最优化理论
Optimality Theory





目 录 (CONTENT)

10

PART TEN

使用导数的最优化方法

Optimization Method Using Derivative

■ 主要内容

- 最速下降法
- 牛顿法
- 共轭梯度法
- 拟牛顿法
- 信赖域法

大连理工大学 Dalian University Of Technology

■ 共轭方向与扩张子空间定理

定义10.3.1 设A是$n\times n$对称矩阵,若R^n 中的两个方向d^1 和d^2满足

$$(d^1)^T Ad^2=0 \qquad (10.3.1)$$

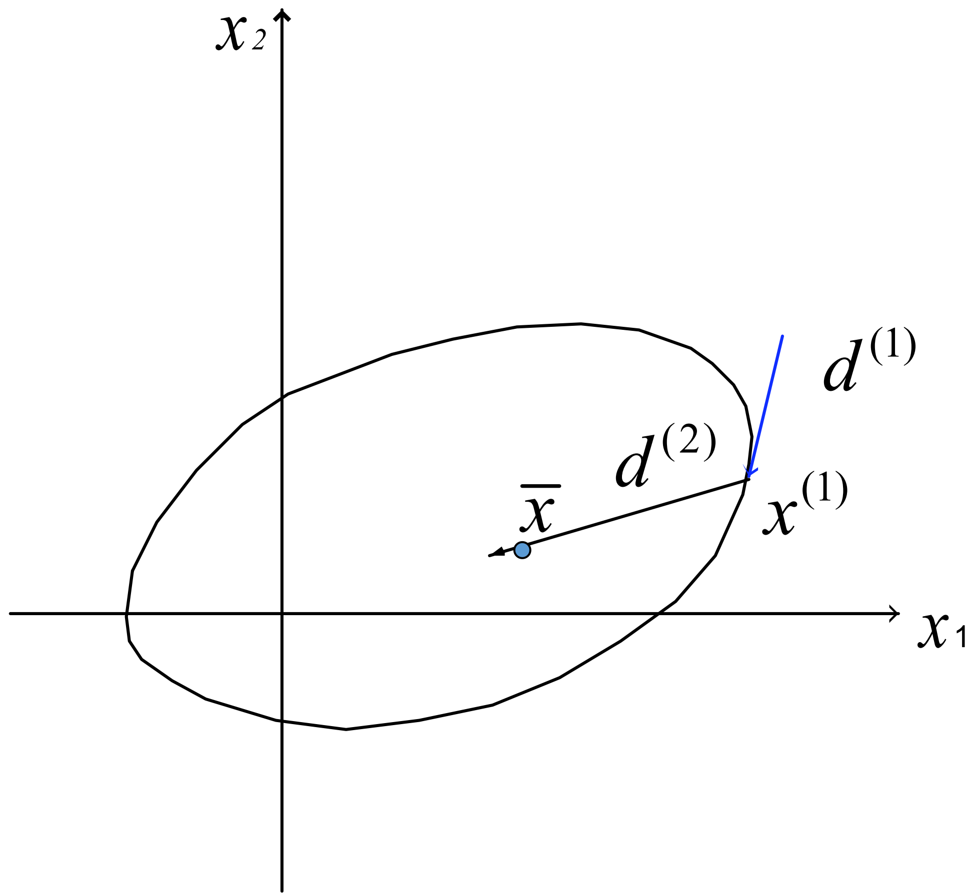
则称这两个方向关于A共轭,或称它们关于A正交.

若$d^{(1)},d^{(2)},\ldots,d^{(k)}$是$R^n$中$k$个方向, 它们两两关于$A$共轭, 即

$$d^{(i)T}Ad^{(j)}=0, i\neq j, \quad i,j=1,2,..k. \qquad (10.3.2)$$

则称这组方向是**A共轭**,或称它们为**A的k个共轭方向**

■ 几何意义



沿着$d^{(1)}$和$d^{(2)}$进行一维搜索，经两次迭代必达到极小点

算法1 共轭方向法

步1(初始化),给定初始点$x^0 \in R^n$, 计算$\nabla f(x^0)$, 给定一个搜索方向$d^0 \neq 0$, 使得$\nabla f(x^0)^T d^0 < 0$; 置$k = 0$

步2(线搜索),求解一维极小化问题$\min\limits_{\alpha \in R} f(x^k + \alpha d^k)$;

$x^{k+1} = x^k + \alpha_k d^k$,若$\nabla f(x^{k+1}) = 0$或$k = n-1$, 停止, 否则转步3

步3(计算共轭方向), 计算一个非零方向$d^{k+1} \in R^n$, 使得

$(d^{k+1})^T G d^j = 0 (\forall j = 0,1,\ldots,k)$; 置$k+1 \to k$转步2

定理10.3.1 设A是n阶对称正定矩阵,$d^{(1)}, d^{(2)}, \ldots, d^{(k)}$是$k$个A共轭的非零向量,则这个向量组线性无关.

定理10.3.2(扩张子空间定理) 设有函数

$$f(x)=\frac{1}{2}x^T Ax+b^T x+c$$

其中A是n阶对称正定矩阵,$d^{(1)},d^{(2)},\dots,d^{(k)}$是A共轭的非零向量. 以任意的$x^{(1)}\in R^n$为初始点, 依次沿$d^{(1)},d^{(2)},\dots,d^{(k)}$进行一维搜索,得到点列$x^{(1)},x^{(2)},\dots,x^{(k+1)}$,则$x^{(k+1)}$是函数$f(x)$在线性流形$x^{(1)}+\Gamma_\kappa$上的唯一极小点.特别地,当$k=n$时,$x^{(k+1)}$是函数$f(x)$在$E^n$上的唯一极小点.其中

$$\Gamma_\kappa=\left\{x\,\middle|\,x=\sum_{i=1}^{k}\lambda_i d^{(i)},\lambda_i\in(-\infty,+\infty)\right\}$$

是$d^{(1)},d^{(2)},\dots,d^{(k)}$生成的子空间.

