02/12/2021

Lecture - 19

$\|\cdot\|_p : \mathbb{R}^n \to \mathbb{R}$

$p \geq 1$

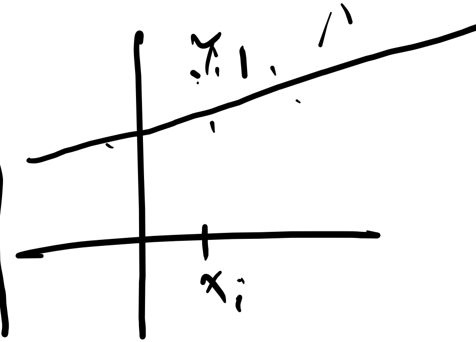
$$\|\vec{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$$

norm → way of measuring size of a vector.

$$\|\vec{x}\|_1 = \sum_{i=1}^{n} |x_i|$$

$$\|\vec{x}\|_2 = \left(\sum_{i=1}^{n} |x_i|^2 \right)^{1/2}$$

$$\|\vec{x}\|_\infty = \max\{ |x_1|, \dots, |x_n| \}$$



$(X_1, Y_1) \; - - - \; (X_n, Y_n)$

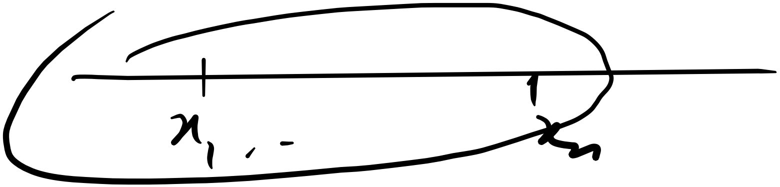
$$Y_{ni} = \beta_0 + \beta_1 X_i \quad \forall \; 1 \leq i \leq n.$$

$$\min_{\beta_0, \beta_1} \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i)^2$$

$$\begin{bmatrix} Y_1 - \beta_0 - \beta_1 X_1 \\ \vdots \\ Y_n - \beta_0 - \beta_1 X_n \end{bmatrix}$$

$$\min_{\beta_0, \beta_1} \sum_{i=1}^{n} |Y_i - \beta_0 - \beta_1 X_i|$$

$$\min_{\beta_0, \beta_1} \max \{ |Y_i - \beta_0 - \beta_1 X_i| \}_i$$



I want to measure <u>central tendency</u>

$$\text{mean} = \frac{x_1 + \cdots + x_n}{n} = y$$

$$\min_{y \in \mathbb{R}} \left(\sum_{i=1}^{n} (x_i - y)^{2} \right)^{1/2}$$

$$\arg\min_{y \in \mathbb{R}} \sum_{i=1}^{n} (x_i - y)^2 = \frac{x_1 + \cdots + x_n}{n}$$

$$\arg\min_{y \in \mathbb{R}} \sum_{i=1}^{n} |x_i - y| = \boxed{\text{median}}$$

$$\arg\min_{y \in \mathbb{R}} |x_i - y|_0 = \text{mode.}$$

$\Big($ ℓ^0 - "norm".

$$\sum |x_i|^0 \to \begin{cases} 1 & \text{if } x_i \neq 0 \\ 0 & \text{if } x_i = 0 \end{cases}$$

$\Big)$

"Chekhov's gun":

Operations that preserve convexity of functions;

(Building new convex functions from old).

(1) Non-negative weighted sums:

If f is convex (on Ω), and $\alpha \geq 0$, then αf is convex.

If f_1, f_2 are convex (on Ω), then $f_1 + f_2$ is convex.

$(\alpha_1 f_1 + \dots + \alpha_k f_k)$ is convex if $\alpha_i \geq 0$, $f_1, \dots f_k$ are convex on Ω.

(2) Composition with an affine mapping.

If $f : \mathbb{R}^n \to \mathbb{R}$ is convex, $A \in \mathbb{R}^{n \times m}$, $\vec{b} \in \mathbb{R}^n$,

$A : \mathbb{R}^m \to \mathbb{R}^n$

$g : \mathbb{R}^m \to \mathbb{R}$

$g(\vec{x}) := f(A\vec{x} + \vec{b})$

then g is convex on $\mathbb{R}^m$.

(3) Pointwise maximum and supremum.

$f(\vec{x}) := \max\{f_1(\vec{x}), \ldots, f_k(\vec{x})\}$

$f_1, \ldots, f_k$ are convex on Ω.

