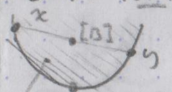
Convexity $f : \Omega \to \mathbb{R}^n$

$C \subseteq \mathbb{R}^n$ is convex set if
$\forall x, y \in C \Rightarrow \theta x + (1-\theta) y \in C$
for $\theta \in [0,1]$

- Convex fctn: $f(\theta x + (1-\theta) y)$ [A]
  $\leq \theta f(x) + (1-\theta) f(y)$ [B]



Concave $= -f$ convex

$epi(f) = \{(x,t) \mid x \in \Omega, t \geq f(x)\}$
$\hookrightarrow f$ convex fctn $\Leftrightarrow epi(f)$ convex set

Differentiable func. f: over open set

1^{st}: $f(y) \geq \underbrace{f(x) + \langle \nabla f(x), y - x \rangle}_{\hat{f}_1(y;x)}$

2^{nd}: $\nabla^2 f(x) \succeq 0$ [P.S.D]

- Strict: If ineqs. never equal for $\theta \notin \{0,1\}$, $x \neq y$.

If f_1, f_2 are convex fctns, then so's the pointwise max:
$f(x) = \max\{f_1(x), f_2(x)\}$
w/ domain $dom(f) = dom(f_1) \cap dom(f_2)$
As f_i is convex, so's domains
$epi(f) = \{(x,t) \mid x \in dom(f), \max(f_1(x), f_2(x)) \leq t\}$
$= \{(x,t) \mid x \in dom(f_1) \cap dom(f_2), f_1(x) \leq t \wedge f_2(x) \leq t\}$
$= epi(f_1) \cap epi(f_2)$

f convex $\Leftrightarrow g(t) = f(x + t v)$ convex over $dom(g) = \{t \in \mathbb{R} \mid g(t) \in dom f\}$
- $dom(f)$ convex
- $dom(g)$ convex

$\Rightarrow g(\theta t_1 + (1-\theta) t_2) = f(x + (\theta t_1 + (1-\theta) t_2) v)$
$\leq \theta g(t_1) + (1-\theta) g(t_2)$

M-Strong convexity
$f(y) \geq \hat{f}_1(y;x) + \underbrace{(y-x)\left(\frac{M}{2} I\right)(y-x)}_{\frac{M}{2}\|y-x\|_2^2}$

$$x^T A x = \sum_{i=1}^n \sum_{j=1}^n A_{ij} x_i x_j$$
$$= \sum_{j=1}^n A_{kj} x_k x_j + \underbrace{\sum_{i \neq k} \sum_{j=1}^n A_{ij} x_i x_j}$$
$$\sum_{i \neq k}^n A_{ik} x_i x_k + \sum_{i \neq k} \sum_{j \neq k} A_{ij} x_i x_j$$
$$= A_{kk} x_k^2 + \sum_{j \neq k} A_{kj} x_k x_j + \sum_{i \neq k} A_{ik} x_i x_k + \sum_{i \neq k} \sum_{j \neq k} A_{ij} x_i x_j$$
$$\frac{\partial}{\partial x_k}[\ldots] = 2 A_{kk} x_k + \sum_{j \neq k} A_{kj} x_j + \sum_{i \neq k} A_{ik} x_i$$
$$= (Ax)_k + (A^T x)_k = [(A + A^T) x]_k$$
$$\Rightarrow \nabla(x^T A x) = (A + A^T) x$$

$$\nabla[\|Ax - b\|_2^2] = \nabla[x^T A^T A x - 2 b^T A x + b^T b]$$
$$= ((A^T A) + (A^T A)^T) x - 2 A^T b + 0$$
$$= 2 A^T (Ax - b)$$
$$\nabla^2[\cdot] = 2 A^T A$$

L-smooth: For $L > 0$, f is L-smooth if...
[1] $h(x)$ convex; $h(x) := \frac{L}{2}\|x\|_2^2 - f(x)$
[2] $\forall x, y \in \mathbb{R}^n$: $f(y) \leq f(x) + \nabla f(x)^T (y - x) + \frac{L}{2}\|y - x\|_2^2$
$\rightarrow$ (f. UB.)
[3] $\|\nabla f(y) - \nabla f(x)\|_2 \leq L \|x - y\|_2^2$

①: $\nabla h = Lx - \nabla f(x)$
$\nabla^2 h = LI - \nabla^2 f(x) \succeq 0 \Leftrightarrow h(x)$ convex

$\nabla_A tr(AB) = B^T$; $\nabla_B tr(AB) = A^T$
$\nabla_A tr(ABBA) = (BBA)^T + (ABB)^T$
Let $\mathcal{L} = \|X - W_2 W_1 X\|_F^2$
$\nabla_{W_1} \mathcal{L} = \nabla_2 [tr(\cancel{XX^T}) + W_2 W_1 X X^T W_1^T W_2^T - 2 W_2 W_1 X X^T]$
$= 2(W_1 X X^T W_1^T W_2^T)^T + 2(W_1 X X^T)^T$
$= 2(W_2 W_1 - I) X X^T W_1^T$
$\nabla_{W_1} \mathcal{L} = 2 W_2^T (W_2 W_1 - I) X X^T$

Monotone: $(\nabla f(y) - \nabla f(x))^T (y - x) \geq 0$
- If $dom(f)$ convex, $\forall x, y \in dom(f)$

Vector Calc.

