Least squares: $\min_{\vec{x}\in\mathbb{R}^n} \|A\vec{x}-\vec{y}\|_2^2$

$\vec{x}^* = (A^TA)^{-1}A^T\vec{y}$

Norm:

① Positive Definite: $f(\vec{x})\ge 0$

② Positive homogeneity: $f(\alpha\vec{x}) = |\alpha| f(\vec{x})$

③ Triangle inequality: $f(\vec{x}+\vec{y}) \le f(\vec{x})+f(\vec{y})$

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

L1: $\|\vec{x}\|_1 = \sum_{i=1}^n |x_i|$

L∞: $\|\vec{x}\|_\infty = \max_{i\in 1\cdots n}|x_i|$

L2: $\|\vec{x}\|_2 = \sqrt{\sum_{i=1}^n x_i^2}$

Cauchy Schwarz inequality

$|\vec{x}^T y| \le \|\vec{x}\|_2\|\vec{y}\|_2$

Projections: U onto V: $\text{proj}_V U = \left(\frac{\vec{U}^T\vec{V}}{\|\vec{V}\|_2}\right)\vec{V}$

Holder's inequality

$1\le p,q\le\infty : \frac{1}{p}+\frac{1}{q} = 1 \to \forall x,y\in\mathbb{R}^n$

$|\vec{x}^T\vec{y}| \le \sum_{i=1}^n |x_iy_i| \le \|\vec{x}\|_p\|\vec{y}\|_q$

Gram Schmidt (lin indep → orthonormal)

GS(lin indep set $\{\vec{a}_1\cdots\vec{a}_k\}$)

$\vec{q}_1 = \frac{\vec{a}_1}{\|\vec{a}_1\|}$

for $i\in\{2\cdots k\}$

$\vec{p}_i = \sum_{j=1}^{i-1}\vec{q}_j(\vec{q}_j^T\vec{a}_i)$ ★ project a onto prev q orthonormal vectors

$\vec{s}_i = \vec{a}_i - \vec{p}_i$ ★ subtract component of a that is in q vectors

$\vec{q}_i = \frac{\vec{s}_i}{\|\vec{s}_i\|_2}$ ★ normalize

end for

return orthonormal set $\{\vec{q}_1\cdots\vec{q}_k\}$

$\text{span}(\vec{a}_1\cdots\vec{a}_k) = \text{span}(\vec{q}_1\cdots\vec{q}_k)$

$\vec{a}_i = \sum_{j=1}^{i} r_{ji}\vec{q}_j \to A = QR$ ← upper triangular

↑ orthonormal columns

Fundamental Thm of Linear Algebra

$U\oplus V = \mathbb{R}^n$ iff ① $\vec{x} = \vec{x}_1+\vec{x}_2$, $\vec{x}_1\in U$, $\vec{x}_2\in V$ ② $\vec{x}_1+\vec{x}_2 = \vec{y}_1+\vec{y}_2 : \vec{x}_1=\vec{y}_1, \vec{x}_2=\vec{y}_2$

$A\in\mathbb{R}^{m\times n}$ $N(A)\oplus R(A^T) = \mathbb{R}^n$

$N(A^T)\oplus R(A) = \mathbb{R}^m$

orthogonal complement: $S^\perp = \{\vec{x}\in\mathbb{R}^n \mid \vec{s}^T\vec{x} = 0\ \forall \vec{s}\in S\}$

$S\oplus S^\perp = \mathbb{R}^n$

Minimum Norm soln: $\min_{\vec{x}\in\mathbb{R}^n}\|\vec{x}\|_2^2$ st $A\vec{x} = \vec{y}$

$\vec{x} = A^T(AA^T)^{-1}\vec{y}$

Symmetric: $A\in S^n$ if $A = A^T$ ★ DIAGONALIZABLE★

Spectral Thm: !!! ★SYMMETRIC MATRICES

① $\lambda_i\in\mathbb{R}$

② Eigenspaces correlated to distinct eigenvalues are orthogonal $\Phi_i = \text{Null}(A-\lambda_i I)$ if $\lambda_i\ne\lambda_j$ $\Phi_i\perp\Phi_j$

③ algebraic multiplicity $\mu_i = \dim(\Phi_i)$

$A = U\Lambda U^T$ U: orthonormal, Λ: diagonal

$(U\Lambda U^T)^k = U\Lambda^k U^T$!! (λ_i, U_i) = eigenvalue eigenvector pair

A=orthonormal → $U^TU = I$, $UU^T = I$

+ square → $U^T = U^{-1}$

Rayleigh coefficients ★ in proof use $A = U\Sigma V^T$

$\lambda_{max}\{A\} = \max_{\vec{x}\in\mathbb{R}^n, \vec{x}\ne\vec{0}}\frac{\vec{x}^TA\vec{x}}{\vec{x}^T\vec{x}} = \max_{\vec{x}\in\mathbb{R}^n, \|\vec{x}\|_2=1}\vec{x}^TA\vec{x}$