$\vec{x} \to x_1, \ldots, \vec{x} \to x_n$

Example:

For $\vec{x} \in \mathbb{R}^n$, $x_{(i)}$ — i^{th} largest component of $\vec{x}$,

$$x_{(1)} \geq \cdots \geq x_{(n)}$$

$$f(\vec{x}) = \boxed{\sum_{i=1}^{r} x_{(i)}}$$ (sum of the r largest elements of $\vec{x}$).

$$= \max\{ \boxed{x_{i_1} + \cdots + x_{i_r}} ;\ 1 \le i_1 < \cdots < i_r \le n \}$$

(maximum of all possible sums of r different components of $\vec{x}$).

$\binom{n}{r}$ terms

Representation of convex functions as pointwise supremum of affine functions

Theorem: Let $f : \Omega \to \mathbb{R}$ (convex domain) be a convex function (with closed epigraph). Then

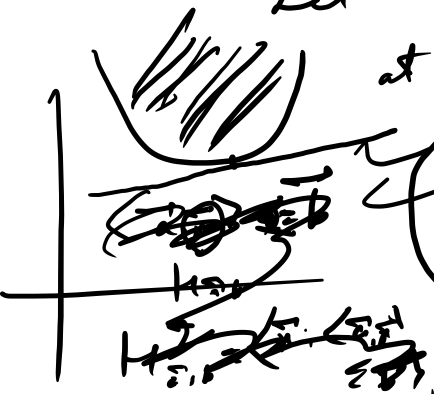
$$f(\vec{x}) = \sup \{ h(\vec{x}) : h \text{ is an affine underestimator of } f \}$$

$$h(\vec{x}) \leq f(\vec{x})$$

Proof: $\mathrm{epi}(f)$ is a closed convex set of $\mathbb{R}^n \times \mathbb{R}$

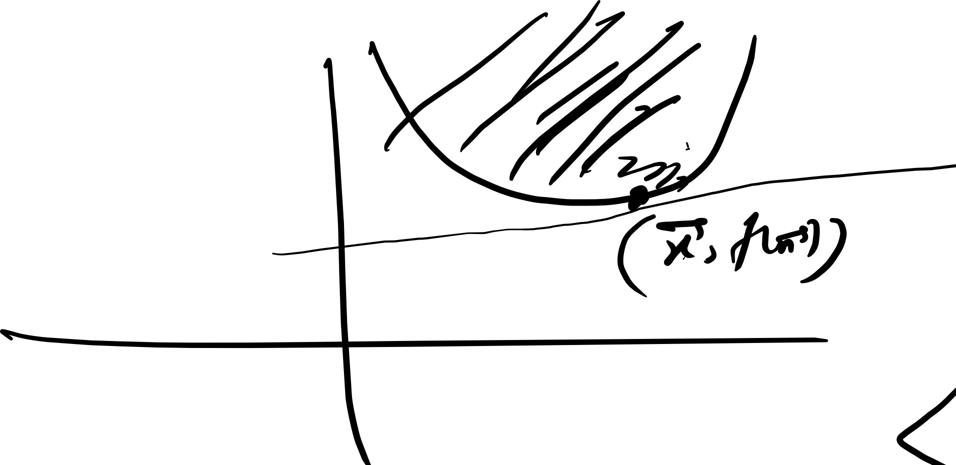
$(\vec{x}, f(\vec{x})) \in \partial\, \mathrm{epi}(f) = \mathrm{bd}(\mathrm{epi}(f))$.

Let $H(\vec{a}, b)$ be a supporting hyperplane at $(\vec{x}, f(\vec{x}))$, $(\vec{a}, b) \neq 0$.



$$\begin{bmatrix}\vec{a}\\ b\end{bmatrix}^T \begin{bmatrix}\vec{x} - \vec{z}\\ f(\vec{x}) - t\end{bmatrix} \le 0 \quad \forall (\vec{z}, t) \in \mathrm{epi}(f)$$

$$[\vec{a}, b]\begin{bmatrix}\vec{x}\\ f(\vec{x})\end{bmatrix} \le [\vec{a}, b]\begin{bmatrix}\vec{z}\\ t\end{bmatrix}$$



$\mathbb{R}^{n+1} \cong \mathbb{R}^n \times \mathbb{R}$

$H_{(\vec{a}, b), \alpha}$

$\left\langle \begin{bmatrix} \vec{a} \\ b \end{bmatrix}, \begin{bmatrix} \vec{x} \\ f(\vec{x}) \end{bmatrix} \right\rangle = \alpha$

$\left\langle \begin{bmatrix} \vec{a} \\ b \end{bmatrix}, \begin{bmatrix} \vec{z} \\ t \end{bmatrix} \right\rangle \leq \left\langle \begin{bmatrix} \vec{a} \\ b \end{bmatrix}, \begin{bmatrix} \vec{x} \\ f(\vec{x}) \end{bmatrix} \right\rangle$

$$\langle \vec{a}, (\vec{x} - \vec{z}) \rangle + b(f(\vec{x}) - t) \leq 0$$

$(\vec{z}, t) \in epi(f)$

$\Leftrightarrow \quad t \geq f(\vec{z})$.