Chain: $g(t) = f(x(t))$
$$\frac{dg}{dt}(t) = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(x(t)) \cdot \frac{dx_i}{dt}(t)$$
$f : n \to 1$
$x : 1 \to n$
$g : 1 \to 1$

Gradient: $f : n \to 1$; $\nabla f : n \to n$
$\nabla f = [\cdot\cdot \ \partial f(x) / \partial x_i \ \cdot\cdot]^T$
- Derivative $\equiv [\nabla f]^T$
- Grad points in direction of steepest ascent
- Rate-of-change by $\|\nabla f\|_2$
- Chain: $\nabla h(x) = [D g(x)]^T \nabla f(g(x))$
  $f : p \to 1$; $g : n \to p$; $h : n \to 1$

Jacobian: $\underline{f} : n \to m$; $D\underline{f} : n \to m \times n$
$$D\underline{f}(x) = \begin{bmatrix} \vdots \\ \nabla f_i(x)^T \\ \vdots \end{bmatrix} = \begin{bmatrix} f_1/x_1 & f_1/x_n \\ f_m/x_1 & f_m/x_n \end{bmatrix}$$
- For $f : n \to 1$, $D f(x) = [\nabla f(x)]^T$
- Chain rule: $\underline{h}(x) = \underline{f}(\underline{g}(x))$
  $D \underline{h}(x) = [D\underline{f}(\underline{g}(x))] \cdot [D \underline{g}(x)]$
  $f : p \to m$
  $g : n \to p$
  $h : n \to m$
- Directional $u \in \mathbb{R}^n$: $D_u f(x) = u^T [\nabla f(x)]$

Hessian: $f : n \to 1$; $\nabla^2 f : n \to n \times n$
$$\nabla^2 f(x) = D(\nabla f)(x) = \begin{bmatrix} /\partial x_1^2 & /\partial x_n \partial x_1 \\ /\partial x_1 \partial x_n & /\partial x_n^2 \end{bmatrix}$$
e.g. $[\nabla^2 f(x)]_{ij} = \partial^2 f / \partial x_i \partial x_j$

Taylor's: $f : n \to 1$ around point x_0
$\hat{f}_1(x; x_0) = f(x_0) + [\nabla f(x_0)]^T (x - x_0)$
$\Rightarrow f(x_0 + \Delta x) - f(x_0) = [\nabla f(x_0)]^T \Delta x$
Hyperplane: $b = p^T x$
$f(x) = \hat{f}_k(x; x_0) + o(\|x - x_0\|_2^k)$
$f(x + \Delta x) = f(x) + [\nabla f(x)]^T \Delta x + o(\|\Delta x\|_2^1)$

$\underline{f} : n \to m \Rightarrow \hat{\underline{f}}_1(x; x_0) = \underline{f}(x_0) + [D\underline{f}(x_0)] \times (x - x_0)$

LinAlg

$det(A) = \prod_{i=1}^n \lambda_i\{A\}$
$diag(\lambda_1 \cdots \lambda_n) = \sum_{i=1}^n \lambda_i (e_i e_i^T)$

$tr(A) = \sum_{i=1}^n \lambda_i\{A\}$
$tr(ABC) = tr(BCA) = tr(CAB)$; $a^T a = tr(a a^T)$
$tr(A + B) = tr(A) + tr(B)$

$det(cA) = c^n det(A)$; $det(A^{-1}) = 1/det(A)$
$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \Rightarrow A^{-1} = \begin{bmatrix} d & -c \\ -b & a \end{bmatrix} / det(A)$
$p_A(\lambda) = \lambda^2 - \lambda Tr(A) + det(A) = 0$
$\lambda = Tr(A) \pm \sqrt{Tr(A)^2 - 4 det(A)} / 2$
$v_1 = \begin{bmatrix} b \\ \lambda_1 - a \end{bmatrix}$, $v_2 = \begin{bmatrix} b \\ \lambda_2 - a \end{bmatrix}$

Fundamental Thm $A \in \mathbb{R}^{m \times n}$
$\mathcal{N}(A) + \mathcal{R}(A^T) = \mathbb{R}^n$
$\mathcal{N}(A^T) + \mathcal{R}(A) = \mathbb{R}^m$
PF| $S^\perp := \{x \in \mathbb{R}^n \mid x^T s = 0, \forall s \in S\}$
s.t. $S \oplus S^\perp = \mathbb{R}^n$
Show $\mathcal{N}(A) = \mathcal{R}(A^T)^\perp$ via $\subseteq$ and $\supseteq$

