Constrained optimization.

$$\min f(x) \quad \text{s.t. } h_i(x)=0 \quad i=1,2\cdots m$$

equality constraints.

1st order necessary conditions

define $L(\vec{x},\vec{\lambda}) = f(\vec{x}) + \sum_{i=1}^{m} \lambda_i h_i(\vec{x})$ $\quad \vec{x}\in \mathbb{R}^n \quad \vec{\lambda} \in \mathbb{R}^m$

: Lagrange multiplies

if x^* is a min, then

$\nabla f(x^*, \lambda^*) = 0$ for some λ^*

$$\nabla_x L = \left[\frac{\partial L}{\partial x_1} \cdots \frac{\partial L}{\partial x_n}\right]$$

$$= \nabla f(\vec{x}) - \Sigma \lambda_i \nabla h_i(\vec{x})$$

$$\nabla_\lambda L = -[h_1(x), \cdots, h_m(x)]$$

2nd order necessary conditions

if x^* is a local min, then

$\nabla L(x^*, \lambda^*) = 0.$

and $y^T \nabla_{xx} L(x^*, \lambda^*) y \geq 0$ for $y \in \mathbb{R}^n$ satisfying $y \neq 0$ and

$\nabla h_i(x^*) y = 0, \quad i=1, \cdots m$

Ex $\quad \min f(x) = x_1^2 + x_2^2 \quad$ s.t. $h(x) = x_1 + x_2 - 1 = 0.$

$L(x,\lambda) = f(x) + \lambda h(x) = x_1^2 + x_2^2 + \lambda(x_1 + x_2 - 1).$

$\nabla L = (2x_1 + \lambda, 2x_2 + \lambda, x_1 + x_2 - 1) = 0$

$$\begin{cases} 2x_1 + \lambda = 0 \\ 2x_2 + \lambda = 0 \\ x_1 + x_2 - 1 = 0 \end{cases} \Rightarrow \begin{cases} x_1 = \frac{1}{2} \\ x_2 = \frac{1}{2} \\ \lambda = -1 \end{cases}$$

Ex: $\min f(x) = -x_1x_2 - x_2x_3 - x_1x_3$

s.t. $x_1 + x_2 + x_3 - 3 = 0$

$L(x,\lambda) = f(x) + \lambda h(x)$

$\nabla_x L = \nabla f(x) + \lambda \nabla h(x) = 0$

$$\begin{cases} -x_2 - x_3 + \lambda = 0 \\ -x_1 - x_3 + \lambda = 0 \\ -x_2 - x_1 + \lambda = 0 \\ x_1 + x_2 + x_3 - 3 = 0 \end{cases} \Rightarrow \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \lambda \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 2 \end{bmatrix} = x^*$$

$$\nabla^2_{xx} f(x^*) = \begin{bmatrix} 0 & -1 & -1 \\ -1 & 0 & -1 \\ -1 & -1 & 0 \end{bmatrix}$$

$\forall y \neq 0$, s.t. $\nabla h(x) y = 0$

$[1, 1, 1] \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = 0 \Rightarrow y_1 + y_2 + y_3 = 0 \qquad y^T \nabla^2_{xx} f y = y_1^2 + y_2^2 + y_3^2 \geq 0.$

Ex: $\min f(x) = x_1^2 + x_2^2 + x_3^2$

s.t. $h_1(x) = x_1 + x_2 + 3x_3 - 2 = 0$

$h_2(x) = 5x_1 + 2x_2 + x_3 - 5 = 0$

$L(x, \lambda) = f(x) + \lambda_1 h_1(x) + \lambda_2 h_2(x)$

$\nabla_x L = \nabla f + \lambda_1 \nabla h_1(x) + \lambda_2 \nabla h_2(x)$

$= [2x_1, 2x_2, 2x_3] + \lambda_1 [1, 1, 3] + \lambda_2 [5, 2, 1].$

Equality & inequality Constrained optimization

$$\min f(x)$$

s.t. $h_i(x)=0 \quad i=1,\cdots m$

& $g_j(x)\le 0 \quad j=1,\cdots,r \quad x\in\mathbb{R}^n$

necessary condition

if x^* is a local min, then $\exists$ constants

λ_i, $i=1,\cdots m$.

$\mu_j \ge 0$, $j=1,\cdots r$.

s.t. $\nabla f(x^*) + \sum_{i=1}^{m}\lambda_i \nabla h_i(x^*) + \sum_{j=1}^{r}\mu_j \nabla g_j(x^*)$

and $\mu_j^* g_j(x^*)=0$, $\forall j=1,\cdots r$.