$\lambda_{min}\{A\} = \min_{\vec{x}\in\mathbb{R}^n, \vec{x}\ne\vec{0}}\boxed{\frac{\vec{x}^TA\vec{x}}{\vec{x}^T\vec{x}}} = \min_{\vec{x}\in\mathbb{R}^n, \|\vec{x}\|_2=1}\vec{x}^TA\vec{x}$

↑ Rayleigh coefc

Positive semidefinite

PSD $A\in S^n_+$ ① $\vec{x}^TA\vec{x}\ge 0\ \forall\vec{x}$ ② $\lambda_i\ge 0$

Positive definite

PD $A\in S^n_{++}$ ① $\vec{x}^TA\vec{x} > 0$ ② $\lambda_i > 0$

$A\in S^n_+$: unique symmetric PSD matrix $B = A^{1/2}$ st $A = B^2$

Principal Components Analysis

Data = rows: $X = \begin{bmatrix}x_1^T\\ \vdots \\ x_n^T\end{bmatrix}$ covariance $C = \frac{1}{n}X^TX = \frac{1}{n}\sum_{i=1}^n x_ix_i^T$

error of projections across data points: $err(\vec{w}_1) = \frac{1}{n}\sum_{i=1}^n\|\vec{x}_i - \vec{w}_1(w_1^Tx_i)\|_2^2$

optimization problem

$\min_{\vec{w}_1\in\mathbb{R}^d, \|\vec{w}_1\|_2=1} err(\vec{w}_1) = \frac{1}{n}\sum_{i=1}^n\|x_i\|_2^2 + \min_{w_1\in\mathbb{R}^d, \|w_1\|_2=1}\frac{1}{n}\sum_{i=1}^n(-x_i^Tw_1)^2$

$= \frac{1}{n}\sum_{i=1}^n\|x_i\|_2^2 - \max_{w_1\in\mathbb{R}^d, \|w_1\|=1}\frac{1}{n}(x_i^Tw_1)^2$

$= \frac{1}{n}\sum_{i=1}^n\|x_i\|_2^2 - \max_{w_1\in\mathbb{R}^d, \|w_1\|=1} w_1^TCw_1 = \frac{1}{n}\sum_{i=1}^n\|x_i\|_2^2 - \lambda_{max}\{C\}$

★ First principal component: eigenvector of largest eigenvalue of $C = \frac{X^TX}{n}$

SINGULAR VALUE DECOMP

$A = U\Sigma V^T = [U_r\ U_{m-r}]\begin{bmatrix}\Sigma_r & 0_{r\times n-r}\\ 0_{m-r\times r} & 0_{m-r\times n-r}\end{bmatrix}\begin{bmatrix}V_r^T\\ V_{n-r}^T\end{bmatrix}$

$= U_r\Sigma_rV_r^T = \sum_{i=1}^r\sigma_iu_iv_i^T$

★ U,V ORTHONORMAL

★ $\Sigma = [\sigma_1 \cdots \sigma_r]$ in descending order

Low Rank Approx

Frobenius Norm $\|A\|_F = \sqrt{\sum_{i=1}^m\sum_{j=1}^n A_{ij}^2} = \sum_{i=1}^r\sigma_i^2$

$\|A\|_F^2 = tr(A^TA)$

$\|UAV\|_F = \|UA\|_F = \|AV\|_F = \|A\|_F$

Eckart Young: $A_k\in\arg\min_{B\in\mathbb{R}^{m\times n}, rank(B)\le k}\|A-B\|_F$

$\|A-A_k\|_F^2\le\|A-B\|_F^2\ \forall B\in\mathbb{R}^{m\times n}$ rank(B)=k

Spectral Norm $\|A\|_2 = \max_{\vec{x}\in\mathbb{R}^n, \|\vec{x}\|_2=1}\|A\vec{x}\|_2 = \sigma_1$

↑ max singular value

Eckart Young:

$A_k \in \arg\min_{B\in\mathbb{R}^{m\times n}, rank(B)\le k}\|A-B\|_2$

$\|A-A_k\|_2\le\|A-B\|_2\ \forall B\in\mathbb{R}^{m\times n}$ rank(B)=k

$\vec{x}^TA\vec{y} = \sum_i\sum_j A_{ij}x_iy_j$ $\begin{bmatrix}x\\ y\end{bmatrix}^T\begin{bmatrix}A & B\\ C & D\end{bmatrix}\begin{bmatrix}x\\ y\end{bmatrix} = x^TAx + x^TBy + y^TCx + y^TDy$

GRADIENT (vector)

$$\nabla f(\vec{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(\vec{x}) \\ \vdots \\ \frac{\partial f}{\partial x_n}(\vec{x}) \end{bmatrix}$$