$s = t - f(\vec{z})$

$$\langle \vec{a}, \vec{x} - \vec{z} \rangle + b(f(\vec{x}) - f(\vec{z}) - s) \leq 0$$

$\forall \boxed{\vec{z} \in \Omega}$

$\in$

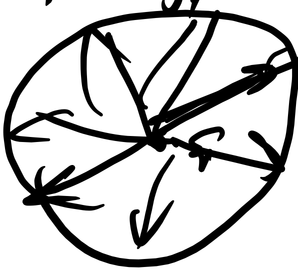
Assume $\vec{x} \in \mathrm{int}(\Omega)$.

(1) Let $s \to \infty$. $b(f(\vec{x}) - f(\vec{z}) - s) \to -\infty$ if $b < 0$

(contradiction)

Thus, $\boxed{b \geq 0}$.

(2) If $b = 0$, then $\langle \vec{a}, \vec{x} - \vec{z} \rangle \leq 0$ $\forall \vec{z} \in \Omega$.



$\vec{z}$ (not possible)

$\vec{a} = \vec{0}$.

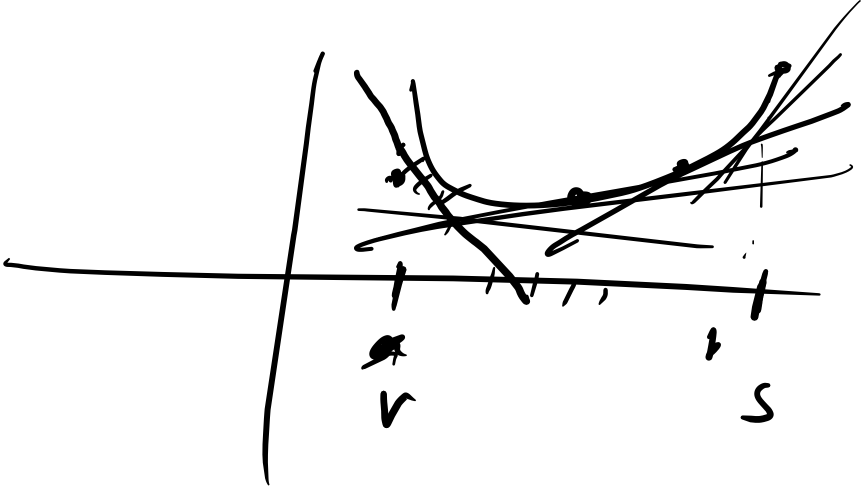
For $\vec{x} \in \text{int}(\Omega)$, the separating hyperplane has ~~posi~~ strictly positive offset.

$$\langle \vec{a}, \vec{x} - \vec{z} \rangle + b\left(f(\vec{x}) - f(\vec{z})\right) \le 0$$

$$\Rightarrow \quad b f(\vec{z}) \ge \langle \vec{a}, \vec{x} - \vec{z} \rangle + b f(\vec{x})$$

$$\Rightarrow \quad f(\vec{z}) \ge \underbrace{\boxed{\left\langle \frac{1}{b}\vec{a}, \vec{x} - \vec{z} \right\rangle + f(\vec{x})}}_{\text{aff}}$$

$\forall \vec{z} \in \Omega$. affine underestimator s.t. $(\vec{x}, f(\vec{x}))$ on this affine hyperplane