KKT conditions.

x^*: stationary point

(KKT point)

λ_j^*, μ_j^*: KKT multiplies.

Lagrange function

$$L(x,\lambda,\mu) = f(x) + \sum_{i=1}^{m}\lambda_i h_i(x) + \sum_{j=1}^{r}\mu_j g_j(x)$$

KKT conditions.

$$\begin{cases} \nabla_x L(x^*,\lambda^*,\mu^*)=0 \\ h_i(x^*)=0 \\ \mu_j^* g_j(x^*)=0. \end{cases}$$

2nd order neccessary condition

$$y^T \nabla_{xx} L(x^*,\lambda^*,\mu^*)\, y \ge 0$$

for all y s.t.

$$\begin{cases} y\neq 0 \\ \nabla h_i(x^*)\, y=0 \quad i=1,\cdots m \\ \nabla g_j(x^*)\, y=0 \quad j\in A(x^*). \end{cases}$$

$A(x^*)\ \{j: g_j(x^*)=0\}$.

if $g_j(x^*)=0$, then constraint $g_j(x^*)\le 0$ is called active (binding).

The above condition become sufficient if

① $y^T \nabla_{xx} L(x^*, \lambda^*, \mu^*) y > 0$ for all y s.t.

$$\begin{cases} y \neq 0 \\ \nabla h_i(x^*) y = 0 & i = 1, \dots m \\ \nabla g_j(x^*) y = 0, & j \in A(x^*) \end{cases}$$

② $\mu_j^* > 0 \quad j \in A(x^*)$

Ex: find the local min of

$$\min f(x) = x_1^2 + x_2^2$$

$$\text{s.t.} \quad g(x) = -x_1 - x_2 + 1 \leq 0.$$

$$L(x, \lambda, \mu) = f(x) + \mu g(x)$$

$$= (x_1^2 + x_2^2) + \mu(-x_1 - x_2 + 1).$$

$$\nabla_x L = \begin{bmatrix} 2x_1 - \mu \\ 2x_2 - \mu \end{bmatrix} = 0$$

$$\mu \geq 0$$

$$\mu g(x) = \mu(-x_1 - x_2 + 1) = 0$$

Case 1 $\quad -x_1 - x_2 + 1 < 0$, Then μ must be 0, Then $\begin{cases} 2x_1 = 0 \\ 2x_2 = 0 \end{cases} \Rightarrow \begin{cases} x_1 = 0 \\ x_2 = 0 \end{cases}$

however, $-0-0+1 \not< 0$, so

Case 2: $\quad -x_1 - x_2 + 1 = 0$, $\quad -\frac{\mu}{2} - \frac{\mu}{2} + 1 = 0$, $\mu = 1 > 0$

$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}$ is a KKT point

$\nabla_{xx} L = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$ is positive definite. So $y^T \nabla_{xx} L \, y \geq 0$

$\forall y \neq 0$ and $g(x^*) y = 0$, if g is active at x^*

$g(x_1^*, x_2^*) = 0$, so g is active at x^*, so x^* satisfies 2nd order necessary condition.

① $y^T \nabla_{xx} L(x^*, \lambda^*, \mu^*) y = 2y_1^2 + 2y_2^2 > 0$

$\forall y \neq 0$, $\nabla g(x^*) y = 0$.

② $\mu > 0$.

x^* satisfies 2nd order sufficient condition.

So x^* is a strict local min.

Ex: Solve

$$\min f(x) = x_1^3 + x_1^2 \quad \text{s.t.} \quad g(x) = -x_1 - 1 \leq 0$$

$L(x,\lambda,\mu) = x_1^3 + x_1^2 + \mu(-x_1-1)$

$\nabla_x L = \begin{bmatrix} 3x_1^2-\mu \\ 2x_2 \end{bmatrix}^T \quad \nabla_{xx} L = \begin{bmatrix} 6x_1 & 0 \\ 0 & 2 \end{bmatrix}$

kkT condition

$$\begin{cases} 3x_1^2-\mu=0 \\ 2x_2=0 \\ \mu(-x_1-1)=0 \end{cases}$$

case 1: $-x_1-1<0$, then $\mu=0 \Rightarrow x_1=0,\ x_2=0 \quad -x_1-1<0$

so $x^*=(0,0)$ is a kkT point $\quad g(x^*)=-1<0$ isn't active.