$\nabla f(\vec{x})^T$ = derivative

$\nabla(\vec{x}^T A \vec{x}) = (A + A^T)\vec{x}$

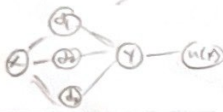
$\boxed{\text{product rule: } \frac{d(uv)}{dx} = u\frac{dv}{dx} + v\frac{du}{dx} \quad \text{quotient rule: } \frac{d\frac{u}{v}}{dx} = \frac{v\frac{du}{dx} - u\frac{dv}{dx}}{v^2}}$

JACOBIAN (matrix)

$$D\vec{f}(\vec{x}) = \begin{bmatrix} \nabla f_1(x)^T \\ \vdots \\ \nabla f_m(\vec{x})^T \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\vec{x}) & \cdots & \frac{\partial f_1}{\partial x_n}(\vec{x}) \\ & \ddots & \\ \frac{\partial f_m}{\partial x_1}(\vec{x}) & & \frac{\partial f_m}{\partial x_n}(\vec{x}) \end{bmatrix}$$

CHAIN RULE: $D\vec{h}(\vec{x}) = [D\vec{f}(\vec{g}(\vec{x}))][D\vec{g}(\vec{x})]$

$$\frac{\partial L}{\partial h_1} = \frac{\partial L}{\partial z}\frac{\partial z}{\partial h_2}\frac{\partial h_2}{\partial p_2} = \frac{\partial L}{\partial h_2}\cdot\frac{\partial h_2}{\partial p_2}$$



$$\frac{dh(x)}{dx} = \sum_{i=1}^{n} \frac{\partial f}{\partial g_i}(\vec{g}(\vec{x}))\cdot\frac{\partial g_i}{\partial x}(x)$$

Level set:
$L_\alpha(f) = \{\vec{x} \in \mathbb{R}^n \mid f(\vec{x}) = \alpha\}$
$L_\alpha(f) = \{\vec{x} \in \mathbb{R}^n \mid f(\vec{x}) \le \alpha\}$

Hessian: second derivative

$$\nabla^2 f(\vec{x}) = D(\nabla f)(\vec{x}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(\vec{x}) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1}(\vec{x}) \\ \vdots & & \\ \frac{\partial^2 f}{\partial x_1 \partial x_n}(\vec{x}) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(\vec{x}) \end{bmatrix}$$

Taylor approximations

$\hat{f}_1(x; x_0) = f(x_0) + \frac{df}{dx}(x_0)(x - x_0) = f(\vec{x}_0) + [\nabla f(\vec{x}_0)]^T(\vec{x} - \vec{x}_0)$

$f_2(x; x_0) = f(\vec{x}_0) + \frac{df}{dx}(x_0)(x - x_0) + \frac{1}{2}\frac{d^2 f}{dx^2}(x_0)(x - x_0)^2$

$= f(\vec{x}_0) + (\nabla f(\vec{x}_0))^T(x - x_0) + \frac{1}{2}(x - x_0)^T[\nabla^2 f(\vec{x}_0)](x - x_0)$

Main Theorem:

$\boxed{\min_{\vec{x} \in \Omega} f(\vec{x}) \text{ w/ soln } \vec{x}^* \to \nabla f(\vec{x}^*) = \vec{0}}$

to use this, must show $f(x)$ = convex

Perturbations

$A(\vec{x} + \delta\vec{x}) = \vec{y} + \delta\vec{y}$ $\quad \frac{\|\delta\vec{x}\|_2}{\|\vec{x}\|_2}$] how big an effect of δx compared to norm of $\vec{x}$

condition #: $K(A) = \frac{\sigma_1(A)}{\sigma_n(A)}$ $\quad$ Large $K(A)$ = perturbation have LARGE effect; small $K(A)$ = little effect

normal equation:

$A^T A\vec{x} = A^T\vec{y} \to K(A^T A) = \frac{\lambda_{max}(A^T A)}{\lambda_{min}(A^T A)}$

Ridge Regression: * use shift property of eigenvalues

$A^T A + \lambda I \longrightarrow$ better conditioned system

* solve $(A^T A + \lambda I)\vec{x} = A^T\vec{y}$

$A \in \mathbb{R}^{m \times n}$ $\quad y \in \mathbb{R}^m$ $\quad \lambda > 0$

$\boxed{\min_{\vec{x} \in \mathbb{R}^n}\{\|A\vec{x} - \vec{y}\|_2^2 + \lambda\|\vec{x}\|_2^2\} \to x^* = (A^T A + \lambda I)^{-1}A^T y}$

Principal components Regression: Ridge thru lens of SVD

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

$$= V\begin{bmatrix} \frac{\sigma_1\{A\}}{\sigma_1\{A\}^2 + \lambda} & & \\ & \ddots & \\ & & \frac{\sigma_n\{A\}}{\sigma_n\{A\}^2 + \lambda} \end{bmatrix} U^T y = \sum_{i=1}^{r} \frac{\sigma_i\{A\}}{\sigma_i\{A\}^2 + \lambda}(u_i^T y)v_i$$