■ 线性共轭梯度法与二次终止性

上述定理表明,对于二次凸函数,若沿一组共轭方向(非零向量)搜索,经有限步迭代必到达极小点.

- 基本思想:把共轭性与最速下降法相结合,利用已知点处的梯度构造一组共轭方向,并沿着这组方向进行搜索,求出目标函数的极小点

- 注意,初始搜索方向选择最速下降方向十分重要,如果选择别的方向作为初始方向,其余方向均按FR方法构造,则极小化正定二次函数时,这样构造出来的一组方向并不能保证共轭性.

例 考虑下列问题

$$\min \quad x_1^2 + \frac{1}{2}x_2^2 + x_3^2$$

取初始点和初始搜索方向分别为

$$x^{(1)} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, d^{(1)} = \begin{pmatrix} -1 \\ -2 \\ 0 \end{pmatrix}$$

- 可以证明,对于正定二次函数,运用FR法时不作矩阵运算就能求出因子β_i

定理10.3.4 对于正定二次函数,FR法中因子β_i具有下述表达式

$$\beta_i = \frac{\|g_{i+1}\|^2}{\|g_i\|^2}, \qquad (i \geq 1, g_i \neq 0) \qquad (10.3.24)$$

- FR法(对二次凸函数)

1, 给定初点$x^{(1)}$, 置$k=1$.

2, 计算$g_k=\nabla f(x^{(k)})$. 若$\|g_k\|=0$, 停止计算, 得点

$\bar{x}=x^{(k)}$; 否则, 进行下一步.

3, 构造搜索方向. 令

$$d^{(k)}=-g_k+\beta_{k-1}d^{(k-1)}$$

其中, 当$k=1$时, $\beta_{k-1}=0$,

当$k>1$时按公式(10.3.24)计算β_{k-1}.

4, 令 $x^{(k+1)} = x^{(k)} + \lambda_k d^{(k)}$

其中按公式(10.3.17)计算步长λ_k.

5, 若$k = n$, 则停止计算, 得点$\bar{x} = x^{(k+1)}$

否则, 置$k := k+1$, 转2

■ 一般函数的共轭梯度法—非线性共轭梯度法

一般迭代格式

$$\begin{cases} x^{k+1} = x^k + \alpha_k d^k \\ d^{k+1} = -g^{k+1} + \beta_k d^k \end{cases} \quad \text{k=0,1,...} \qquad (10.3.3.1)$$

其中初始方向$d^0 = -g^0$,步长参数α_k由一维搜索得到,β_k的计算公式通常有如下几种:

1, $$\beta_k = \frac{(g^{k+1})^T g^{k+1}}{(g^k)^T g^k} \qquad \text{(Fletcher-Reeves(FR))}$$

2, $$\beta_k = \frac{g_{k+1}^T (g_{k+1} - g_k)}{g_k^T g_k}$$ ----PRP(Polak-Ribiere-Polyar)

3, $$\beta_{\mathrm{k}} = \frac{g_{k+1}^T (g_{k+1} - g_k)}{(d^k)^T (g_{k+1} - g_k)}$$ -----SW(Sorenson-Wolfe)

4, $$\beta_k = \frac{(d^k)^T \nabla^2 f(x^{k+1}) g_{k+1}}{(d^k)^T \nabla^2 f(x^{k+1}) d^k}$$ ----Daniel

5, $$\beta_{\mathrm{k}} = \frac{g_{k+1}^T g_{k+1}}{(d^k)^T g_k}$$ -----Dixon

FR共轭梯度法

1, 给定初始点$x^{(1)}$, 允许误差$\varepsilon > 0$. 置

$y^{(1)} = x^{(1)}, d^{(1)} = -\nabla f(y^{(1)}), k = j = 0.$

2, 若$\left\|\nabla f(y^{(j)})\right\| < \varepsilon$, 则停止计算, 否则作一维搜索,

求λ_j满足 $f(y^{(j)} + \lambda_j d^{(j)}) = \min\limits_{\lambda} f(y^{(j)} + \lambda d^{(j)})$

令 $y^{(j+1)} = y^{(j)} + \lambda_j d^{(j)}$

3,如果$j < n$,转步4,否则,转5

4,令 $d^{(j+1)} = -\nabla f(y^{(j+1)}) + \beta_j d^{(j)}$

其中 $$\beta_j = \frac{\left\|\nabla f(y^{(j+1)})\right\|^2}{\left\|\nabla f(y^{(j)})\right\|^2}$$

置$j := j+1$,转步2.

5,令 $x^{(j+1)} = y^{(n+1)}, y^{(1)} = x^{(k+1)}, d^{(1)} = -\nabla f(y^{(1)})$

置$j = 1, k := k+1$,转步2.