$\nabla_{xx} L(x^*,\lambda^*,\mu^*) = \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix}$ isn't positive definite, however $y^T \nabla_{xx} L\, y = 2y_2^2 \geq 0$, $\forall y \neq 0$

So 2nd order necessary condition satisfied

$\forall y \neq 0$

① $y^T \nabla_{xx} L(x^*,\lambda^*,\mu^*) y = 2y_2^2 > 0$

$\forall y \neq 0$

2nd order sufficient condition not satisfied

no conclusion is made.

$$\begin{cases} f(x) = x_1^3 + x_2^2 \\ g(x) = -x_1 - 1 \\ x^* = (0,0) \end{cases}$$

let $x=(-\varepsilon,0)$ $\quad f(x)<f(x^*)$

So x^* isn't a local minium.

Case 2: $-x_1-1=0$

$$\begin{cases} 3x_1^2-\mu=0 \\ 2x_2=0 \\ -x_1-1=0 \end{cases} \Rightarrow \begin{cases} x_1=-1 \\ x_2=0 \\ \mu=3 \end{cases}$$

$x^*=(-1,0)$ is a kkT point.

$g(x^*)=0$, g is active at x^*.

$$(*)\begin{cases} y \neq 0 \\ \nabla g(x^*) y = 0 \Rightarrow -y_1=0 \Rightarrow y_1=0 \end{cases}$$

$y^T \nabla_{xx} L(x^*,\lambda^*,\mu^*) y = -6y_1^2 + 2y_2^2 > 0$ for every y satisfies condition $(*)$.

So 2nd order necessary condition is satisfied and part ① of sufficient also satisfied.

② $\mu_j > 0,\ \forall j \in A(x^*)$ ✓

sufficient cond satisfied

so $(-1, 0)$ is a strict local min.

Ex: Solve

$$\min f(x) = x_1^2 + x_2^2$$

$$\text{s.t.}\quad g_1(x) = -x_1 - x_2 + 1 \le 0$$

$$g_2(x) = -x_1 + 2 \le 0$$

$$L(x, \lambda, \mu) = x_1^2 + x_2^2 + \mu_1(-x_1 - x_2 + 1) + \mu_2(-x + 2)$$

$$\nabla_x L = \begin{bmatrix} 2x_1 - \mu_1 - \mu_2 \\ 2x_2 - \mu_2 \end{bmatrix}^T$$

$$\nabla_{xx} L = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

KKT conditions.

$$\begin{cases} 2x_1 - \mu_1 - \mu_2 = 0 \\ 2x_2 - \mu_1 = 0 \\ \mu_1(-x_1 - x_2 + 1) = 0 \\ \mu_2(-x_1 + 2) = 0 \end{cases}$$

Case 1: $\begin{cases} -x_1 - x_2 + 1 < 0 \\ -x_1 + 2 < 0 \end{cases}$

$\Rightarrow \mu_1 = 0,\ \mu_2 = 0$

$\Rightarrow x_1 = 0,\ x_2 = 0$

$x^* = (0, 0)$

$g(x^*) = 1 \nless 0$ rejected

Case 2: $\begin{cases} -x_1 - x_2 + 1 = 0 \\ -x_1 + 2 < 0 \end{cases}$

$\Rightarrow \mu_2 = 0 \Rightarrow \begin{cases} 2x_1 - \mu_1 = 0 \\ 2x_2 - \mu_1 = 0 \\ -x_1 - x_2 + 1 = 0 \end{cases} \Rightarrow \begin{array}{l} x_1 = x_2 = \frac{1}{2} \\ \mu_1 = 1 \end{array}$

$x^* = (\frac{1}{2}, \frac{1}{2})$ $g_2(x^*) \nless 0$, rejected.

Case 3. $\begin{cases} -x_1 - x_2 + 1 < 0 \\ -x_1 + 2 = 0 \end{cases}$

$\Rightarrow \mu_1 = 0$

$\begin{cases} 2x_1 - \mu_2 = 0 \\ 2x_2 = 0 \\ -x_1 + 2 = 0 \end{cases} \Rightarrow \begin{cases} x_1 = 2 \\ x_2 = 0 \\ \mu_2 = 4 \end{cases}$ so $x^* = (2, 0)$ is a KKT point

Since $\nabla_{xx} L = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$ is SPD, 2nd necessary cond are satisfied

① $y^T \nabla_{xx} L \, y > 0, \ \forall y \neq 0$

② $g_1(x^*) \neq 0$, not active

$g_2(x^*) = 0$, active

$\mu_2 = 4 > 0$

So 2nd order sufficient conds are Satisfied

So $x^* = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$ is a strict local min

case 4: $\begin{cases} -x_1 - x_2 + 1 = 0 \\ -x_1 + 2 = 0 \end{cases}$

$$\begin{cases} 2x_1 - \mu_1 - \mu_2 = 0 \\ 2x_2 - \mu_1 = 0 \\ -x_1 - x_2 + 1 = 0 \\ -x_1 + 2 = 0 \end{cases} \Rightarrow \begin{cases} x_1 = 2 \\ x_2 = -1 \\ \mu_1 = -2 \\ \mu_2 = 6 \end{cases}$$

$\mu_1 < 0$. rejected.