$rk(A + B) \leq rk(A) + rk(B)$
PF| $R(A+B) = \{\underbrace{(A+B)x}_{Ax + Bx} \mid x \in \mathbb{R}^n\} \subseteq R(A) + R(B)$
$rk(A+B) = dim(R(A+B))$
$\leq dim(\{Ax : x \in \mathbb{R}^n\}) + dim(\{By : y \in \mathbb{R}^n\})$
$= rk(A) + rk(B)$ ▭

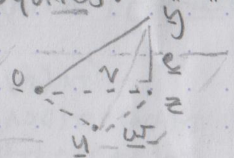
For $A \in \mathbb{R}^{m \times n}$, $rk(A) = \min(m,n) \geq r$
$n = r + dim(\mathcal{N}(A))$

Convexity of $\|X\|_F^2$: For $\theta \in [0,1]$
$\|\theta X + (1-\theta) Y\|_F^2$
$= \langle \theta X + (1-\theta) Y, = \rangle$
$= \theta^2 \|X\|_F^2 + (1-\theta)^2 \|Y\|_F^2 + 2\theta(1-\theta)\langle X, Y \rangle$
$\leq [\quad " \quad] + 2\theta(1-\theta)\|X\|_F \|Y\|_F$
$\leq [\quad " \quad] + \theta(1-\theta)(\|X\|_F^2 + \|Y\|_F^2)$
$= \theta \|X\|_F^2 + (1-\theta)\|Y\|_F^2$

$ab \leq \frac{1}{2}(a^2 + b^2)$

proj $\underline{a}$ onto $\underline{b} = (\underline{a}^T \underline{b}) \underline{b}$

Regression | Least-Squares: $\min_x \|Ax - y\|_2^2$



Uniqueness:
$\|y-u\|_2^2 = \|v\|_2^2$
$= \|w\|^2 + \|e\|^2$
$z - u > 0$
$> \|e\|_2^2 = \|y - z\|_2^2 \qquad w = z - u$
$\therefore$ z closest pt to y in $R(A)$

$e = y - z$
$z = \mathrm{proj}_{R(A)}\, y$
$\in R(A)$
$v = y - u$
$u \in R(A)$

$\} \; e \perp \mathrm{Col}(A)$

$0 = \langle A, y - z \rangle$
$= A^T(y - Ax^*)$
$= A^T y - A^T A x^*$
$\therefore x^* = (A^TA)^{-1} A^T y$
exists if A full rk

Min-norm: $\min_x \|x\|_2^2$ s.t. $Ax = y$

- Overdetermined system: ∞ solns
  $\hookrightarrow m > n$ for $A \in \mathbb{R}^{m\times n}$ [Wide]

Let $x = u + v$ for $y = Ax$
$u \in N(A) \Rightarrow Au = 0 \qquad = Au + Av$
$v \in R(A^T) \Rightarrow v = A^T w \qquad = 0 + AA^T w$
$\Rightarrow w^* = (AA^T)^{-1} y$
$\therefore x^* = u^* + v^* = A^T(AA^T)^{-1} y$
$\Rightarrow \infty$ choices for u^*
$\Rightarrow \infty$ solns x^*

Ridge: $\min_{x \in \mathbb{R}^n} \left[\|Ax - y\|_2^2 + \lambda \|x\|_2^2 \right]$

$\nabla_x f = 2A^TAx - 2A^Ty + 2\lambda x$
$0 = 2(A^TA + \lambda I)x - 2A^Ty$
$\Rightarrow x^* = (A^TA + \lambda I)^{-1} A^T y$

- SVD POV: $\tilde{x}_i = (\sigma_i / \sigma_i^2 + \lambda)\, \tilde{y}_i$
  for $\tilde{x} = V^T x$, $\tilde{y} = U^T y$
  $\hookrightarrow$ If $\lambda \gg \sigma_i$ then don't blowup/stay small

Kernel Equivilence:

$\lambda A^T + A^TAA^T = A^T(\lambda I + AA^T)$
$= (\lambda I + A^TA)A^T$
$\therefore (\lambda I + A^TA)^{-1} A^T = A^T(\lambda I + AA^T)^{-1}$

Matrix Form

$$\min_{X \in \mathbb{R}^{n\times m}} \|AX - B\|_F^2$$
$$= \min_{x_i \in \mathbb{R}^n} \sum_{i=1}^{k} \|Ax_i - b_i\|_2^2$$
$$= \left[\min_{x,y} \{f(x) + g(y)\} = \{\min_x f(x) + \min_y g(y)\} \right]$$
$$\sum_{i=1}^{k} \min_{x_i \in \mathbb{R}^n} \|Ax_i - b_i\|_2^2$$

Tikhonov

$$\min_x \left[\|W_1(Ax - y)\|_2^2 + \|W_2(x - x_0)\|_2^2 \right]$$
$$x^* = (A^T W_1^2 A + W_2^2)^{-1} (A^T W_1^2 y + W_2^2 x_0)$$