可以证明,对一般函数,共轭梯度法在一定条件下是收敛的,

FR算法中使用精确线搜索，我们有如下收敛性结果

定理 假设函数$f: R^n \to R$有下界，梯度$\nabla f(x)$是Lipschitz连续的.在FR共轭梯度法中，步长参数α_k是由精确线搜索确定的，并且满足充分下降条件(即Armijo条件). 若$\forall k \geq 0, g^k \neq 0$, 则

$$\liminf_{k\to+\infty} \left\| g^k \right\| = 0$$

(Armijo条件:选择步长λ_k满足

$$f(x^k + \lambda_k d^k) \leq f(x^k) + c_1 \lambda_k \nabla f(x^k)^T d^k \quad)$$

■ 拟牛顿条件和算法步骤

基本思想:

牛顿法成功的关键在于利用了Hessian矩阵提供的曲率信息，而计算Hessain矩阵工作量大，并且有的目标函数的Hessian矩阵很难计算，甚至不好求出，这就导致**仅利用目标函数一阶导数的方法**。

拟牛顿法就是利用目标函数值f和一阶导数g的信息,构造出目标函数的曲率近似，而不需要明显形成Hessian矩阵，同时具有收敛速度快的优点。

大连理工大学 Dalian University Of Technology

■ 拟牛顿条件和算法步骤

牛顿法的迭代公式为

$$x^{(k+1)}=x^{(k)}+\lambda_k d^{(k)} \qquad (10.4.1)$$

其中$d^{(k)}$是在$x^{(k)}$处的牛顿方向

$$d^{(k)}=-\nabla^2 f(x^{(k)})^{-1}\nabla f(x^{(k)}) \qquad (10.4.2)$$

λ_k是从$x^{(k)}$出发沿牛顿方向搜索的最优步长.

为构造$\nabla^2 f(x^{(k)})^{-1}$的近似矩阵H_k,先分析$\nabla^2 f(x^{(k)})^{-1}$与一阶导数的关系.

■ 拟牛顿条件和算法步骤

设在第k次迭代后,得点$x^{(k+1)}$,将$f(x)$在点$x^{(k+1)}$Taylor展开

$$f(x) \approx f(x^{(k+1)}) + \nabla f(x^{(k+1)})^T (x - x^{(k+1)}) + \frac{1}{2}(x - x^{(k+1)})^T \nabla^2 f(x^{(k+1)})(x - x^{(k+1)})$$

于是在$x^{(k+1)}$附近

$$g(x) = \nabla f(x) \approx \nabla f(x^{(k+1)}) + \nabla^2 f(x^{(k+1)})(x - x^{(k+1)}) \qquad (10.4.3)$$

令$x = x^{(k)}$,则

$$\nabla f(x^{(k)}) \approx \nabla f(x^{(k+1)}) + \nabla^2 f(x^{(k+1)})(x^{(k)} - x^{(k+1)})$$

■ 拟牛顿条件和算法步骤

记 $$p^{(k)}=x^{(k+1)}-x^{(k)} \tag{10.4.4}$$

$$q^{(k)}=\nabla f(x^{(k+1)})-\nabla f(x^{(k)}) \tag{10.4.5}$$

则

$$q^{(k)}\approx\nabla^2 f(x^{(k+1)})p^{(k)} \tag{10.4.6}$$

设Hessian矩阵$\nabla^2 f(x^{(k+1)})$可逆, 则

$$p^{(k)}\approx\nabla^2 f(x^{(k+1)})^{-1}q^{(k)} \tag{10.4.7}$$

于是, 计算出$p^{(k)}$和$q^{(k)}$, 可根据(10.4.7)估计在$x^{(k+1)}$处的Hessian矩阵的逆. 令H_{k+1}取代牛顿法中的Hessian阵的逆$\nabla^2 f(x^{(k+1)})^{-1}$, 则$H_{k+1}$满足

■ 拟牛顿条件和算法步骤

$$p^{(k)}=H_{k+1}q^{(k)} \qquad (10.4.8)$$

(10.4.8)称为**拟牛顿条件(方程)**，也称为**割线方程**. 怎样确定满足这个条件的H_{k+1}?

■ 对称秩1校正

当$\nabla^2 f(x^{(k)})^{-1}$是n阶对称正定矩阵时,满足拟牛顿条件的矩阵H_{k+1}也应是n阶对称矩阵.于是,构造如此的近似矩阵的一般策略是:

H_1取为任意一n阶对称正定矩阵(如单位阵I),

然后通过修正H_k给出H_{k+1}.

■ 对称秩1校正

$$H_{k+1} = H_k + \Delta H_k \qquad (10.4.9)$$

ΔH_k称为**校正矩阵**. 确定ΔH_k的一个方法是令

$$\Delta H_k = \alpha_k Z^{(k)} Z^{(k)T} \qquad (10.4.10)$$

α_k是一常数, $Z^{(k)}$是n维列向量.

$Z^{(k)}$的选择应使(10.4.8)得到满足, 令

$$p^{(k)} = H_k q^{(k)} + \alpha_k Z^{(k)} Z^{(k)T} q^{(k)} \qquad (10.4.11)$$

从而

$$Z^{(k)} = \frac{p^{(k)} - H_k q^{(k)}}{\alpha_k Z^{(k)T} q^{(k)}} \qquad (10.4.12)$$

■ 对称秩1校正

(10.4.11)等号两端左乘$q^{(k)T}$，整理得

$$q^{(k)\mathrm{T}}(p^{(k)}-H_k q^{(k)})=\alpha_k(Z^{(k)T}q^{(k)})^2 \qquad (10.4.13)$$

利用(10.4.10),(10.4.12-13),(10.4.9)可写成

$$H_{k+1}=H_k+\frac{(p^{(k)}-H_k q^{(k)})(p^{(k)}-H_k q^{(k)})^T}{q^{(k)\mathrm{T}}(p^{(k)}-H_k q^{(k)})} \qquad (10.4.14)$$

