Lagragian Duality (Bloomberg)

Linear programs (linear objective & constraints) were the focus

Nonlinear programs

Convex: $\forall c_1, c_2 \in C \quad \forall \theta \in [0,1]$

$$\theta c_1 + (1-\theta) c_2 \in C$$

A function f: convex if dom f is a convex set and if for all x,y ∈ dom f and $0 \le \theta \le 1$

$$f(\theta x + (1-\theta) y) \le \theta f(x) + (1-\theta) f(y)$$

Eg. ① $x \to ax+b$ convex concave

② $x \to |x|^p$ for $p \geq 1$ convex for all p.

③ $x \to e^{ax}$ convex on R $\forall R$.

④ norm

⑤ max

$\boxed{\text{Strictly convex}}$: exclude end points.

convex: local minimum $\to$ global optimum

strictly convex: local minimum $\to$ unique global minimum.

$\boxed{\text{General optimization Problem}}$

$$\min \quad f_0(x)$$

$$\text{subj to} \quad f_i(x) \leq 0 \quad i=1 \dots m$$

$$h_i(x) = 0 \quad i=1 \dots p.$$

$x \in R^n$ $\boxed{\text{optimization variable}}$

f_0 $\boxed{\text{objective function}}$

domain nonempty

Set of points in constraint feasible set.

If x feasible and $f_i(x) = 0$,

then $f_i(x) \le 0$ $\boxed{\text{active}}$ at x,

$\boxed{\text{Optimal value}}$ $p^* = \inf \{ f_0(x) \mid x \text{ satisfies all constraints} \}$

x^* $\boxed{\text{optimal point}}$

$\boxed{\text{Lagrangian}}$: $L(x, \lambda) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x)$

$\lambda \ge 0$ (Lagrangian multipliers

dual variable)

$$\sup_{\lambda \ge 0} L(x, \lambda) = \sup_{\lambda \ge 0} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) \right)$$

$$= \begin{cases} f_0(x) & \text{when } f_i(x) \le 0 \ \forall i \\ \infty & \text{o.w.} \end{cases}$$

just another form of the original obj

So $\boxed{\text{primal form}}$ $p^* = \inf_x \sup_{\lambda \geq 0} L(x, \lambda)$

$\boxed{\text{dual form}}$: sup inf - sup

$$\downarrow$$

$$d^* = \sup_{\lambda \geq 0} \inf_x L(x, \lambda)$$

$\boxed{\text{weak duality}}$: $p^* \geq d^*$

$$\sup_{z \in Z} \inf_{w \in W} f(w, z) \leq \inf_{w \in W} \sup_{z \in Z} f(w, z)$$

Proof: $\inf_{w \in W} f(w, z) \leq f(w_0, z_0) \leq \sup_{z \in Z} f(w_0, z)$

$$\Rightarrow \inf_w f(w, z_0) \leq \sup_{z_0 \in Z} f(w, z_0)$$

$$\sup_{z_0 \in Z} \inf_{w \in W} f(w, z_0) \leq \inf_{w_0 \in W} \sup_{z \in Z} f(w_0, z)$$

$\boxed{\text{Duality gap}}$: $p^* - d^*$

$\boxed{\text{strong duality}}$: $p^* = d^*$

$\boxed{\text{Lagrangian dual function}}$

$$g(\lambda) = \inf_x L(x, \lambda) = \inf_x \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) \right)$$

$$p^* \geq \sup_{\lambda \geq 0} g(\lambda) = d^*$$

$$p^* \geq g(\lambda) \quad \forall \lambda$$

max $g(\lambda)$

subj $\lambda \geq 0$

Complementary Slackness

def of strong duality

$$\lambda_i^* f_i(x^*) = 0$$

$$f_0(x^*) = g(\lambda^*) = \inf L(x, \lambda^*)$$

$$\leq L(x^*, \lambda^*)$$

$$= f_0(x^*) + \sum_{i=1}^{m} \underbrace{\lambda_i^* f_i(x^*)}_{\leq 0}$$

$$\leq f_0(x^*)$$

$$\Rightarrow \sum_{i=1}^{m} \underbrace{\lambda_i^* f_i(x^*)}_{\leq 0}$$

Convex Optimization

Convex: A set is convex if

$$\{x, x'\} \in S \Rightarrow [x, x'] \in S$$

$$\{ax + (1-a)x' : a \in [0,1]\}$$

Eg. ① All of $\mathbb{R}^d$

② Empty set

③ Half spaces $\{x \in \mathbb{R}^d, a^T x \le b\}$

④ Intersection of convex sets.

⑤ polyhedra $\{x \in \mathbb{R}^d : A^T x \le b\}$

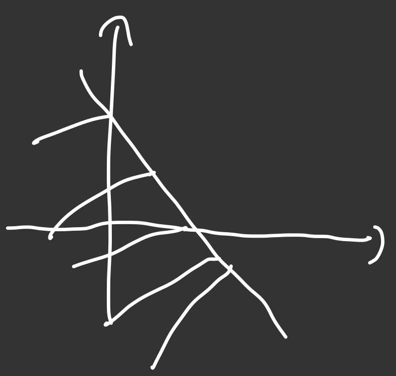
$= \bigcap_{i=1}^{m} \{x \in \mathbb{R}^d : a_i^T x \le b_i\}$

⑥ convex hulls.

$= \{ \sum_{i=1}^{k} a_i \cdot x_i : k \in \mathbb{N}, x_i \in S, a_i \ge 0, \sum_{i=1}^{k} a_i = 1 \}$,

Proof half space is convex.

$H := \{x \in \mathbb{R}^d : a^T x \le b\}$.