Barrier Method.

$\min f(x)$

$x \in S.$

S: feasible set

$x \in S$: feasible point

let $\delta(x) = \begin{cases} 0 & x \in S \\ \infty & x \notin S \end{cases}$

$\min f(x) \longrightarrow \min f(x) + \delta(S)$

$\min f(x) \quad \text{s.t. } g_i(x) \geq 0, \; i=1,\cdots,m.$

$S = \{x \in \mathbb{R}^n : g_i(x) \geq 0, \; i=1,\cdots m\}.$

$S^\circ = \{x \in \mathbb{R}^n : g_i(x) > 0, \; i=1,\cdots m\}, \neq \phi.$

$\phi(x) = \sum_{i=1}^{m} \log g_i(x).$

$\min f(x) + \mu \phi(x).$

Barrier term. (μ: barrier parameter).

let $\boxed{\beta(x,\mu) = f(x) + \boxed{\mu \phi(x)}}$ Barrier function.

$\min f(x) \quad \text{s.t. } g_i(x) \geq 0 \quad i=1,\cdots m$

$\Downarrow$

$\min \beta(x,\mu). \quad \mu_k \to 0.$

Logrithm barrier function : $-\sum \log g_i(x).$

inverse barrier function : $\sum \frac{1}{g_i(x)}$

Ex: $\min f(x)$ $\quad$ $x \in [1,4].$ $\Rightarrow$ $\min f(x)$, $g_1(x) = x-1 \geq 0$, $g_4(x) = 4-x \geq 0$

$-\mu(\log(x-1) + \log(4-x)).$

$\min f(x) - \mu_k [\log(x-1) + \log(4-x)].$

$\mu_0 = 1 \Rightarrow x_0$

Ex: solve

$$\min f(x) = x_1 - 2x_2$$

$$\text{s.t.} \quad 1 + x_1 - x_2^2 \geq 0$$

$$x_2 \geq 0$$

(logrithm)

Barrier function

$$\beta(\vec{x}, \mu) = f(\vec{x}) - \mu \sum \log g_i(\vec{x})$$

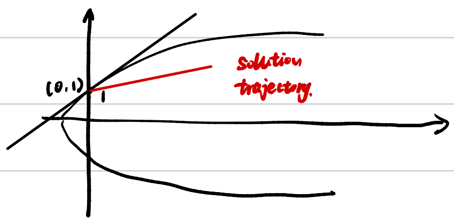
$$= x_1 - 2x_2 - \mu[\log(x_1 + 1 - x_2^2) + \log x_2]$$

for find μ

$$\nabla_{\vec{x}} \beta(\vec{x}, \mu) = \left[1 - \mu \frac{1}{1 + x_1 - x_2^2},\ -2 - \mu\left[\frac{-2x_2}{1 + x_1 - x_2^2} + \frac{1}{x_2}\right].\right]$$

let $\nabla = 0 \Rightarrow \begin{cases} x_1 = \frac{\sqrt{1+2\mu} + 3\mu - 1}{2} \\ x_2 = \frac{1 + \sqrt{1+2\mu}}{2} \end{cases}$

Verify. $[x_1, x_2]$ is a local min. let $\mu \to 0$ $x_1 \to 0$, $x_2 \to 1$, $[0, 1]$ is a local min.



$\min f(x)$ s.t $g_i(x) \geq 0$

$$\beta(\vec{x}, \mu) = f(x) - \mu \sum \log g_i(x)$$

$$\nabla_x \beta(\vec{x}, \mu) = \nabla f(x) - \mu \sum \frac{\nabla g_i(\vec{x})}{g_i(\vec{x})}$$

$$= \nabla f(\vec{x}) - \sum \frac{\mu}{g_i(\vec{x})} \nabla g_i(\vec{x}).$$

$$L(\vec{x}, \vec{\lambda}, \vec{\mu}) = f(\vec{x}) + \sum \mu_i g_i(\vec{x})$$

$$\nabla_{\vec{x}} L(\vec{x}, \vec{\lambda}, \vec{\mu}) = \nabla f(\vec{x}) + \sum \mu_i \nabla g_i(\vec{x}).$$