Tikhonov regression: diff weights on rows

$W_1 W_2$ diagonal $\to \min_{\vec{x} \in \mathbb{R}^n}\{\|W_1(A\vec{x} - \vec{y})\|_2^2 + \|W_2(\vec{x} - \vec{x}_0)\|_2^2\}$

$x^* = (A^T W_1^2 A + W_2^2)^{-1}(A^T W_1^2 y + W_2^2 x_0)$

MLE: $\operatorname{argmax}_{\vec{x} \in \mathbb{R}^n} P_{\vec{x}}(\vec{y}) = \operatorname{argmin}_{\vec{x} \in \mathbb{R}^n} \|\Sigma_w^{-1/2}(A\vec{x} - \vec{y})\|_2^2$

CONVEXITY!!

convex combination: $\vec{x} = \sum_{i=1}^{k}\theta_i\vec{x}_i$ $\quad \theta_i \ge 0, \ \sum_{i=1}^{k}\theta_i = 1$

convex set: $\forall \vec{x}_1, \vec{x}_2 \in C$ $\ \theta \in [0,1]$: $\theta\vec{x}_1 + (1-\theta)\vec{x}_2 \in C$

* for every 2 points, line segment contained in C

convex hull: set of all convex combos of points in S

$\operatorname{conv}(S) = \{\sum_{i=1}^{k}\theta_i x_i \mid k \in \mathbb{N}, \theta_1 \cdots \theta_k \ge 0, \ \sum\theta_i = 1, \ x_1 \cdots x_k \in S\}$

$\operatorname{conv}(S) = \bigcup_{A \in S, |A| \le n+1} \operatorname{conv}(A)$

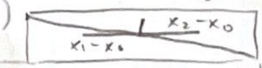
Hyperplane: $\{\vec{x} \in \mathbb{R}^n \mid a^T\vec{x} = b\} = \{\vec{x} \in \mathbb{R}^n \mid a^T(x - x_0) = 0\}$

Pos half space

$\{\vec{x} \in \mathbb{R}^n \mid a^T\vec{x} \ge b\}$ or $\{\vec{x} \in \mathbb{R}^n \mid a^T(x - x_0) \ge 0\}$

Neg half space

$\{\vec{x} \in \mathbb{R}^n \mid a^T\vec{x} \le b\}$ or $\{\vec{x} \in \mathbb{R}^n \mid a^T(x - x_0) \le 0\}$



$x_1 - x_0$ = obtuse = neg dot prod (H_-)

$x_2 - x_0$ = acute = pos dot prod (H_+)

* set of PSD matrices ∈ Convex

① A is PSD if $\vec{x}^T A\vec{x} \ge 0 \ \ \forall \vec{x} \in \mathbb{R}^n$

② check convex combos: $x^T(\theta A_1 + (1-\theta)A_2)\vec{x}$

$= \theta\underbrace{x^T A_1 x}_{\ge 0} + (1-\theta)\underbrace{\vec{x}^T A_2\vec{x}}_{\ge 0} \ge 0$ ✓

separating Hyperplane thm

$a^T(x - x_0) > 0 \ \ \forall x \in C$
$a^T(x - x_0) < 0 \ \ \forall x \in D$

convex functions

f is convex if

① dom(f) = Ω is convex

② $\forall \vec{x}_1, \vec{x}_2 \in \Omega$ $\ \theta \in [0,1]$

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

Jensen inequality: $f\left(\sum_{i=1}^{k}\theta_i\vec{x}_i\right) \le \sum_{i=1}^{k}\theta_i f(\vec{x}_i)$

Epigraph: all points in Ω that lie above f

$\operatorname{epi}(f) = \{\vec{x}, t) \mid \vec{x} \in \Omega, t \ge f(\vec{x})\}$

f is convex fn iff epi(f) is convex set

1st order: f convex iff $\forall \vec{x}, \vec{y} \in \Omega$: $f(\vec{y}) \ge f(\vec{x}) + [\nabla f(\vec{x})]^T(\vec{y} - \vec{x})$

2nd order: f convex iff $\nabla^2 f(x) \succeq 0$ PSD

$f(\vec{x}) = \underbrace{\vec{x}^T Q\vec{x} + b^T\vec{x} + c}_{\text{quadratic}}$ $\quad$ convex iff Q is PSD

strict convexity:
$f(\theta x_1 + (1-\theta)x_2) < \theta f(x_1) + (1-\theta)f(x_2)$: first + second order > and ≻ PD

Affine function (both convex & concave)

a) $f(\vec{x}) = a^T\vec{x} + b \to f(\alpha\vec{x} + (1-\alpha)) = \alpha f(x) + (1-\alpha)f(x)$

⑤ $f(A\vec{x} + b) \to \vec{f}(\alpha\vec{x} + (1-\alpha)\vec{y}) = \alpha\vec{f}(\vec{x}) + (1-\alpha)\vec{f}(\vec{y})$