---秩1校正公式

利用秩1校正极小化函数$f(x)$,在第k次迭代中,令搜索方向

$$d^{(k)}=-H_k\nabla f(x^{(k)}) \qquad (10.4.15)$$

■ 对称秩1校正

然后沿$d^{(k)}$方向搜索, 求步长λ_k, 满足

$$f(x^{(k)}+\lambda_k d^{(k)})=\min_{\lambda\geq 0} f(x^{(k)}+\lambda d^{(k)})$$

确定后继点

$$x^{(k+1)}=x^{(k)}+\lambda_k d^{(k)} \qquad (10.4.16)$$

■ 对称秩2校正

•1,DFP算法(变尺度法)

定义校正矩阵

$$\Delta H_k = \frac{p^{(k)} p^{(k)T}}{p^{(k)\mathrm{T}} q^{(k)}} - \frac{H_k q^{(k)} q^{(k)\mathrm{T}} H_k}{q^{(k)\mathrm{T}} H_k q^{(k)}} \qquad (10.4.17)$$

则

$$H_{k+1} = H_k + \frac{p^{(k)} p^{(k)T}}{p^{(k)\mathrm{T}} q^{(k)}} - \frac{H_k q^{(k)} q^{(k)\mathrm{T}} H_k}{q^{(k)\mathrm{T}} H_k q^{(k)}} \qquad (10.4.18)$$

DFP (Davidon-Fletcher-Powell)公式

■ 对称秩2校正

- DFP算法

$Step1$,给定初始点$x^{(k)} \in E^n$,允许误差$\varepsilon > 0$,

$Step2$,置$H_1 = I_n$,计算出在$x^{(1)}$处的梯度

$$g_1 = \nabla f(x^{(1)})$$

置$k = 1$

$Step3$,令 $d^{(k)} = -H_k g_k$

$Step4$,从$x^{(k)}$出发,沿$d^{(k)}$进行一维搜索,求λ_k,使

$$f(x^{(k)} + \lambda_k d^{(k)}) = \min_{\lambda \geq 0} f(x^{(k)} + \lambda d^{(k)})$$

令$x^{(k+1)} = x^{(k)} + \lambda_k d^{(k)}$,

■ 对称秩2校正

$Step5$,检验是否满足收敛准则,若

$$\left\|\nabla f(x^{(k+1)})\right\| \le \varepsilon,$$

停止,得$\bar{x} = x^{(k+1)}$,否则,转$Step6$

$Step6$,若$k = n$,则令$x^{(1)} = x^{(k+1)}$,转$Step2$;否则,转$Step7$

$Step7$,令$g_{k+1} = \nabla f(x^{(k+1)}),\ p^{(k)} = x^{(k+1)} - x^{(k)},\ q^{(k)} = g_{k+1} - g_k$

由公式(10.4.18)计算H_{k+1},置$k := k+1$,转$Step3$

■ 对称秩2校正

例1 用DFP方法求解下列问题

$$\min \quad 2x_1^2 + x_2^2 - 4x_1 + 2$$

初始点及初始矩阵分别为

$$x^{(1)} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, H_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

在点$x = (x_1, x_2)^T$的梯度

$$g = \begin{pmatrix} 4(x_1 - 1) \\ 2x_2 \end{pmatrix}$$

■ 对称秩2校正

第1次迭代

在点$x^{(1)}$处的梯度 $g_1 = \begin{pmatrix} 4 \\ 2 \end{pmatrix}$

令搜索方向

$$d^{(1)} = -H_1 g_1 = \begin{pmatrix} -4 \\ -2 \end{pmatrix}$$

从$x^{(1)}$出发,沿方向$d^{(1)}$进行一维搜索,求步长λ_1:

$$\min_{\lambda \geq 0} f(x^{(1)} + \lambda d^{(1)})$$

得到

$$\lambda_1 = \frac{5}{18}$$

■ 对称秩2校正

令

$$x^{(2)}=x^{(1)}+\lambda_1 d^{(1)}=\begin{bmatrix}2\\1\end{bmatrix}+\frac{5}{18}\begin{bmatrix}-4\\-2\end{bmatrix}=\begin{bmatrix}8/9\\4/9\end{bmatrix}$$

$$g_2=\begin{bmatrix}4(\frac{8}{9}-1\\2\frac{4}{9}\end{bmatrix}=\begin{bmatrix}-4/9\\8/9\end{bmatrix}$$

■ 对称秩2校正

第2次迭代

$$p^{(1)} = \lambda_1 d^{(1)} = \begin{pmatrix} -10/9 \\ -5/9 \end{pmatrix}$$

$$q^{(1)} = g_2 - g_1 = \begin{pmatrix} -40/9 \\ -10/9 \end{pmatrix}$$

计算H_2

$$H_2 = H_1 + \frac{p^{(1)} p^{(1)T}}{p^{(1)\mathrm{T}} q^{(1)}} - \frac{H_1 q^{(1)} q^{(1)\mathrm{T}} H_1}{q^{(1)\mathrm{T}} H_1 q^{(1)}}$$

■ 对称秩2校正

$$
=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}+\frac{\begin{pmatrix}-10/9\\ -5/9\end{pmatrix}\begin{pmatrix}-10/9 & -5/9\end{pmatrix}}{\begin{pmatrix}-10/9 & -5/9\end{pmatrix}\begin{pmatrix}-40/9\\ -10/9\end{pmatrix}}
$$

$$
+\frac{\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\begin{pmatrix}-40/9\\ -10/9\end{pmatrix}\begin{pmatrix}-40/9 & -10/9\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}}{\begin{pmatrix}-40/9 & -10/9\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\begin{pmatrix}-40/9\\ -10/9\end{pmatrix}}
$$