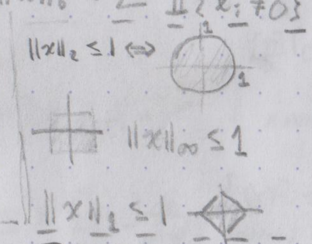
Norms: $f: X \to \mathbb{R}$ norm iff:

1.) $f(x) \geq 0$; $f(x) = 0 \iff x = 0$
2.) $f(\alpha x) = |\alpha| f(x)$
3.) $f(x + y) \leq f(x) + f(y)$

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

$\|x\|_\infty = \max_i |x_i|$

$\|x\|_0 = \sum \mathbb{1}\{x_i \neq 0\}$



$\|x\|_2 \leq 1 \iff$; $\|x\|_\infty \leq 1$; $\|x\|_1 \leq 1$

Ineq.: Cauchy-Swartz

$|x^T y| \leq \|x\|_2 \|y\|_2 \cos\theta$
$\leq \|x\|_2 \|y\|_2$

Holders:
$1 < p, q < \infty$ s.t. $\frac{1}{p} + \frac{1}{q} = 1$
$|x^T y| \leq \sum_{i}^{n} |x_i y_i|$
$\leq \|x\|_p \|y\|_q$

$\|x\|_2 / \sqrt{n} \leq \|x\|_\infty \leq \|x\|_2 \leq \underbrace{\|x\|_1}_{\langle \mathbb{1}, |x| \rangle \leq \|\mathbb{1}_n\|_2 \|x\|_2} \leq \sqrt{n} \|x\|_2 \leq n \|x\|_\infty$

Orthonormal doesn't effect ℓ^2:
$\|Ax - b\|_2^2 = (Ax - b)^T QQ^T (Ax - b) = \|Q^T(Ax - b)\|_2^2$

$\|A\|_F = \sqrt{\mathrm{tr}(A^TA)} = \sqrt{\sum_i^m \sum_j^n A_{ij}^2}$
$\mathrm{tr}(\cdot)$ cyclical: $A^TA = V\Sigma^2V^T$
$= \sqrt{\mathrm{tr}(V^T V \Sigma^2)} = \sqrt{\sum_{i=1}^{r} \sigma_i^2}$

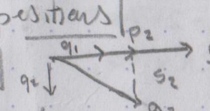
$\|A\|_2 = \max_{x \in \mathbb{R}^n, \|x\|_2 = 1} \underbrace{\|Ax\|_2}_{\sqrt{x^TA^TAx}}$
$= \sqrt{\lambda\{A^TA\}} = \sigma_{max}$

$A_k \in \mathrm{argmin}_{\mathrm{rk}(B) \leq k} \|A - B\|_{2/F}$

$A_k = \sum^{k} \sigma_i u_i v_i^T$
e.g. $\|A - A_k\|_2 = \sigma_{k+1} \leq \|A - B\|_2$
$\forall B \in \mathbb{R}^{m\times n}: \mathrm{rk}(B) \leq k$

Decompositions

QR



$q_1 = a_1 / \|a_1\|_2$
For i in $[2, k]$:
$p_i = \sum_{j=1}^{i-1} q_j (q_j^T a_i)$
$s_i = a_i - p_i$
$q_i = s_i / \|s_i\|$

$r_{ij} = \langle a_i, q_j \rangle$
$r_{ii} = \|q_i\|_2$
$a_i = \sum_{j=1}^{i} r_{ij} q_j$

Eigen. $A \in S^n$

$A = U \Lambda U^T = \sum_i \lambda_i u_i u_i^T$

$A^TA = \sum_{i=1}^{r} \sum_{j=1}^{r} \sigma_i^2 v_i \cancel{u_i^T u_j} v_j^T$
PSD follows non-neg Eigens

SVD $A = U\Sigma V^T$, $\Sigma \in \mathbb{R}^{m\times n}$

$\mathrm{span}(U) = [U_r \; U_{m-r}] \begin{bmatrix} \Sigma_r & \\ & 0 \end{bmatrix} \begin{bmatrix} V_r^T \\ V_{n-r}^T \end{bmatrix}$
$= \mathrm{col}(A)$
$\mathrm{span}(V_r^T) = \sum_{i=1}^{r} \sigma_i u_i v_i^T$
$= \mathrm{row}(A)$

$\mathrm{span}(U_{m-r}) = N(A^T)$
$\mathrm{span}(V_{n-r}) = N(A)$

$(\lambda_i, u_i) = \mathrm{Eigens}(AA^T)$
$(\lambda_i, v_i) = \mathrm{Eigens}(A^TA)$
$\sigma_i = \sqrt{\lambda_i}$; $u_i = Av_i / \sigma_i$
$\|u_i\|_2^2 = \frac{v^TA^TAv}{\sigma_i^2}$
$= \frac{\lambda}{\sigma^2} v_i^T v_i = 1$