© $tr(A^T X) + b \to f(\alpha x + (1-\alpha)y) = \alpha f(x) + (1-\alpha)f(y)$

$\min_{\vec{x} \in \Omega} f(\vec{x})$ convex iff Ω convex set & convex function

$\nabla f(\vec{x}) = \vec{0} \to \vec{x}^*$ = global minimizer

$\boxed{\text{GRADIENT DESCENT: } \vec{x}_{t+1} = \vec{x}_t - \eta\nabla f(x_0)}$

for least squares $\min_{\vec{x} \in \mathbb{R}^n}\|A\vec{x} - \vec{y}\|_2^2$

$\vec{x}_{t+1} - \vec{x}^* = (I - 2\eta A^T A)(\vec{x}_t - \vec{x}^*)$

$|\text{eig}(I - \eta(A^T A)| < 1 \to$ for convergence of least squares

Stochastic gradient descent:

$\vec{x}_{t+1} = \vec{x}_t - \eta\nabla f_i(\vec{x})$

* use decaying step size $\eta_t = \lim_{t \to \infty}\eta_t = 0$

Duality!

primal $p^* = \min_{\vec{x} \in \mathbb{R}^n} f_0(\vec{x})$

$f_i(\vec{x}) \le 0 \quad \forall i \in \{1 \cdots m\}$

$h_j(\vec{x}) = 0 \quad \forall j \in \{1 \cdots p\}$

Linear program

$p^* = \min_{\vec{x} \in \mathbb{R}^n} \vec{c}^T \vec{x}$

$A\vec{x} = \vec{y}$

$\vec{x} \ge \vec{0}$

convex

$d^* = \max_{\vec{\lambda} \in \mathbb{R}^n, \vec{\nu} \in \mathbb{R}^m} -\vec{y}^T \vec{\nu}$

$\vec{c} - \vec{\lambda} + A^T \vec{\nu} = \vec{0}$

$\vec{\lambda} \ge \vec{0}$

simplex algorithm

at least 1 optimal pt of LP is a vertex

* check vertices and compute value w/ obj function

Lagrangian:

$\mathcal{L}(\vec{x}, \vec{\lambda}, \vec{\nu}) = f_0(\vec{x}) + \sum_{i=1}^{m} \lambda_i f_i(\vec{x}) + \sum_{i=1}^{p} \nu_j h_j(\vec{x})$

$p^* = \min_{\vec{x} \in \mathbb{R}^n} \max_{\lambda \in \mathbb{R}^m, \nu \in \mathbb{R}^p} \mathcal{L}(\vec{x}, \vec{\lambda}, \vec{\nu})$

Dual function

$g(\vec{\lambda}, \vec{\nu}) = \min_{\vec{x} \in \mathbb{R}^n} \mathcal{L}(\vec{x}, \vec{\lambda}, \vec{\nu})$

Dual problem:

$d^* = \max_{\lambda \in \mathbb{R}^m, \nu \in \mathbb{R}^p} \min_{\vec{x} \in \mathbb{R}^n} \mathcal{L}(x, \lambda, \nu)$

$= \max_{\lambda \in \mathbb{R}^m, \nu \in \mathbb{R}^p} g(\lambda, \nu)$ st $\lambda_i \ge 0 \ \forall i = \{1 \cdots m\}$

ALWAYS CONVEX no matter what primal is

$f_0(\vec{x}) \ge p^*$ and $g(\vec{\lambda}, \vec{\nu}) \le d^*$

$f_0(\vec{x}) \ge d^*$ and $g(\vec{\lambda}, \vec{\nu}) \le p^*$

$p^* \ge d^* \rightarrow$ weak duality $\leftarrow$ ALWAYS HOLDS

$p^* = d^* \rightarrow$ strong duality

$p^* - d^*$ = duality gap

min max inequality

$\min_{x \in X} \max_{y \in Y} F(x, y) \ge \max_{y \in Y} \min_{x \in X} F(x, y)$

$p^* = \min_{x \in \mathbb{R}^n} \max_{\lambda \in \mathbb{R}^m, \nu \in \mathbb{R}^p} L(\vec{x}, \vec{\lambda}, \vec{\nu}) \ge \max_{\lambda \in \mathbb{R}^m, \nu \in \mathbb{R}^p} L(\vec{x}, \vec{\lambda}, \vec{\nu}) = d^*$

Quadratic Program

quadratic obj + affine constraints

$p^* = \min_{\vec{x} \in \mathbb{R}^n} \frac{1}{2} \vec{x}^T H \vec{x} + \vec{c}^T \vec{x}$

$A\vec{x} \le \vec{y}$

$C\vec{x} = \vec{z}$

CONVEX iff $H \in S_+^n$ (PD)