$$\lambda_i(\vec{\mu}) = \frac{\mu}{g_i(\vec{x}(\mu))}.$$

estimate of the KKT multiplies

Ex: $\min f(x) = x_1^2 + x_2^2$

s.t $x_1 - 1 \geq 0$

$x_2 + 1 \geq 0$

barrier function

$B(\vec{x},\mu) = x_1^2 + x_2^2 - \mu \log(x_1-1) - \mu \log(x_2+1)$

$$\begin{cases} 2x_1 - \frac{\mu}{x_1-1} = 0 \\ 2x_2 - \frac{\mu}{x_2-1} = 0 \end{cases} \Rightarrow \begin{cases} x_1 = \frac{1+\sqrt{1+2\mu}}{2} \\ x_2 = \frac{-1+\sqrt{1+2\mu}}{2} \end{cases}$$

as $\mu \to 0$, $x_1 \to 1$, $x_2 \to 0$. $\vec{\lambda}^* = [1,0]^T$

$$\lambda_i(\mu) = \frac{\mu}{g_i(\vec{x}(\mu))}$$

$$\lambda_1(\mu) = \frac{\mu}{x_1(\mu)-1} = \frac{2\mu}{-1+\sqrt{1+2\mu}} \to 2$$

$$\lambda_2(\mu) = \frac{2\mu}{1+\sqrt{1+2\mu}} \to 0$$

Penalty methods

$\min f(x)$

s.t. $g_i(x) = 0$, $i=1,\cdots,m$.

Penalty term

$$\psi(x) \begin{cases} =0 & x \text{ is feasible} \\ >0 & \text{otherwise} \end{cases}$$

$$\min \underbrace{f(x) + \rho\,\psi(x)}_{\text{penalty function}} \quad (\rho > 0)$$

$\rho\psi(x)$: penalty term

eg. quadratic loss

$$\psi(x) = \frac{1}{2}\sum_{i=1}^{m} g_i^2(x)$$

$$\min_{x\in\mathbb{R}^n} f(x) + \frac{\rho}{2}\sum g_i^2(x)$$

Ex: $\min f(x) = -x_1 x_2$

s.t. $g(x) = x_1 + 2x_2 - 4 = 0$

let $\pi(x,\rho) = f(x) + \frac{\rho}{2} g^2(x)$

fixed ρ

$$\nabla_x \pi(x,\rho) = \nabla f(x) + \rho g \nabla g$$

$$= \begin{bmatrix} -x_2 \\ -x_1 \end{bmatrix}^T + \rho(x_1+2x_2-4)\begin{bmatrix} 1 \\ 2 \end{bmatrix}^T$$

$$= \begin{bmatrix} -x_2 + \rho(x_1+2x_2-4) \\ -x_1 + 2\rho(x_1+2x_2-4) \end{bmatrix} = 0$$

$\Rightarrow x_1 = 2x_2 \Rightarrow -2x_2 + 2\rho(4x_2-4) = 0$

$$\begin{cases} x_1 = \frac{8\rho}{4\rho-1} \\ x_2 = \frac{4\rho}{4\rho-1} \end{cases}$$

$$g(\vec{x}(\rho)) = x_1(\rho) + 2x_2(\rho) - 4$$

$$= \frac{16\rho}{4\rho-1} - 4$$

$$= \frac{4}{4\rho-1} \neq 0$$

let $\rho \to \infty$

$$\begin{cases} x_1 \to 2 \\ x_2 \to 1 \end{cases}$$

$$x^* = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

in practice, take $\rho_k \to \infty$

$$\min \pi(x,\rho_k)$$
$$\Downarrow$$
$$x_k$$
$$\Downarrow$$
$$\min \pi(x,\rho_{k+1}).$$

$\pi(x,\rho) = f(x) + \frac{\rho}{2}\sum_{i=1}^{m} g_i^2(x)$

$\nabla_x \pi(x,\rho) = \nabla f(x) + \rho \sum g_i(x)\nabla g_i(x)$

$= \nabla f(x) + \sum \rho g_i(x) \nabla g_i(x)$

$\lambda_i = \rho g_i(x(\rho))$

$\lim_{\rho\to\infty} \lambda_i(\rho) = \lambda_i^*$

in this case, $\lambda = \rho g(x(\rho))$

$= \frac{4\rho}{4\rho-1} \to 1 =: \lambda^*$

$\min f(x)$

s.t. $g_i(x) \geq 0$, $i=1,\cdots,m$

Let $\varphi(x) = \frac{1}{2}\sum_{i=1}^{m}[\min(g_i(x),0)]^2$
$\begin{cases} =0, & x \text{ is feasible} \\ >0, & \text{otherwise} \end{cases}$

