w_j \right)$$

$Q_{ij} = [Q]_{i,j}$

$$
= \begin{pmatrix}
\frac{\partial}{\partial w_1}\left(w_1 \times 2 \sum_{j \neq 1} Q_{1j} w_j + w_1 Q_{11} w_1 \right) = \underbrace{2 \sum_{j\neq 1} Q_{1j} w_j + 2 Q_{11} w_1}_{= 2 Q_{1,:}\, w} \\
\vdots \\
2 Q_{d,:}\, w
\end{pmatrix}
$$

$$
= 2Q \underline{w}
$$

$$
\nabla^2 f \triangleq \begin{bmatrix}
\frac{\partial}{\partial x_1}\left(\frac{\partial f}{\partial x_1}\right) & \cdots & \frac{\partial}{\partial x_d}\left(\frac{\partial f}{\partial x_1}\right) \\
 & & \\
\frac{\partial}{\partial x_n}\left(\frac{\partial f}{\partial x_1}\right) & \cdots & \frac{\partial}{\partial x_d}\left(\frac{\partial f}{\partial x_d}\right)
\end{bmatrix}
$$

$$
\nabla_w^2 \left(a^T w \right) = 0_{d \times d}
$$

$$
\nabla_w^2 \left(w^T Q w \right) =
\begin{pmatrix}
\frac{\partial}{\partial w_1}\left(2 \sum Q_{1j} w_j\right) = 2Q_{11} & 2Q_{12} & \cdots & 2Q_{1d} \\
 & & \ddots & \\
 & & & 2Q_{dd}
\end{pmatrix}
$$

$$
= 2Q
$$

$$\nabla \mathcal{E} = \frac{1}{N}\left(-2y^T X\right) + \frac{1}{N} 2 X^T X w$$

$$\nabla^2 \mathcal{E} = \frac{1}{N} \underbrace{2 X^T X}_{\succeq 0}$$

$$\Rightarrow \quad z^T X^T X z = \|Xz\|^2 \geq 0$$

$$\Rightarrow \quad \text{Convex} \quad \checkmark$$

(2) Considering Cross-entropy loss function problem for $D = \{(x_i, y_i)\}_{i=1}^N$ where $y_i \in \{0,1\}$ and the loss function is

$$\ell(y_i, h(x_i)) \triangleq -y_i \log h(x_i) - (1-y_i) \log(1 - h(x_i))$$

with $h(x_i) \triangleq \sigma(w^T x_i)$ and

$$\sigma(z) = \frac{1}{1+e^{-z}}$$

Show that this is a convex problem, i.e.

$$\min_{\underline{w}} \sum_{i=1}^N \ell(y_i, h(x_i)) \quad \text{is convex.}$$

$$J = +\sum_{i=1}^{N} \ell(y_i, h(x_i))$$

$$= -\sum_{i=1}^{N} y_i \log \sigma(w^T \underline{x}_i) + (1-y_i) \log\left(1-\sigma(w^T x_i)\right)$$

we need to show $-\log \sigma(w^T \underline{x}_i)$

and $-\log(1-\sigma(w^T x_i))$ are both convex.

(based on checking ∇^2 for both functions)

$$\nabla_w \left(-\log \sigma(w^T x_i)\right)$$

$$= \frac{-1}{\sigma(w^T x_i)} \nabla_w \left(\sigma(w^T \underline{x}_i)\right)$$

$$= \frac{-1}{\sigma(w^T x_i)} \sigma'(w^T x_i) \nabla_w \left(w^T \underline{x}_i\right)$$

$$= -(1-\sigma(w^T \underline{x}_i))\, \underline{x}_i$$

$$\sigma'(z) = \frac{\partial}{\partial z}\left(\frac{1}{1+e^{-z}}\right)$$

$$= \frac{-1}{(1+e^{-z})^2} \times (-e^{-z})$$

$$= \underbrace{\frac{e^{-z}}{1+e^{-z}}}_{1-\sigma(z)} \times \underbrace{\frac{1}{1+e^{-z}}}_{\sigma(z)}$$

$$\nabla_w^2 \left(-\log \sigma(w^T \underline{x}_i)\right) = \nabla_w \left(-(1-\sigma(w^T \underline{x}_i)) \underline{x}_i\right)$$

$$= \underline{x}_i \begin{bmatrix} \frac{\partial}{\partial w_1}\left(\sigma(w^T \underline{x}_i)\right) = \sigma(w^T \underline{x}_i) \times (1-\sigma(w^T \underline{x}_i))\, x_1 \\ \vdots \\ \\ \end{bmatrix}$$

$$= \underbrace{\sigma(w^T \underline{x}_i)\left[1-\sigma(w^T \underline{x}_i)\right]}_{\geq 0} \underline{x}_i \underline{x}_i^T$$

$$\underline{x}_i \underline{x}_i^T \succeq 0$$

$$z^T \underline{x}_i x_i^T z = (x_i^T z)^2 \geq 0$$

$$\Downarrow$$

$-\log \sigma(w^T \underline{x}_i)$ is convex in w

Similarly we can show that $-\log(1-\sigma(w^T \underline{x}_i))$ is convex.

$$\nabla_w \left(-\log(1-\sigma(w^T \underline{x}_i))\right) = \frac{-1}{1-\sigma(w^T \underline{x}_i)} \nabla_w \left(1-\sigma(w^T \underline{x}_i)\right)$$

$$= \frac{1}{\cancel{1-\sigma(w^T x_i)}} \sigma(w^T x_i) \left[\cancel{1-\sigma(w^T x_i)}\right] \underline{x}_i$$

$$= \sigma(w^T \underline{x}_i)\, \underline{x}_i$$

$$\nabla^2\left(-\log\left(1-\sigma(w^T x_i)\right)\right)$$

$$= \underbrace{\sigma(w^T x_i)\left(1-\sigma(w^T x_i)\right)}_{\geq 0} \underbrace{x_i x_i^T}_{\succeq 0}$$

$$\Downarrow$$

convex.

$\nabla J = X^T(\underline{\hat{y}} - \underline{y})$ $\quad \hat{y}_i = \sigma(w^T x_i) = h(x_i)$

$\nabla^2 J = X^T R X$ $\quad R$ is diagonal

$R_{ii} = h(x_i)\left(1-h(x_i)\right)$

Example (3) : Ridge Regression :

$$\min \ \|\underline{y} - X\underline{w}\|^2 \quad (*)$$
$$\text{s.t.} \quad \|w\|_2^2 \leq t$$

$\left.\begin{array}{ll} y - X\underline{w} & \text{is affine} \\ \|\cdot\|^2 & \text{is convex} \end{array}\right\} \Rightarrow$ obj func. is convex.

$g(\underline{w}) = \|w\|_2^2 - t$ is convex

and therefore $(*)$ is a convex opt. problem