$\nabla^2(\frac{1}{2}\vec{x}^T H \vec{x} + \vec{c}^T \vec{x}) = H$

Quadratically constrained quadratic programs

$p^* = \min_{\vec{x} \in \mathbb{R}^n} \frac{1}{2} \vec{x}^T H \vec{x} + \vec{c}^T \vec{x}$

st $\frac{1}{2} \vec{x}^T P_i \vec{x} + b_i^T \vec{x} + c_i \le 0 \quad \forall i = \{1 \cdots m\}$

$\frac{1}{2} \vec{x}^T Q_i \vec{x} + d_i^T \vec{x} + f_i = 0 \quad \forall i = \{1 \cdots p\}$

CONVEX iff $Q_1 = \cdots = Q_p = 0$, $H, P_1 \cdots P_m \in S_+^n$ (PD)

Second order cone programs = CONVEX!! ALWAYS

$p^* = \min_{\vec{x} \in \mathbb{R}^n} \vec{c}^T \vec{x} \leftarrow$ linear obj

st : $\|A_i \vec{x} - \vec{y}_i\|_2 \le \vec{b}_i^T \vec{x} + z_i \quad \forall i \in \{1 \cdots n\}$

affine function of $\vec{x}$ in norm

REFORMULATIONS

$p^* = \min_{x \in \mathbb{R}^n} \sum_{i=1}^{m} \|A_i \vec{x} - \vec{y}_i\|_2 =$

$= \min_{\vec{x} \in \mathbb{R}^n, \vec{s} \in \mathbb{R}^m} \sum_{i=1}^{m} s_i$ st $\|A_i \vec{x} - y_i\|_2 \le s_i \quad \forall i \in \{1 \cdots m\}$

$p^* = \min_{x \in \mathbb{R}^n} \max_{i \in \{1 \cdots m\}} \|A_i \vec{x} - \vec{y}_i\|_2$

$= \min_{z \in \mathbb{R}^n, s \in \mathbb{R}} s$ st $\|A_i \vec{x} - \vec{y}_i\|_2 \le s \quad \forall i \in \{1 \cdots m\}$

$p^* = \min_{\vec{x}} \|A\vec{x} - \vec{y}\|_1$

$= \min_{\vec{x}, \vec{z}} \vec{z}^T \vec{1}$ st $z_i \ge a_i^T \vec{x} - y_i \quad i = \{1 \cdots m\}$

$z_i \ge -(a_i^T \vec{x} - y_i)$

STRONG DUALITY

SLATERS

$p^* = \min f_0(\vec{x})$

$f_i(\vec{x}) \le 0 \quad 1 \le i \le m$

$Ax = \vec{b}$ affine equality constraints

* if there exists point $\vec{x}_0 \in$ rel interior (domain) st $f_i(\vec{x}) < 0$ for all $1 \le i \le m \rightarrow$ strong duality holds

sufficient to have $\vec{x}_0$ st $f_1(\vec{x}_0) \le 0$ ... and $f_{k+1}(x_0) < 0$ ... $f_m(\vec{x}_0) < 0$

↓ if you have $Ax \le b$ we need strict feasibility for at least 1 inequality constraint $f_k(\vec{x}_0) < 0$

KKT condition (= set of conditions for OP

① Primal feasibility: $f_i(\vec{x}) \le 0 \quad \forall i \in \{1 \cdots m\}$

$h_j(\vec{x}) \le 0 \quad \forall j \in \{1 \cdots p\}$

② Dual feasibility $\lambda_i \ge 0 \quad \forall i \in \{1 \cdots m\}$

$\nu_i \ge 0 \quad \forall j \in \{1 \cdots p\}$

③ stationary : $\vec{0} = \nabla_x L(\vec{x}, \vec{\lambda}, \vec{\nu})$

④ complementary slackness: $\lambda_i f_i(\vec{x}) = 0 \quad \forall i \in \{1 \cdots m\}$

$\nu_j h_j(\vec{x}) = 0 \quad \forall j \in \{1 \cdots p\}$

If strong duality holds $\rightarrow$ KKT condns necessary for optimality

↳ If (x^*, λ^*, ν^*) fulfill KKT $\rightarrow$ P is convex

strong duality holds $\rightarrow (x^*, \lambda^*, \nu^*)$ optimal

L1 - L∞ duality:

$\max_{\vec{u} : \|\vec{u}\|_1 \le 1} \vec{u}^T \vec{z} = \|\vec{z}\|_\infty$

$\max_{\vec{u} : \|\vec{u}\|_\infty \le 1} \vec{u}^T \vec{z} = \|\vec{z}\|_1$

$p^* = \min_x \|A\vec{x} - \vec{y}\|_1 + \mu \|\vec{x}\|_\infty$

↓

$p^* = \min_x \vec{z}^T \vec{1} + \mu t$

st $t \ge x_j$, $t \ge -x_j \quad \forall j = 1 \cdots n$

$z_i \ge a_i^T x - y_i$, $z_i \ge -(a_i^T x - y_i) \quad i = 1 \cdots m$

Now look at dual

$\mathcal{L}(x, u, \nu) = u^T (Ax - y) + \nu^T x$

$g(u, \nu) = \min_x \mathcal{L}(x, u, \nu)$

$= \begin{cases} -u^T y & \text{if } A^T \vec{u} + \vec{\nu} = 0 \\ -\infty & \text{otherwise} \end{cases}$