$\pi(x,\rho) = f(x) + \frac{\rho}{2}\sum_{i=1}^{m}[\min g(x),0]^2$

$\nabla \pi(x,\rho) = \nabla f(x) + \rho\sum_{j=1}^{m} \min(g(x),0)\ \nabla g_i(x)$

$\min f(x)$

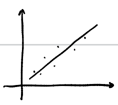
s.t. $h_i(x)=0$, $i=1,\cdots,r$

$g_j(x) \geq 0$, $j=1,\cdots,m$

$\min \pi(x,\rho)$

$\pi(x,\rho) = f(x) + \frac{\rho}{2}\left\{\sum_{i=1}^{r} h_i^2(x) + \sum_{j=1}^{m}[\min g(x),0]^2\right\}$.

Nonlinear least-square problems.

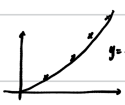


(x_i, y_i) $i=1,\cdots,m$

$y = ax+b$ $\quad \sum_{i=1}^{m}[(ax_i+b)-y_i]^2 := f(a,b)$ $\quad \min f(a,b)$.

general:

$\min_w f(w) = \sum_{i=1}^{m} r_i^2(w)$



$y = ae^{bx}$

$f(a,b) = \frac{1}{2}\sum_{i=1}^{m}(ae^{bx_i}-y_i)^2$

$= \frac{1}{2}\sum_{i=1}^{m} r_i^2(w)$ $\quad r_i(w) = ae^{bx_i}-y_i$ $\quad w=\begin{bmatrix} a \\ b \end{bmatrix}$

(not in x!)
r_i is non-linear in $w=\begin{bmatrix} a \\ b \end{bmatrix}$, (non-linear least square problems)

$\min \frac{1}{2}\sum r_i^2(w) = f(w)$ $\quad w \in \mathbb{R}^n$

$\nabla f(w) = \sum r_i(w)\nabla r_i(w)$

let $R(w) = \begin{bmatrix} r_1(w) \\ r_2(w) \\ \vdots \\ r_m(w) \end{bmatrix}$

$J = \nabla R(w) = \begin{bmatrix} \frac{\partial r_1}{\partial w_1} & \cdots & \frac{\partial r_1}{\partial w_n} \\ \vdots & & \\ \frac{\partial r_m}{\partial w_1} & \cdots & \frac{\partial r_m}{\partial w_n} \end{bmatrix}$: $m\times n$

$\nabla f = J^T R \qquad f = \frac{1}{2}\sum_{i=1}^{m} r_i^2(w) \qquad \nabla f = \sum_{i=1}^{m} r_i(w)\nabla r_i(w)$

$$r(w)\nabla r(w) = \begin{bmatrix} r\frac{\partial r}{\partial w_1} \\ \vdots \\ r\frac{\partial r}{\partial w_n} \end{bmatrix}^T$$

$$\nabla[r\nabla r(w)] = \begin{bmatrix} \frac{\partial r}{\partial w_1}\frac{\partial r}{\partial w_1} + r\frac{\partial^2 r}{\partial w_1^2} \\ \frac{\partial r}{\partial w_1}\frac{\partial r}{\partial w_2} + r\frac{\partial^2 r}{\partial w_1 \partial w_2} \end{bmatrix}$$

$$= [\nabla r]^T \nabla r + r\nabla^2 r$$

$$\nabla^2 f = \sum_{i=1}^{m} (\nabla r_i)^T \nabla r_i + r_i \nabla^2 r_i$$

$$= \sum_{i=1}^{m} \begin{bmatrix} \frac{\partial r_i}{\partial w_1} \\ \vdots \\ \frac{\partial r_i}{\partial w_n} \end{bmatrix} \left[\frac{\partial r_i}{\partial w_1} \cdots \frac{\partial r_i}{\partial w_n}\right] + \sum_{i=1}^{m} r_i \nabla^2 r_i$$

$$= J^T J + \sum_{i=1}^{m} r_i \nabla r_i^2$$

fit (x_i, y_i) by $l(w) = w_1 + w_2 x$

$$f(w) = \sum_{i=1}^{n} (w_1 + w_2 x_i - y_i)^2$$

$$= \sum_{i=1}^{n} r_i^2(w), \quad r_i = w_1 + w_2 x_i - y_i$$

$$R = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix} = \begin{bmatrix} w_1 + w_2 x_1 - y_1 \\ \vdots \\ w_1 + w_2 x_n - y_n \end{bmatrix}$$