■ 对称秩2校正

$$=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}+\frac{1}{18}\begin{pmatrix}4 & 2\\ 2 & 1\end{pmatrix}-\frac{1}{17}\begin{pmatrix}16 & 4\\ 4 & 1\end{pmatrix}$$

$$=\frac{1}{306}\begin{pmatrix}86 & -38\\ -38 & 305\end{pmatrix}$$

令

$$d^{(2)}=-H_2g_2=-\frac{1}{306}\begin{pmatrix}86 & -38\\ -38 & 305\end{pmatrix}\begin{pmatrix}-4/9\\ 8/9\end{pmatrix}$$

$$=\frac{12}{51}\begin{pmatrix}1\\ -4\end{pmatrix}$$

■ 对称秩2校正

从$x^{(2)}$出发,沿方向$d^{(2)}$进行一维搜索, 求步长λ_2:

$$\min_{\lambda \geq 0} f(x^{(2)} + \lambda d^{(2)})$$

得到

$$\lambda_2 = \frac{17}{36}$$

令

$$x^{(3)} = x^{(3)} + \lambda_2 d^{(2)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

■ 对称秩2校正

$$g_3 = \nabla f(x^{(3)}) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

于是得最优解

$$(x_1, x_2) = (1, 0)$$

大连理工大学 Dalian University Of Technology

■ DFP算法的正定性及二次终止性

定理10.4.1若$g_i \neq 0, i=1,2,\ldots,n$,则DFP方法构造的矩阵$H_i(i=1,2,\ldots,n)$为对称正定矩阵.

定理10.4.2 设用DFP方法求解下列问题

$$\min \ f(x)=\frac{1}{2}x^T Ax+b^T x+c$$

其中A为n阶对称正定矩阵.取初点$x^{(1)} \in E^n$,令H_1为n阶对称正定矩阵,则成立:

1, $p^{(i)T} Ap^{(j)}=0, \ 1\le i<j\le k$ (10.4.25)

2, $H_{k+1}Ap^{(i)}=p^{(i)}, \quad 1\le i\le k$ (10.4.26)

其中$p^{(i)}=x^{(i+1)}-x^{(i)}=\lambda_i d^{(i)}, \lambda_i \neq 0, k\le n.$

- 推论:在Th10.4.2的条件下,必有

$$H_{n+1} = A^{-1}$$

•DFP方法中构造出来的搜索方向是一组A共轭方向
•DFP方法具有二次终止性.

•若f是E^n上的二次连续可微实函数,对任意的$\hat{x} \in E^n$存在常数$m > 0$,使得当 $x \in C(\hat{x}) = \{x \mid f(x) \le f(\hat{x})\}$,

$y \in E^n$时有 $m\|y\|^2 \le y^T \nabla^2 f(x) y$

则DFP方法产生的序列$\{x^{(k)}\}$或终止于f 在E^n上的唯一极小点.

■ BFGS公式及 Broyden簇

若用不含二阶导数的矩阵B_{k+1}近似Henssian 矩阵$\nabla^2 f(x^{(k+1)})$, 则由(10.4.6)给出另一种形式的拟牛顿条件, 即

$$q^{(k)}=B_{k+1}p^{(k)} \qquad (10.4.32)$$

可得修正公式

$$B_{k+1}=B_k+\frac{q^{(k)}q^{(k)T}}{q^{(k)\mathrm{T}}p^{(k)}}-\frac{B_k p^{(k)}p^{(k)\mathrm{T}}B_k}{p^{(k)\mathrm{T}}B_k p^{(k)}} \qquad (10.4.33)$$

---关于矩阵B的BFGS修正公式

■ BFGS公式及 Broyden簇

设B_{k+1}可逆,则由(10.4.32)可知

$$p^{(k)}=B_{k+1}^{-1}q^{(k)}$$

于是B_{k+1}^{-1}满足拟牛顿条件(10.4.8),故可令

$$H_{k+1}=B_{k+1}^{-1} \qquad (10.4.34)$$

于是,利用$Sherman-Morrison$公式

可得关于H的BFGS公式:

$$H_{k+1}^{BFGS}=H_k+\left(1+\frac{q^{(k)T}H_kq^{(k)}}{p^{(k)\mathrm{T}}q^{(k)}}\right)\frac{p^{(k)}p^{(k)T}}{p^{(k)\mathrm{T}}q^{(k)}} - \frac{p^{(k)}q^{(k)\mathrm{T}}H_k+H_kq^{(k)}p^{(k)\mathrm{T}}}{p^{(k)\mathrm{T}}q^{(k)}} \qquad (10.4.35)$$

■ BFGS公式及 Broyden簇

上述公式由Broyden,Fletcher,Goldfarb,Shanno(1970)给出.

DFP和BFGS公式都有由$p^{(k)}$和$H_k q^{(k)}$构成的对称秩2校正,故此两公式的加权组合仍具有相同的形式

定义

$$H_{k+1}^{\phi} = (1-\phi)H_{k+1}^{\mathrm{DFP}} + \phi H_{k+1}^{\mathrm{BFGS}} \qquad (10.4.36)$$

---Broyden簇

大连理工大学 Dalian University Of Technology

■ BFGS公式及 Broyden簇

显示表达式

$$H_{k+1}^{\phi} = H_k + \frac{p^{(k)}p^{(k)T}}{p^{(k)\mathrm{T}}q^{(k)}} - \frac{H_k q^{(k)} q^{(k)\mathrm{T}} H_k}{q^{(k)\mathrm{T}} H_k q^{(k)}} + \phi v^{(k)} v^{(k)\mathrm{T}}$$

$$= H_{k+1}^{DFP} + \phi v^{(k)} v^{(k)\mathrm{T}} \qquad (10.4.37)$$

其中

$$v^{(k)} = \left(q^{(k)\mathrm{T}} H_k q^{(k)}\right)^{1/2} \left(\frac{p^{(k)}}{p^{(k)\mathrm{T}} q^{(k)}} - \frac{H_k q^{(k)}}{q^{(k)\mathrm{T}} H_k q^{(k)}} \right)$$