Gram-Schmidt $m-r$ vectors. Not unique

$Av_i = \sigma_i u_i$; $A^T u_i = \sigma_i v_i$

$A^\dagger = V_r \Sigma^\dagger U_r^T = \sum v_i u_i^T / \sigma_i$

PCA $X \in \mathbb{R}^{n\times d} = [-x_i^T-]$

Covariance: $C = \frac{1}{n} X^T X \in S^d$

$\min_{w \in \mathbb{R}^d, \|w\| = 1} \mathrm{err}(w) = \frac{1}{n} \sum_{i=1}^{n} \|x_i - w(w^T x_i)\|_2^2$
$= \frac{1}{n} \sum (\|x_i\|_2^2 - (x_i^T w)^2)$
$= \frac{1}{n} \sum \|x_i\|_2^2 + \max_w w^T X^T X w / n$
$\lambda_{max}\{C\}$

Spectral: Let $A \in S^n$, e.g. $A^T = A$

1.) $\forall \lambda\{A\} \in \mathbb{R}$
2.) Eigenspaces $\Xi_i = N(\lambda_i I - A)$ are orthogonal
3.) Eigendecomp exists if equal alg + geo. multiplicities
- Geo. $\dim(\Xi_i := N(\lambda_i I - A))$
- Alg. power of $(x - \lambda_i)$ in $\det(xI - A) =: p_A(x)$

- $N(A) = \{u_i : \lambda_i = 0\}$
- $R(A) = \{u_i : \lambda_i \neq 0\}$

$\lambda_{max} = \max_{x \in \mathbb{R}^n} \frac{x^TAx}{x^Tx}$

P.S.D.: $\forall x$, $x^TAx \geq 0$
$\hookrightarrow$ Denoted as $A \succeq 0$, $A \in S_+^n$
- $\lambda\{A\} \geq 0$
- $\exists B \succeq 0: A = B^TB$, i.e. $B = U\Lambda^{1/2}U^T$
- $0 \leq A_{ii} = B_{:i}^2$ — Alt: $x^TAx \geq 0$; $e_i^TAe_i = A_{ii} \geq 0$

Descent Methods

Newton's: Use 2^{nd}-ord approximate

$$x_{t+1} = x_t - [\nabla^2 f(x_{(t)})]^{-1} \cdot \nabla f(x_{(t)})$$

- If $f(\cdot)$ is quadratic, converge after single step
- Affine Invariance: Let $T \in \mathbb{R}^{n \times n}$ be non-singular & $g(y) := f(Ty)$

Then $\nabla g = T^T \nabla f(Ty)$

$\nabla^2 g = T^T (\nabla^2 f(Ty)) T$

Hence, Newton step looks like

$$-\nabla^2 g(y)^{-1} \nabla g(y) = -T^{-1} \nabla^2 f(Ty)^{-1} \nabla f(Ty)$$

So if we start $y_0 = T^{-1} x_0$ and apply Newton's on $g(y)$ we'll get $y_{t+1} = T^{-1} x_{t+1}$

GD | Consider $F(x) = \frac{1}{2}\|Ax - b\|_2^2 + \frac{1}{2}\lambda\|x\|_2^2$

$\nabla F = 0 = A^TA x - A^Tb + \lambda x$

$\Rightarrow x^* = (A^TA + \lambda I)^{-1} A^T b$

$x_{t+1} = x_t - \eta((A^TA + \lambda I)x_t - A^Tb)$

$= (I - \eta(A^TA + \lambda I))x_t + \eta A^T b$

$x_{t+1} - x^* = (\;\;)x_t + \underbrace{\eta(A^TA + \lambda I)(A^TA + \lambda I)^{-1}A^Tb - (A^TA + \lambda I)^{-1}A^Tb}_{(\eta(A^TA + \lambda I) - I)\,x^*}$

$= (I - \eta(A^TA + \lambda I))(x_t - x^*)$

$= (\;\;)^t (x_0 - x^*)$

Converge $\Leftrightarrow |\lambda_{max}\{I - \eta(A^TA + \lambda I)\}| < 1$

$|1 - \eta(\sigma_1^2 + \lambda)|$

Consider $g(x) = \frac{1}{2}x^TMx + x^Tb + c$; $M \in S_{++}$

$\nabla g = 0 = \frac{1}{2}(M + M^T)x - b \Rightarrow x^* = M^{-1}b$

$x_{t+1} = x_t - \eta(Mx_t - b)$

$x_{t+1} - x^* = (I - \eta M)x_t + \underbrace{\eta(MM^{-1})b - M^{-1}b}_{(\eta M - I)\,x^*}$