$d^* = \max_{\vec{u}} -\vec{u}^T \vec{y}$

st $\|\vec{u}\|_\infty \le 1$

$\|A^T \vec{u}\|_1 \le \mu$

$\max_{\|\vec{u}\|_\infty \le 1} \{\vec{a}_i^T \vec{x}_0 + r \vec{a}_i^T \vec{u}\} \le b_i$

↓ based on $L_1 - L_\infty$ duality

$a_i^T x_0 + r\|a_i\|_1 \le b_i \quad \forall i$

STRATEGY

① show P is convex + differentiable

② show Slater's holds $\rightarrow$ strong duality

③ Find Lagrangian + gradient of Lagrangian

④ compute KKT

⑤ solve for optimal vars + check cases

LASSO REGRESSION

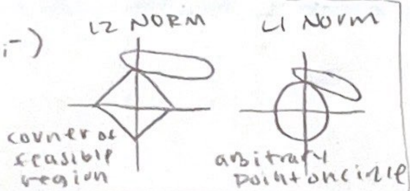
$$\min_{\vec{x}\in\mathbb{R}^n} \{ \underbrace{\|A\vec{x}-\vec{y}\|_2^2 + \lambda\|\vec{x}\|_1}_{f_0(\vec{x}) = \text{convex}} \}$$

$$x^*_{lasso} = \begin{cases} \frac{\vec{a}^T y - \lambda}{\|\vec{a}\|_2^2} & \text{if } \vec{a}^T\vec{y} > \lambda \\ \frac{\vec{a}^T\vec{y} + \lambda}{\|\vec{a}\|_2^2} & \text{if } \vec{a}^T\vec{y} < -\lambda \\ 0 & -\lambda \le \vec{a}^T\vec{y} \le \lambda \end{cases}$$

least L1 Norm: $\min_{\vec{x}\in\mathbb{R}^n} \|\vec{x}\|_2^2$ st $A\vec{x}=\vec{y}$

Formulate as LP

$$\min_{x^+, x^- \in\mathbb{R}^n} \sum_{i=1}^n (x_i^+ + x_i^-)$$
st $A(x^+ - x^-) = y$
$x^+ \ge \vec{0}$
$x^- \ge \vec{0}$



SVM

SVM = affine function $g_{\vec{w},b}$ $\vec{x} \mapsto \vec{w}^T\vec{x} - b$ which separates data into classes

if $y_i = 1 \to g_{\vec{w},b}(\vec{x}_i) > 0$
if $y_i = -1 \to g_{\vec{w},b}(\vec{x}_i) < 0$
$\to y_i \overbrace{g_{\vec{w},b}(\vec{x}_i)}^{f(x)} > 0$

HARD MARGIN SVM: $y_i f(x_i) > 0$ for all i

pick $\vec{w}, b$ w/ largest margin between hyperplane $H_{w,b}$ & closest point

solve $\max_{\vec{w}\in\mathbb{R}^d, b\in\mathbb{R}} \min_{i\in\{1\cdots n\}} \text{dist}(H_{\vec{w},b}, \vec{x}_i)$
st $y_i f(\vec{x}_i) > 0 \ \forall i \in \{1\cdots n\}$

* $\text{Dist}(H_{w,b}, \vec{x}) = \frac{|\vec{w}^T x - b|}{\|\vec{w}\|_2} = \frac{|f(\vec{x})|}{\|\vec{w}\|_2}$

* $(\vec{x}_i, y_i) =$ support vector if $\lambda_i > 0$

$$\max_{\vec{w}\in\mathbb{R}^d, b\in\mathbb{R}} \frac{1}{\|\vec{w}\|_2} \min_{i\in\{1\cdots n\}} |f(\vec{x}_i)|$$
st $y_i f(x_i) > 0 \ \forall i\{1\cdots n\}$

* apply slack vars and simplify

$$\boxed{\min_{\vec{w}\in\mathbb{R}^d, b\in\mathbb{R}} \frac{1}{2}\|\vec{w}\|_2^2 \quad \text{st } y_i(\vec{w}^T x_i - b) \ge 1 \ \forall i \in\{1\cdots n\}}$$

usually contribute to optimal soln and on/violate margin

$$\mathcal{L}(\vec{w}, b, \vec{\lambda}) = \frac{1}{2}\|\vec{w}\|_2^2 + \sum_{i=1}^n \lambda_i (1 - y_i(\vec{w}^T\vec{x}_i - b))$$
↳ apply KKT