(10.4.38)

■ BFGS公式及 Broyden簇

定理10.4.3 设

$$\min \ f(x) = \frac{1}{2} x^T A x + b^T x + c$$

其中A为n阶对称正定矩阵. 则对于Broyden方法, 成立:

1, $p^{(i)T} A p^{(j)} = 0, \ 1 \le i < j \le k$

2, $H_{k+1} A p^{(i)} = p^{(i)}, \quad 1 \le i \le k$

■ 拟牛顿条件和算法步骤

算法　拟牛顿法

1(初始化),给定初始点$x^0 \in R^n$,正定矩阵$H_0 \in R^{n\times n}$,$\varepsilon \in (0,1)$;

计算$g^0 = \nabla f(x^0)$,置$k=0$.

2(平稳性检验),若$\|g^k\| \le \varepsilon$,则停止,否则,

计算搜索方向$d^k = -H_k g^k$,

3(线搜索),沿射线$R(x^k, d^k) = \{x \mid x = x^k + \lambda d^k, \lambda \ge 0\}$进行线搜索,

求出步长λ_k,令$x^{k+1} = x^k + \lambda_k d^k$

4(修正拟牛顿方程),计算$g^{k+1} = \nabla f(x^{k+1})$,对$H_k$校正，得

H_{k+1}使满足拟牛顿条件，令$k \leftarrow k+1$，转2

■ 算法步骤

- 线搜索方法：每次迭代时产生一搜索方向，在此方向上进行精确或不精确一维搜索，得到下一迭代点。

- 缺点：可能由于步长过大导致算法失败，特别当问题病态时。

■ 算法步骤

- 信赖域方法：

  在每次迭代时，强制性要求新迭代点与当前迭代点之间的距离不超过某一控制量。实际上是，在以当前迭代点为中心的邻域内对一近似于原问题的简单模型求极值。

- 优点：算法稳定性好、收敛性强。

■ 算法步骤

- 问题：

$$\min_{x\in R^n} \quad f(x)$$

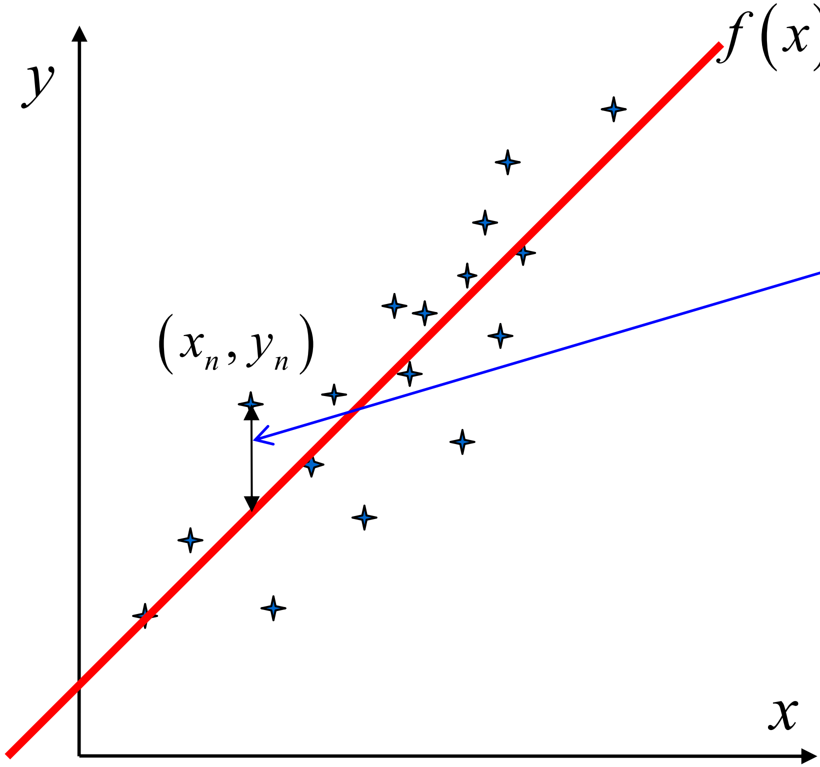
- 基本思想：给定初始迭代点，确定一个以其为中心的邻域，在此域内优化目标函数的二次逼近式，得到下一迭代点。

■ 算法收敛性

- 在一定条件下，信赖域方法具有全局收敛性。
- 设 $f(x)$ 是 R^n 上的实函数，x_1 是给定初始点，$S=\{x \mid f(x)\le f(x_1)\}$ 是有界闭集，$f(x),\nabla f(x)$ 和 $\nabla^2 f(x)$ 在 S 上连续，用信赖域方法求得序列 $\{x_k\}$，则

$$\lim_{k\to\infty} \|\nabla f(x_k)\| = 0$$

■ 一元线性回归/单变量线性回归



$$y_1 \approx wx_1 + b$$
$$y_2 \approx wx_2 + b$$
$$\ldots$$
$$y_N \approx wx_N + b$$

- 回归误差：

$$e_n = y_n - (wx_n + b)$$

- 均方和误差：

$$\sum_{n=1}^{N} e_n^2 = \sum_{n=1}^{N} (y_n - wx_n - b)^2$$

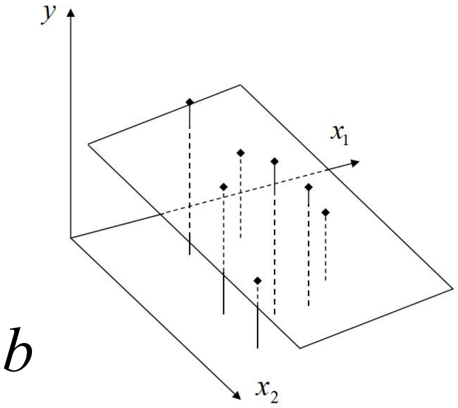
✓ 欧式距离
✓ 能量...