$h \le f$

affine hyperplane

Let's if possible,

$(\sup h)(r) < f(r)$

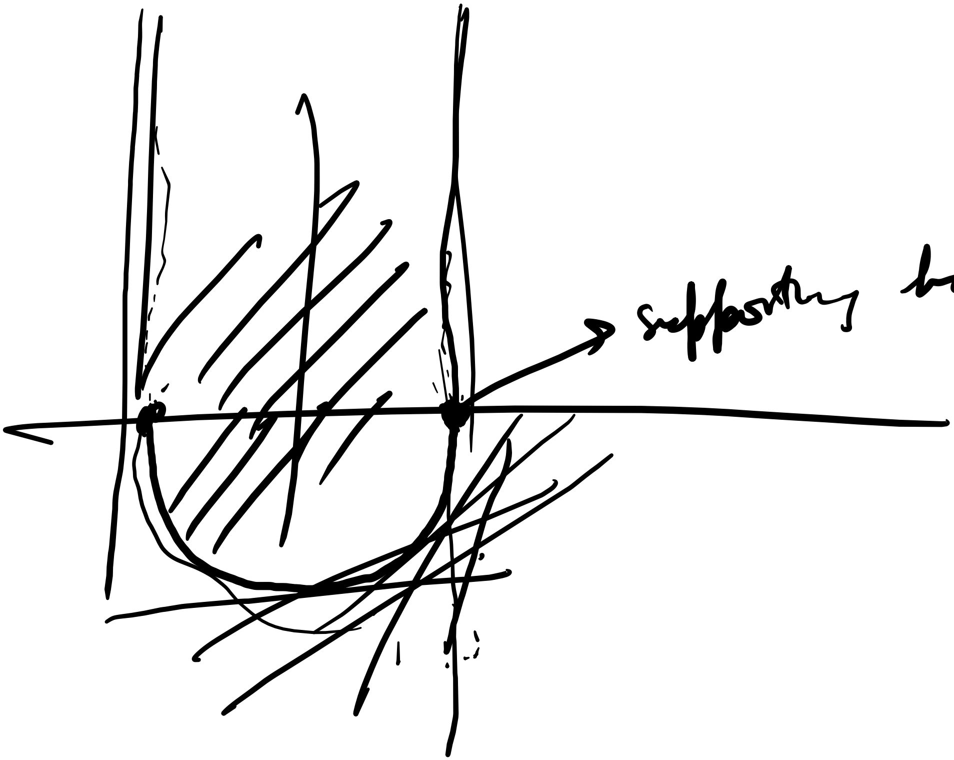
$(r, \boxed{(\sup h)(r)}) \notin \operatorname{epi}(f)$

$(\sup h)(\vec{x}) = f(\vec{x})$

$\forall \vec{x} \in \operatorname{int}(S)$

By separating hyperplane theorem, $\exists\, h$ s.t. $h(\vec{x}) \le f(\vec{x})$

s.t. $h(r) > (\sup h)(r)$



supporting hyperplane at (1, 0)

Minimization:

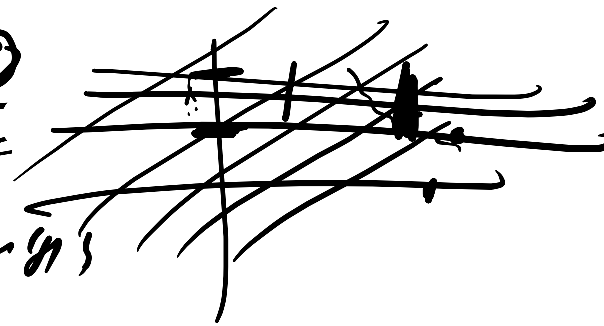
Supremum of an arbitrary family of convex functions is convex.

$$(\vec{x}, \vec{y}) \in \Omega \subseteq \mathbb{R}^n \times \mathbb{R}^m$$

f is convex on Ω.

$C \to$ nonempty convex set in $\mathbb{R}^m$,

$$g(\vec{x}) = \inf_{\vec{y} \in C} f(\vec{x}, \vec{y}) =$$



$\mathrm{dom}(g) = \{\vec{x} \in \mathbb{R}^n : (\vec{x}, \vec{y}) \in \mathrm{dom}(f)\}$

If $g(\vec{x}) > -\infty$, for some $\vec{x} \in \mathrm{dom}(g)$,

then $g(\vec{x}) > -\infty$, for all $\vec{x} \in \mathrm{dom}(g)$.

and g is convex.

Proof: $\vec{x}_1, \vec{x}_2 \in \mathrm{dom}(g)$. $((\vec{x}_1, \vec{\xi}_1) \in \mathrm{dom}(f)$, $(\vec{x}_2, \vec{\xi}_2) \in \mathrm{dom}(f)$,

given $\varepsilon > 0$, $\exists\ \vec{\xi}_1, \vec{\xi}_2$ s.t.

$f(\vec{x}_1, \vec{\xi}_1) \leq g(\vec{x}_1) + \varepsilon$,

$f(\vec{x}_2, \vec{\xi}_2) \leq g(\vec{x}_2) + \varepsilon$,

$$g(\theta \vec{x}_1 + (1-\theta)\vec{x}_2)$$

$$= \inf_{y \in C} f(\theta \vec{x}_1 + (1-\theta)\vec{x}_2, \vec{y})$$

$$\leq f(\theta \vec{x}_1 + (1-\theta)\vec{x}_2, \theta \vec{y}_1 + (1-\theta)\vec{y}_2)$$

$$\leq \theta f(\vec{x}_1, \vec{y}_1) + (1-\theta) f(\vec{x}_2, \vec{y}_2)$$

$$\leq \theta g(\vec{x}_1) + (1-\theta) g(\vec{x}_2) + \varepsilon$$

$$\forall \varepsilon > 0,$$

$$\Rightarrow g(\theta \vec{x}_1 + (1-\theta)\vec{x}_2) \leq \theta g(\vec{x}_1) + (1-\theta) g(\vec{x}_2)$$