$= (I - \eta M)(x_t - x^*)$

$= (I - \eta M)^t (x_0 - x^*)$

Converge $\Leftrightarrow |\sigma_i\{I - \eta M\}| < 1$

$\Rightarrow \eta < 2/\lambda_{max}\{M\}$

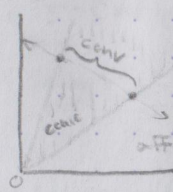
Possible to converge with larger LR (η) IF $(x_0 - x^*) \in$ EigenSpace$(I - \eta M)$ corresponding to Eigenval. $\in (-1, 1)$

Analysis | $= \bigcup_{\{A \subseteq S : |A| \le n+1\}} \text{conv}(A)$

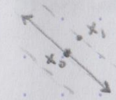
$\text{conv}(S) = \{\sum \theta_i x_i \mid \theta \ge 0,\ x_i \in S,\ \mathbf{1}^T\theta = 1\}$

$\text{conic}(S) = \{\sum \theta_i x_i \mid \theta \ge 0,\ x_i \in S\}$

$\text{aff}(S) = \{\sum \theta_i x_i \mid \theta_i \in \mathbb{R},\ x_i \in S,\ \mathbf{1}^T\theta = 1\}$



SVM | $y_i = \begin{cases} +1 : w^Tx_i + b > 0 \\ -1 : \quad \text{"} \quad < 0 \end{cases}$



$\equiv \underbrace{y_i(w^Tx_i + b)}_{|w^Tx_i + b|} > 0$

$x_0 \in H = \{x \mid w^Tx + b = 0\}$

Let $x_1 = x_0 + kw$ s.t. $x_1 \in \{x \mid w^Tx + b = 1\}$

So $w^Tx_1 + b = 1$

$\cancel{w^Tx_0} + w^Tkw + b = 1 \Rightarrow k = 1/\|w\|_2^2$

$\therefore \|x_0 - x_1\|_2 = k \cdot \|w\| = 1/\|w\|_2$

Soft Margin:

- Small C: Maximizes margin; underfits; robust to outliers; flat boundary
- Big C: keeps slack vars small; overfits; sensitive to outliers; more sinuous
  - $C \to \infty \equiv$ Hard-Margin

KKT | $\xi_i \ge 1 - y_i((w^*)^Tx_i - b^*)\ \forall i$

$\lambda_i, \mu_i, \xi_i \ge 0$.

$0 = \nabla_w \mathcal{L} = w^* - \sum_{i=1}^{n} \lambda_i^* y_i x_i$

$0 = \nabla_b \mathcal{L} = \sum_{i=1}^{n} \lambda_i^* y_i$

$0 = \nabla_\xi \mathcal{L} = C \cdot \mathbf{1} - \mu^* - \lambda^*$

$\mu_i \xi_i = 0 \wedge \lambda_i(1 - y_i((w)^Tx_i - b) - \xi) = 0$

1] $\lambda_i^* = 0 \Rightarrow \mu_i^* = C > 0 \Rightarrow \xi_i = 0$

$w^* = \sum \lambda_i^* y_i x_i^*$ has zilch from x_i

2] $\lambda_i^* = C \Rightarrow \mu_i^* = 0$; $y_i((w^*)^Tx_i - b) = 1 - \xi_i \le 1$

↳ On or violates margin

3] $\lambda_i^* \in (0, C) \Rightarrow \mu_i^* \in (0, C)$; $\xi_i = 0$

↳ Exactly on margin

Hyperplane | $\dim(H) = \dim(\{x \mid c^Tx = 0\})$

$= \dim(N(c^T))$

Rank-Nullity: $\dim(N(c^T)) + \underbrace{\dim(R(c^T))}_{1} = n$

$\Rightarrow \dim(H) = n - 1$

Given x^* as one side of $H = \{x \mid c^T(x - x_0) = 0\}$. Find vector on other side.

wlog. $c^T(x^* - x_0) > 0$

$-c^T(x^* - x_0) < 0$

$c^T(x_0 - x^*) < 0$

$x_0 - x^* = z - x_0 \Rightarrow z = 2x_0 - x^*$

Linear Separability: Given data $x_i \in \mathbb{R}^n$ and binary labels $y_i \in \{-1, 1\}$ we want to find $H = \{x \mid h^Tx + b = 0\}$ to separate $y_i = -1$ from $y_j = +1$