SOFT MARGIN SVM

$l_{hinge}(z) = \begin{cases} 0 & z\le 0 \\ z & z>0 \end{cases}$

reformulate to unconstrained and add hinge loss to relax penalty

$$\min_{\vec{w}\in\mathbb{R}^d, b\in\mathbb{R}} \left[\frac{1}{2}\|\vec{w}\|_2^2 + C\sum_{i=1}^n \underbrace{l_{hinge}(1 - y_i g_{w,b}(\vec{x}_i))}_{\max((1-y_i g_{w,b}(x_i)), 0)} \right]$$

* add slack variable

$$\min_{\vec{w}\in\mathbb{R}^d, b\in\mathbb{R}, \xi\in\mathbb{R}^n} \frac{1}{2}\|\vec{w}\|_2^2 + C\sum_{i=1}^n \xi_i$$
st $\xi_i \ge 0 \quad \forall i = 1\cdots$
$\xi_i \ge 1 - y_i(w^T x_i - b)$
$\forall i \in\{1\cdots n\}$

$$\mathcal{L}(w, b, \xi, \lambda, \mu) = \frac{1}{2}\|\vec{w}\|_2^2 + C\sum_{i=1}^n \xi_i + \sum_{i=1}^n \mu_i(-\xi_i) + \sum_{i=1}^n \lambda_i(1 - y_i(\vec{w}^T\vec{x}_i - b) - \xi_i)$$

COORDINATE DESCENT

COORDINATE DESCENT: minimize $f(x)$ wrt each coordinate: using ith coordinate minimizer value to minimize subsequent coordinates

$$x_i^{(t+1)} \in \underset{x_i\in\mathbb{R}}{\text{argmin}}\ f(x^{(t+1)}_{1:i-1}, x_i, x^{(t)}_{i+1:n})$$

* converges for $f(\vec{x}) = g(\vec{x}) + \sum_{i=1}^n h_i(x_i)$
* compute partials $\frac{\partial f}{\partial x_i}(\vec{x}^*)$

NEWTONS METHOD

second order method for strictly convex w/ positive definite Hessians

UPDATE RULE: $\vec{x}^{(t+1)} = \vec{x}^{(t)} - [\nabla^2 f(\vec{x}^{(t)})]^{-1}[\nabla f(\vec{x}^{(t)})]$

Damped update rule: $\vec{x}^{(t+1)} = \vec{x}^{(t)} - \eta[\nabla^2 f(\vec{x}^{(t)})]^{-1}[\nabla f(\vec{x}^{(t)})]$

lineq constraints: solve
$$\begin{bmatrix} \nabla^2 f(\vec{x}^{(t)}) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \vec{v}^{(t)} \\ \vec{v} \end{bmatrix} = \begin{bmatrix} -\nabla f(\vec{x}^{(t)}) \\ 0 \end{bmatrix}$$
for $\vec{v}^{(t)}$

$x^{(t+1)} \leftarrow \vec{x}^{(t)} + \vec{v}^{(t)}$

$f_1(\vec{x}, \vec{x}_0) =$ first order approx

second order Taylor approx:
$$\hat{f}_2(\vec{x}, \vec{x}^{(t)}) = f(\vec{x}^{(t)}) + [\nabla f(\vec{x}^{(t)})]^T(x - \vec{x}^{(t)}) + \frac{1}{2}(\vec{x} - \vec{x}^{(t)})^T[\nabla^2 f(\vec{x}^{(t)})](\vec{x} - \vec{x}^{(t)})$$

* when diagonal entries of diagonal matrix are non zero → full rank

* max of 2 convex functions also convex
* vectors in range A must be orthogonal to null space

$p^* = \min_{x\in\mathbb{R}^3} (\vec{x}^T M\vec{x} - 2\vec{b}^T\vec{x}) \to$ if M not full rank and $\vec{b}$ not in range of $M \to p^* = -\infty$
and $\vec{b}$ IN range of $M \to p^* =$ finite

* when converting $\max(a, 0)$ to slack var → $t \ge 0$, $t \ge a$ (greater than quantities maxxing)

P is polyhedron: $P = \{\vec{x} : \vec{a}_i^T\vec{x} \le b_i \forall i = 1\cdots m\}$

Euclidean ball: $B(\vec{x}_0, R) = \{\vec{x}\in\mathbb{R}^n : \|\vec{x} - \vec{x}_0\|_2 \le R\}$
$x = x_0 + u$ where $\|\vec{u}\|_2 \le R$

↳ largest ball $\max_{\|\vec{u}\|_2 \le R} \underline{(a_i^T(\vec{x}_0 + \vec{u}))} \le b_i$
max this with $a_i^T\left(\frac{R}{\|\vec{a}_i\|_2}\vec{a}_i\right) = R\|\vec{a}_i\|_2$
↳ condn: $a_i^T x_0 + R\|a_i\|_2 \le b_i \ \forall i = 1\cdots m$

$$\min_{x_0, R} -R$$
st $a_i^T x_0 + R\|a_i\|_2 \le b_i$