- <u>最小二乘法</u>

<u>(Least Square Method)</u>

$$\boxed{(w^*, b^*) = \arg\min_{w,b} \sum_{n=1}^{N} (y_n - wx_n - b)^2}$$

■ 多元线性回归 (Multivariate Linear Regression)



$$
\begin{aligned}
y_1 &\approx x_1 w + b \\
y_2 &\approx x_2 w + b \\
&\dots \\
y_N &\approx x_N w + b
\end{aligned}
\quad \xrightarrow{\text{一元到多元}} \quad
\begin{aligned}
y_1 &\approx x_{1,1}w_1 + x_{1,2}w_2 + \dots + x_{1,D}w_D + b \\
y_2 &\approx x_{2,1}w_1 + x_{2,2}w_2 + \dots + x_{2,D}w_D + b \\
&\dots \\
y_N &\approx x_{N,1}w_1 + x_{N,2}w_2 + \dots + x_{N,D}w_D + b
\end{aligned}
$$

一般形式

$$
\underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}}_{\mathbf{y}} \approx \underbrace{\begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,D} & 1 \\ x_{2,1} & x_{2,2} & \cdots & x_{2,D} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{N,1} & x_{N,2} & \cdots & x_{N,D} & 1 \end{bmatrix}}_{\mathbf{X}} \underbrace{\begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_D \\ b \end{bmatrix}}_{\mathbf{w}}
$$

$$
\boxed{\mathbf{w}^* = \arg\min_{\mathbf{w}} \|\mathbf{y} - \mathbf{X}\mathbf{w}\|_2^2}
$$

多元最小二乘法

■ 问题求解/优化

$$\mathbf{w}^* = \arg\min_{\mathbf{w}} \|\mathbf{y} - \mathbf{X}\mathbf{w}\|_2^2$$

➢ **闭合解/解析解 (Closed-form Solution/ Analytic Solution):**

$$J(\mathbf{w}) = \|\mathbf{y} - \mathbf{X}\mathbf{w}\|_2^2 = (\mathbf{y} - \mathbf{X}\mathbf{w})^T(\mathbf{y} - \mathbf{X}\mathbf{w})$$

$$\frac{\partial J(\mathbf{w})}{\partial \mathbf{w}} = 2\mathbf{X}^T(\mathbf{y} - \mathbf{X}\mathbf{w}) = 2\mathbf{X}^T\mathbf{X}\mathbf{w} - 2\mathbf{X}^T\mathbf{y} = \mathbf{0}$$

$$\mathbf{w}^* = \underline{\left(\mathbf{X}^T\mathbf{X}\right)^{-1}}\mathbf{X}^T\mathbf{y} \quad \Rightarrow \quad \mathbf{w}^* = \left(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}\right)^{-1}\mathbf{X}^T\mathbf{y}$$

✓ 正则化

✓ 岭回归

➢ **梯度下降法 (Gradient Descent):**

$$\frac{\partial J(\mathbf{w})}{\partial \mathbf{w}} = 2\mathbf{X}^T(\mathbf{y} - \mathbf{X}\mathbf{w}) \quad \Rightarrow \quad \mathbf{w} \leftarrow \mathbf{w} - \eta\frac{\partial J(\mathbf{w})}{\partial \mathbf{w}}$$

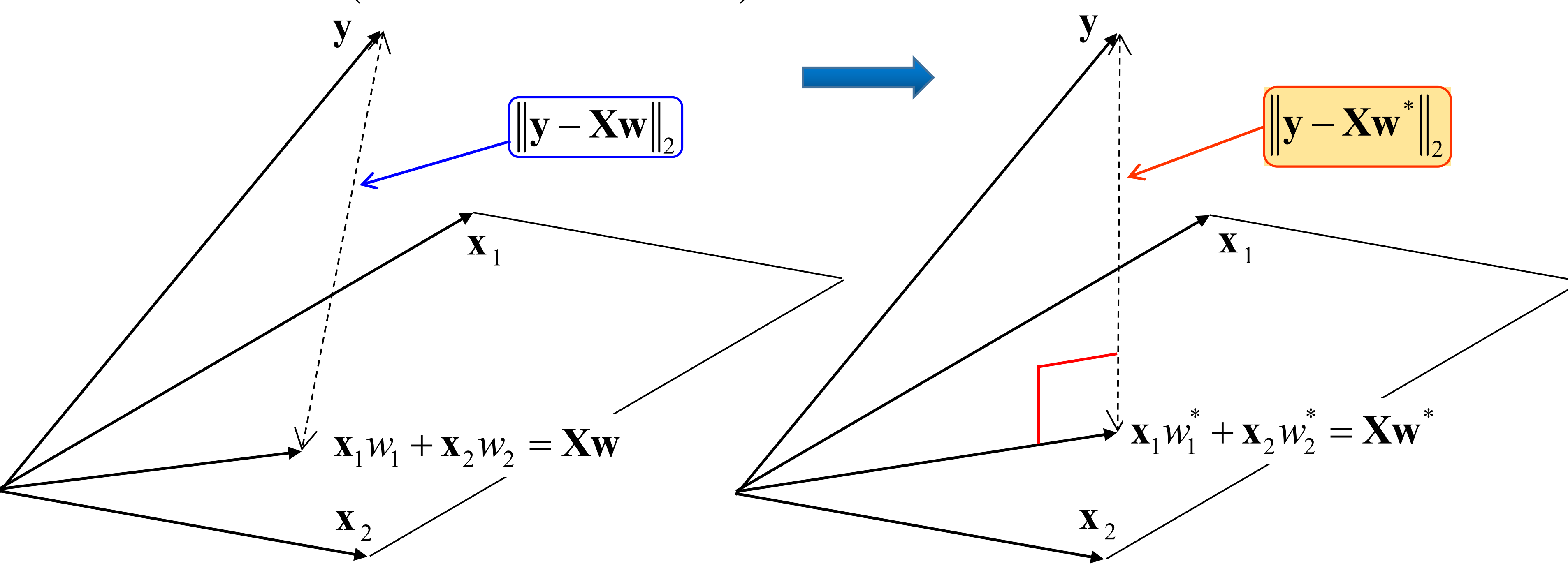
https://www.bilibili.com/video/av41473299?from=search&seid=8346799062692305739