$$J = \nabla R \qquad \nabla f = J^T R = \begin{bmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} r_1 \\ \vdots \\ r_n \end{bmatrix}$$

$$= \begin{bmatrix} \Sigma r_i \\ \Sigma x_i r_i \end{bmatrix} = 0$$

$$R = J\vec{w} - \vec{b}$$

$$J^T J \vec{w} = J^T \vec{b}$$

let $f(w) = \frac{1}{2}\sum_{i=1}^{m} r_i^2(w)$

$$\nabla f = \Sigma r_i \nabla r_i$$

$$R = \begin{bmatrix} r_1 \\ \vdots \\ r_m \end{bmatrix}$$

$$J = \begin{bmatrix} \nabla r_1 \\ \vdots \\ \nabla r_m \end{bmatrix}$$

$$\frac{d}{dw} \cdot \frac{1}{2} r(w)^2 = r \cdot r'$$

$$\frac{d^2}{dw^2} \frac{1}{2} r(w)^2 = r' \cdot r' + r \cdot r''$$

$$\nabla^2 f = J^T J + \sum_{i=1}^{m} r_i \nabla^2 r_i$$

linear least square problems

$$r_i(w) = a_{i1} w_1 + a_{i2} w_2 + \cdots a_{in} W_n - b_i$$

$$R = Aw - b$$

$$J = A$$

$$\nabla f = J^T R = A^T(Aw - b) = A^T A w - A^T b = 0$$

$A^T A w = A^T b$: normal equation

$$\nabla^2 f = A^T A \geq 0$$

nonlinear least squares

Newton's method

$x_{k+1} = x_k - [\nabla^2 f(x_k)]^{-1} [\nabla f(x_k)]$

$\nabla f(w_k) = J^T(w_k) R(w_k)$

$\nabla^2 f(w_k) = J^T(w_k) J(w_k) + \sum_{i=1}^{m} r_i(w_k) \nabla^2 r_i(w_k)$

Gauss-Newton method

$\nabla^2 f = J^T R$ $\qquad\qquad$ $\nabla^2 f \approx J^T J$

$r_i / \nabla^2 r_i$ is small: Small residence case

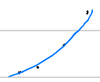
① $k=0$, w_0

② $p_k = -(J^T(w_k) J(w_k))^{-1} J^T(w_k) R(w_k)$.

③ $x_{k+1} = x_k + \alpha_k p_k$

find α_k by line search

Ex: $(x_1, y_1) \cdots (x_n, y_n)$



fit by $y = w_1 e^{w_2 t}$

$\vec{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$

$f(w) = \sum_{i=1}^{n} (w_1 e^{w_2 x_i} - y_i)^2 = \sum_{i=1}^{n} r_i(\vec{w})^2$, where $r_i(\vec{w}) = w_1 e^{w_2 x_i} - y_i$

$R = \begin{bmatrix} r_1 \\ \vdots \\ r_n \end{bmatrix} = \begin{bmatrix} w_1 e^{w_2 x_1} - y_1 \\ \vdots \\ w_1 e^{w_2 x_n} - y_n \end{bmatrix}$

$\nabla r_i(\vec{w}) = [e^{w_2 x_i}, x_i w_1 e^{w_2 x_i}]$

$\nabla^2 r_i(\vec{w}) = \begin{bmatrix} 0 & x_i e^{w_2 x_i} \\ x_i e^{w_2 x_i} & x_i^2 w_1 e^{w_2 x_i} \end{bmatrix}$

$J = \begin{bmatrix} \nabla r_1 \\ \vdots \\ \nabla r_n \end{bmatrix} = \begin{bmatrix} e^{w_2 x_1} & x_1 w_1 e^{w_2 x_1} \\ \vdots & \vdots \\ e^{w_2 x_n} & x_n w_1 e^{w_2 x_n} \end{bmatrix}$

$\nabla f = J^T R = \begin{bmatrix} e^{w_2 x_1} & \cdots & e^{w_2 x_n} \\ x_1 w_1 e^{w_2 x_1} & \cdots & x_n w_1 e^{w_2 x_n} \end{bmatrix} \begin{bmatrix} w_1 e^{w_2 x_1} - y_1 \\ \vdots \\ w_1 e^{w_2 x_n} - y_n \end{bmatrix}$