Let $x, x' \in H$ be given, $a \in [0,1]$

want: $ax + (1-a)x' \in H$

show $a^T(ax + (1-a)x') \in H$

$= a^T a x + a^T (1-a) x'$

$= a^T b + (1-a) b$

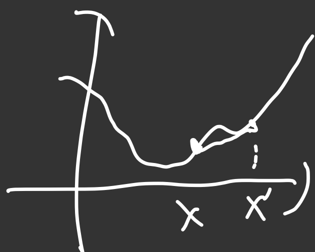
$\le b$

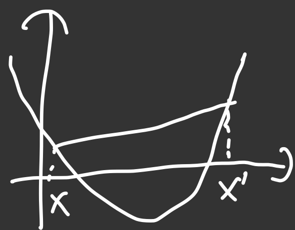
$\boxed{\text{Epigraph:}}$ area above curve.

$$epi(f) := \{(x,y) \in R^{d+1} \; y \geq f(x)\}$$

A function is convex if its epigraph is convex

$\boxed{\text{Convex function}}$ $f((1-\alpha)x + \alpha x') \leq (1-\alpha) \cdot f(x) + \alpha \cdot f(x')$



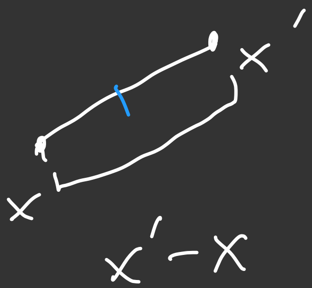


$f(x) = b^T x$ for any $b \in R^d$

$f(x) = \|x\|$ for any norm $\|\cdot\|$

$f(x) = x^T A x$ for symmetric positive semidefinite A

$f(x) = \ln(\sum_{i=1}^{d} \exp(x_i))$ which approximates $\max_i x_i$



Line segment

point betwn x x'

$$x + (x' - x)a$$

Proof $f(x) = \|x\|$ convex

Let $x, x' \in R^d$ $\quad \alpha \in [0,1]$

$$f(\alpha x + (1-\alpha) x')$$

$$= \|\alpha x + (1-\alpha) x'\| \le \|\alpha x\| + \|(1-\alpha) x'\|$$

$$\le \alpha \|x\| + (1-\alpha) \|x'\|$$

$$\alpha f(x) + (1-\alpha) f(x')$$

Eg

$\boxed{\text{Summations}}$; if $(f_1, \cdots f_k)$ convex and $(\alpha_1 \cdots \alpha_k)$ nonnegative.

$$x \to \alpha_1 f_1(x) + \cdots + \alpha_k f_k(x) \text{ convex}$$

$\boxed{\text{Affine composition}}$: $f(x)$ convex

$$\to f(Ax+b) \quad \text{convex}$$

$\boxed{\text{Maxima}}$ max convex

Suppose l_i convex

$\hat{R}$ convex, since

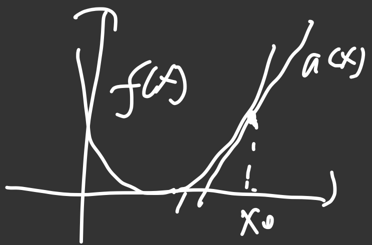
$$\hat{R}(w) = \frac{1}{n}\sum_{i=1}^{n} l(w^T x_i, y_i)$$

$$= \frac{1}{n}\sum_{i=1}^{n} l_i(w).$$

Convexity of $\boxed{\text{differentiable}}$ equations.

If $f: R^d \to R$ differentiable, then f convex if and only if

$$f(x) \geq f(x_0) + \nabla f(x_0)^T (x - x_0)$$



Increasing slopes:

$$(\nabla f(x) - \nabla f(y))^T (x - y) \geq 0$$

$\boxed{\text{Twice differentiable functions}}$

$$\nabla^2 f(x) \geq 0$$

Strict convexity

Function values: $\forall x \neq y \quad \forall \alpha \in (0,1)$

$$f(\alpha x + (1-\alpha) y) < \alpha f(x) + (1-\alpha) f(y)$$

Derivatives: $\forall x \neq y$

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

Hessian: $\forall x$

$$\forall x \quad \nabla^2 f(x) \succ 0$$

Strong convexity

at least as quadratic

Function values: $\forall x \neq y \quad \forall \alpha \in (0,1)$

$$f(\alpha x + (1-\alpha) y) \leq \alpha f(x) + (1-\alpha) f(y) - \frac{\lambda \alpha (1-\alpha)}{2} \|x-y\|^2$$

Derivatives: $\forall x \neq y$

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

Hessian: $\forall x$

$$\forall x \quad \nabla^2 f(x) \succeq \cancel{0}\ \lambda I$$

Logistic loss $z \to \ln(1+\exp(-z))$

strictly convex

Squared loss $z \to \frac{1}{2}(1-z)^2$

strongly convex

Convex: local optimum

$\to$ global optimum

$\boxed{\beta\text{-smooth}}$: $f(w') \leq f(w) + \nabla f(w)^T(w'-w) + \frac{\beta}{2}\|w'-w\|^2$

Theorem.

$$\underbrace{\min_{i \leq t} \|\nabla f(w_{i-1})\|^2}_{\text{norm of gradient descent}} \leq \frac{1}{t}\sum_{i=1}^{t}\|\nabla f(w_{i-1})\|^2$$

$$\leq \underbrace{\frac{2\beta}{t}\left(f(w_0) - \min_w f(w)\right)}_{O(\frac{1}{t})}$$

$\boxed{\text{Stochastic GD}}$

take $|S| = B$ (minibatch)

$$\frac{1}{B} \sum_{i \in S} l'(y_i f(x_i; w)) \, y_i \nabla_w f(x_i; w)$$