LP: $p^* = \min_{h, b, z} \; -z$

s.t. $h^Tx_i + b \le 0 \quad \forall i : y_i = +1$

$h^Tx_i + b \ge z \quad \forall i : y_i = -1$

$\Rightarrow$ Possible (Feasible) iff $z^* > 0$

Chebyshev Center: $P = \{x \mid a_i^Tx \le b_i, \forall i\}$ $[m]$

WTF biggest ball $B(x_0, r) \in P$

$x \in B(x_0, r) \equiv x = x_0 + u$ for $\|u\|_2 \le r$

should sat. P; $a_i^T(x_0 + u) \le b_i$

$a_i^Tx_0 + \max_{\|u\| \le r} a_i^Tu \le b_i$

$r \cdot a_i/\|a_i\|_2$

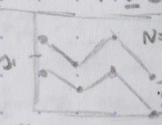
$\therefore p^* = \min_{x_0, r} \; -r$

s.t. $a_i^Tx_0 + r\|a_i\|_2 \le b_i$

$r \ge 0$

Shalem SOCP

Given $\{(y_i, x_i, c_i)\}_{i=0}^{n+1}$



$$\min_{z, t} \sum_{i=1}^{n+1} t_i$$

s.t. $y_i - \frac{c_i}{2} \le z_i \le y_i + \frac{c_i}{2} \quad i \in [0, n+1]$

$\left\| \begin{bmatrix} x_i \\ z_i \end{bmatrix} - \begin{bmatrix} x_{i-1} \\ z_{i-1} \end{bmatrix} \right\|_2^2 \le t_i \quad i \in [1, n+1]$

Duality

Lagrangian: Incorporate constraints into problem itself

$$p^* = \min_{x \in \mathbb{R}^n} \max_{\substack{\lambda \in \mathbb{R}^m_+ \\ \nu \in \mathbb{R}^p}} \underbrace{\mathcal{L}(\underline{x}, \underline{\lambda}, \underline{\nu})}_{f_0(\underline{x}) + \sum_{i=1}^m \lambda_i f_i(\underline{x}) + \sum_{j=1}^p \nu_j h_j(\underline{x})}$$

$$d^* = \max_{\substack{\lambda \in \mathbb{R}^m_+ \\ \nu \in \mathbb{R}^p}} \underbrace{\min_{\underline{x} \in \mathbb{R}^n} \mathcal{L}(\underline{x}, \underline{\lambda}, \underline{\nu})}_{g(\underline{\lambda}, \underline{\nu})}$$

$g(\underline{\lambda}, \underline{\nu})$; concave always

Pf] pt-wise min over family of concave fctns

Minimax ineq: For any sets X, Y & $F: X \times Y \to \mathbb{R}$

$$\min_x \max_y F(x,y) \geq \max_y \min_x F(x,y)$$

Pf] $u(y) = \min_{\underline{x}'} F(x', y) \leq F(x,y) \leq \max_{y'} F(x, y') =: v(x)$

$\max_y u(y) \leq \min_x v(x)$

Implies weak duality always holds;

$p^* \geq d^* \equiv p^* - d^* \geq 0$] Positive gap

Slater's: If there exists a point in the relative interior of feasible set, then strong duality holds. Formally,

- Non-affine constraints strict: $f_i(x) < 0$
- Affine: $f_i(x) \leq 0$; $h_j(x) = 0$

× Only works for convex problems

Convex Problems:

min [convex] ≡ max [concave]

KKT Conditions

1] Primal Feasibility. $\tilde{\underline{x}}$ feasible for P;

$f_i(\tilde{\underline{x}}) \leq 0 \quad \forall i \in [m]$

$h_j(\tilde{\underline{x}}) = 0 \quad \forall j \in [p]$

2] Dual Feasibility. $(\tilde{\underline{\lambda}}, \tilde{\underline{\nu}})$ feasible for D;

$\tilde{\lambda}_i \geq 0 \quad \forall i \in [m]$

3] Complementary Slackness.

$\tilde{\lambda}_i f_i(\tilde{\underline{x}}) = 0 \quad \forall i \in [m]$

4] Stationary. $\tilde{\underline{x}} = \inf_{\underline{x}} \mathcal{L}(\underline{x}, \tilde{\underline{\lambda}}, \tilde{\underline{\nu}})$

$\equiv$

$\nabla \mathcal{L} = \underline{0}$

- If P is convex problem w/ differentiable constraints + obj fctn, & solns to KKT are found ⟹ strong duality
- If convexity holds ⟹ KKT suffices for optimality

QCQP]

$$\min_{\underline{x}} \tfrac{1}{2} \underline{x}^T H \underline{x} + \underline{c}^T \underline{x}$$

s.t. $\tfrac{1}{2} \underline{x}^T P_i \underline{x} + \underline{b}_i^T \underline{x} + r_i \leq 0$

$\underline{d}_i^T \underline{x} + f_i = 0$

where $H, P_1 \cdots P_m \in \mathbb{S}_+$

SOCP]

$$\min_{\underline{x}} \underline{c}^T \underline{x} \text{ s.t. } \|A_i \underline{x} - y_i\|_2 \leq \underline{b}_i^T \underline{x} + z_i$$

- Encode affine $A_i \underline{x} = y_i$ with $\underline{b}_i = \underline{0}, z_i = 0$ as ["] $\equiv \|A_i \underline{x} - y_i\|_2 \leq \underline{0}$

LP ⊂ QP ⊂ QCQP ⊂ SOCP ⊂ SDP ⊂ Conic

LP]

$$\min_{\underline{x} \in \mathbb{R}^n} c^T \underline{x} + d \text{ s.t. } A\underline{x} = y, \; \underline{x} \geq \underline{0}$$