$= \begin{bmatrix} \sum_{i=1}^{n} e^{w_2 x_i} (w_1 e^{w_2 x_i} - y_i) \\ \sum_{i=1}^{n} x_i w_1 e^{w_2 x_i} (w_1 e^{w_2 x_i} - y_i) \end{bmatrix}$

$\nabla^2 f \approx J^T J$

Levenberg-Marquardt method (*)

approximate $\nabla^2 f$ by $J^T J + \beta I$, $\beta > 0$

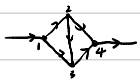
$[J^T(w_k) J(w_k) + \beta_k I] p_k = -J^T(w_k) R(w_k)$

Structed secant method

$\nabla^2 f = J^T J + \sum r_i \nabla^2 r_i$

approximate $\nabla^2 r_i$ by B satisfying a secant condition

Vehicle routing



integer programing / discrete optimization

min total travel time. $\sum_{ij} X_{ij} T(X_{ij})$ $ij = 12, 13, 23, 24, 34$

X_{ij} : # cars on the road

$T(X_{ij})$: travel time of each car on ij

$T(X_{ij}) = t_{ij} + \alpha_{ij} \frac{X_{ij}}{1 - \frac{X_{ij}}{C_{ij}}}$

t_{ij}: travel time with light traffic

C_{ij}: capacity of # cars

α_{ij}: coefficient.

constrains: $X_{ij} \geq 0$

$X_{ij} \leq C_{ij}$

$X_{12} + X_{13} = X$ $X_{12} - X_{23} - X_{24} = 0$

$X_{24} + X_{34} = X$ $X_{13} + X_{23} - X_{34} = 0$.

Maximum likelihood estimator

Samples $X_1 \cdots X_N$

from probablity distribution $g(X, \theta)$. determine θ so that $g(X, \theta)$ best fits the data $X_1 \cdots X_N$

$L(\theta) = \prod_{i=1}^{N} g(X_i, \theta) = P(X_1 \cdots X_N | \theta)$.

assuming $X_1 \cdots X_N$ are independent.

max $L(\theta)$.

$LL(\theta) = \log L(\theta)$

$= \sum_{i=1}^{N} \log g(X_i, \theta)$

Ex: $g(X, \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X-\mu)^2}{2\sigma^2}}$

$\theta = (\mu, S)$ $S = \sigma^2$

$$LL(\theta) = \sum_{i=1}^{N} \log g(X_i, \theta) = \sum_{i=1}^{N} -\frac{1}{2}\log(2\pi S) - \frac{(X-\mu)^2}{2S}$$

$$= \sum_{i=1}^{N} -\frac{1}{2}[\log\sigma^2 + \log(2\pi)] - \frac{(X-\mu)^2}{2S}$$

$\frac{\partial LL}{\partial \mu} = \sum \frac{X_i - \mu}{S}$

$\frac{\partial LL}{\partial \mu} = \sum -\frac{1}{2}\frac{1}{S} + \frac{(X_i-\mu)^2}{2S}$

$= \frac{1}{2S}\left[\sum_{i=1}^{N} \frac{(X_i-\mu)^2}{S} - N\right] = 0.$

$\Rightarrow \sum X_i = N\mu$

$\Rightarrow \mu = \frac{\sum X_i}{N}$

$S = \frac{\sum (X_i - \mu)^2}{N}$

$= \frac{1}{N} (X_i^2 - 2\mu X + \mu^2)$

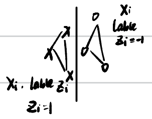
$= \frac{1}{N} [\sum X_i^2 - 2\mu \sum X_i + N\mu^2]$

$= \frac{1}{N} [\sum X_i^2 - N\mu^2]$

$= \frac{\sum X_i^2}{N} - \mu^2$

Support Vector machine

classfiction problem.



X: Category 1

O: Category 2

Seperating plane: $ax+b$

find w, b s.t. the data are seperated as much as possible.

$$\min \Big\| \sum_{z_i=1} \alpha_i x_i - \sum_{z_i=-1} \beta_i x_i \Big\|^2$$

s.t. $\sum \alpha_i = 1$

$\sum \beta_i = 1$

$\alpha_i \geq 0$

$\beta_i \geq 0$

$\alpha_i \leq 1$

$\beta_i \leq 1$

$\Rightarrow \alpha_i^*,\ \beta_i^*$

$$x^* = \sum_{z_i=1} \alpha_i^* x_i$$

$$y^* = \sum_{z_i=-1} \beta_i^* x_i$$