$\mathcal{L}(\cdot) = \underline{c}^T \underline{x} - \underline{\lambda}^T \underline{x} + \underline{\nu}^T (A\underline{x} - y)$

$= (\underline{c} + \underline{\lambda} + A^T \underline{\nu})^T \underline{x} - \underline{\nu}^T y$

$$g(\cdot) = \begin{cases} -\underline{\nu}^T y & : \underline{c} - \underline{\lambda} + A^T \underline{\nu} = \underline{0} \\ -\infty & : \text{o/w} \end{cases}$$

$\therefore d^* = \max_{\lambda, \underline{\nu}} -\underline{\nu}^T y$ s.t. $\underline{c} - \underline{\lambda} + A^T \underline{\nu} = \underline{0}$, $\underline{\lambda} \geq \underline{0}$

Polyhedron: Intersection of finite set of half-spaces

- Set of form $\{\underline{x} \in \mathbb{R}^n_+ \mid A\underline{x} = y\}$] LP
- Polygon: Bounded polyhedron
- $\underbrace{\text{Has vertex}}$ ⟺ doesn't contain line

E.g. bounded & soln exists ⟹ LP || has opt val. achieved at vertex

QP]

$$\min_{\underline{x}} \tfrac{1}{2} \underline{x}^T H \underline{x} + \underline{c}^T x \text{ st } A\underline{x} \leq y, \; Cx = z$$

$\underbrace{\text{where } H \in \mathbb{S}^n_+}$

- If not, use $\hat{H} = \tfrac{1}{2}(H + H^T)$
- Needed for convexity; $\nabla^2 f = H \succeq 0$

1] $\underline{c} \in \mathcal{N}(H) \setminus \{0\} \Rightarrow \exists \lambda\{H\} = 0$.

Let $\underline{v}$ be unit e-vec for " st $\underline{c}^T \underline{v} \neq 0$

Let $x_t = -t \cdot \operatorname{sgn}(\underline{c}^T \underline{v}) \underline{v}$

so $\underline{c}^T \underline{x}_t = -t \cdot |\underline{c}^T \underline{v}| \therefore p^* = -\infty$

2] $\underline{c} \in \mathcal{R}(H) \Rightarrow \exists \underline{x}_0 \neq \underline{0}$ s.t. $\underline{c} = -H\underline{x}_0$

$f_0(\underline{x}) = \tfrac{1}{2}(\underline{x} - \underline{x}_0)^T H (x - x_0) - \tfrac{1}{2} \underline{x}_0^T H \underline{x}$

min. if $\underline{x}^* - \underline{x}_0 \in \mathcal{N}(H) \equiv \underline{x}^* = -\underbrace{H^\dagger}_{U\Lambda^\dagger U^T} \underline{c}$

↗ Vertex of polyhedron

Extreme Pt: Let $K \subseteq \mathbb{R}^n$. $\underline{x} \in K$ extreme if $\neg(\exists y, z \in K)\, \theta \in [0,1]$ st $\underline{x} = \theta y + (1-\theta) z$

Re-write "Tricks"

$x = x_+ - x_-$ for $x_+, x_- \geq 0$

where $x_+ = \max(0, x)$, $x_- = -\min(0, x)$ | $|x| = x_+ + x_-$

$|x| \equiv t$ s.t. $t \geq x$, $t \geq -x$

$$\min_{\underline{x}} \left[\max_{i \in [n]} x_i - \min_{j \in [n]} x_j \right] \equiv \min_{\underline{x}, t, u} [t - u] \text{ s.t. } t \geq x_i \; \forall i \in [n], \; u \leq x_i \; "$$

$G\underline{x} \leq h \equiv G\underline{x} + \underline{s} = \underline{h}$ st $\underline{s} \geq \underline{0}$

$\equiv G\underline{x}_+ - G\underline{x}_- + \underline{s} = \underline{h}$ st $\underline{x}_+, \underline{x}_-, \underline{s} \geq \underline{0}$

$$\underline{a}^T \underline{x} = b \equiv \begin{cases} \underline{a}^T \underline{x} \leq b \\ \underline{a}^T \underline{x} \geq b \equiv -\underline{a}^T \underline{x} \leq -b \end{cases}$$

$\|X\underline{w} - y\|_\infty \leq \varepsilon \equiv -\mathbb{1}\varepsilon \leq X\underline{w} - y \leq \mathbb{1}\varepsilon$

$\|x\|_1^2 \equiv \left(\sum_{i=1}^n v_i\right)^2 = \underline{v}^T Q \underline{v}$ s.t. $-x_i \leq v_i \leq x_i \; \forall i \in [n]$ where $Q = \mathbb{1}_{n \times n}$

Biggest Box:

$B(\underline{x}_0, r) = \{\underline{x}_0 + r\underline{u} \mid \|u\|_\infty \leq 1\}$