■ 子空间X上的正交投影

$$\mathbf{e}=\mathbf{y}-\mathbf{X}\mathbf{w}=\begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}-\left(\begin{bmatrix} x_{1,1} \\ \dots \\ x_{N,1} \end{bmatrix} w_1+\begin{bmatrix} x_{1,2} \\ \dots \\ x_{N,2} \end{bmatrix} w_2+\dots+\begin{bmatrix} x_{1,D} \\ \dots \\ x_{N,D} \end{bmatrix} w_D\right)$$

$$=\mathbf{y}-\left(\mathbf{x}_1 w_1+\mathbf{x}_2 w_2+\dots+\mathbf{x}_d w_D\right)$$

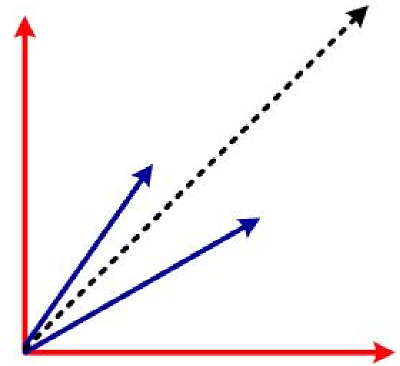
$$\boxed{\mathbf{w}^* = \arg\min_{\mathbf{w}} \|\mathbf{y}-\mathbf{X}\mathbf{w}\|_2^2}$$



线性回归-->最小二乘几何解释

■ 子空间

$$\underbrace{\begin{bmatrix}1\\0\\0\end{bmatrix},\begin{bmatrix}0\\1\\0\end{bmatrix},\begin{bmatrix}0\\0\\1\end{bmatrix}}_{\text{完备正交基}}\rightarrow\underbrace{\begin{bmatrix}1\\1\\0\end{bmatrix},\begin{bmatrix}0\\1\\0\end{bmatrix},\begin{bmatrix}1\\0\\1\end{bmatrix}}_{\text{完备基}}$$

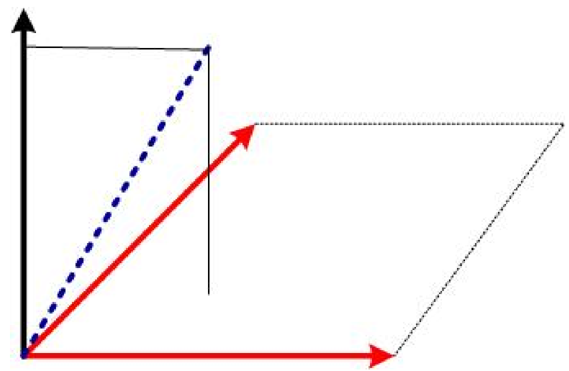


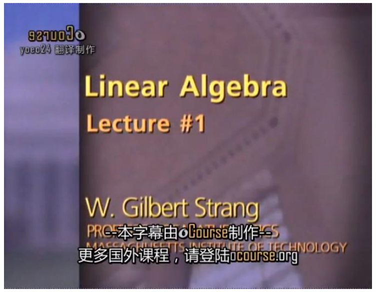
列秩（rank）

子空间

正交补空间

$$\underbrace{\begin{bmatrix}1\\0\\0\end{bmatrix},\begin{bmatrix}0\\1\\0\end{bmatrix},\begin{bmatrix}0\\0\\1\end{bmatrix}}_{\mathbf{a}_1x_1+\mathbf{a}_2x_2+\mathbf{a}_3x_3}\rightarrow\underbrace{\begin{bmatrix}1\\0\\0\end{bmatrix},\begin{bmatrix}0\\1\\0\end{bmatrix}}_{\mathbf{a}_1x_1+\mathbf{a}_2x_2}\rightarrow\begin{bmatrix}0\\0\\1\end{bmatrix}$$





MIT公开课《线性代数》：
https://www.bilibili.com/video/av24368594

线性回归-->最小二乘概率解释

高斯噪声假设

$$\mathbf{e}=\mathbf{y}-\mathbf{X}\mathbf{w} \quad \Rightarrow \quad \begin{bmatrix} e_1 \\ \vdots \\ e_N \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} - \begin{bmatrix} x_{1,1}w_1 + x_{1,2}w_2 + \ldots + x_{1,D}w_D \\ \vdots \\ x_{N,1}w_1 + x_{N,2}w_2 + \ldots + x_{N,D}w_D \end{bmatrix}$$

- 零均值等方差高斯噪声假设

$$e_n \sim \mathrm{N}\left(0,\sigma^2\right) \Rightarrow p\left(e_n\right) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{e_n^2}{2\sigma^2}\right)$$

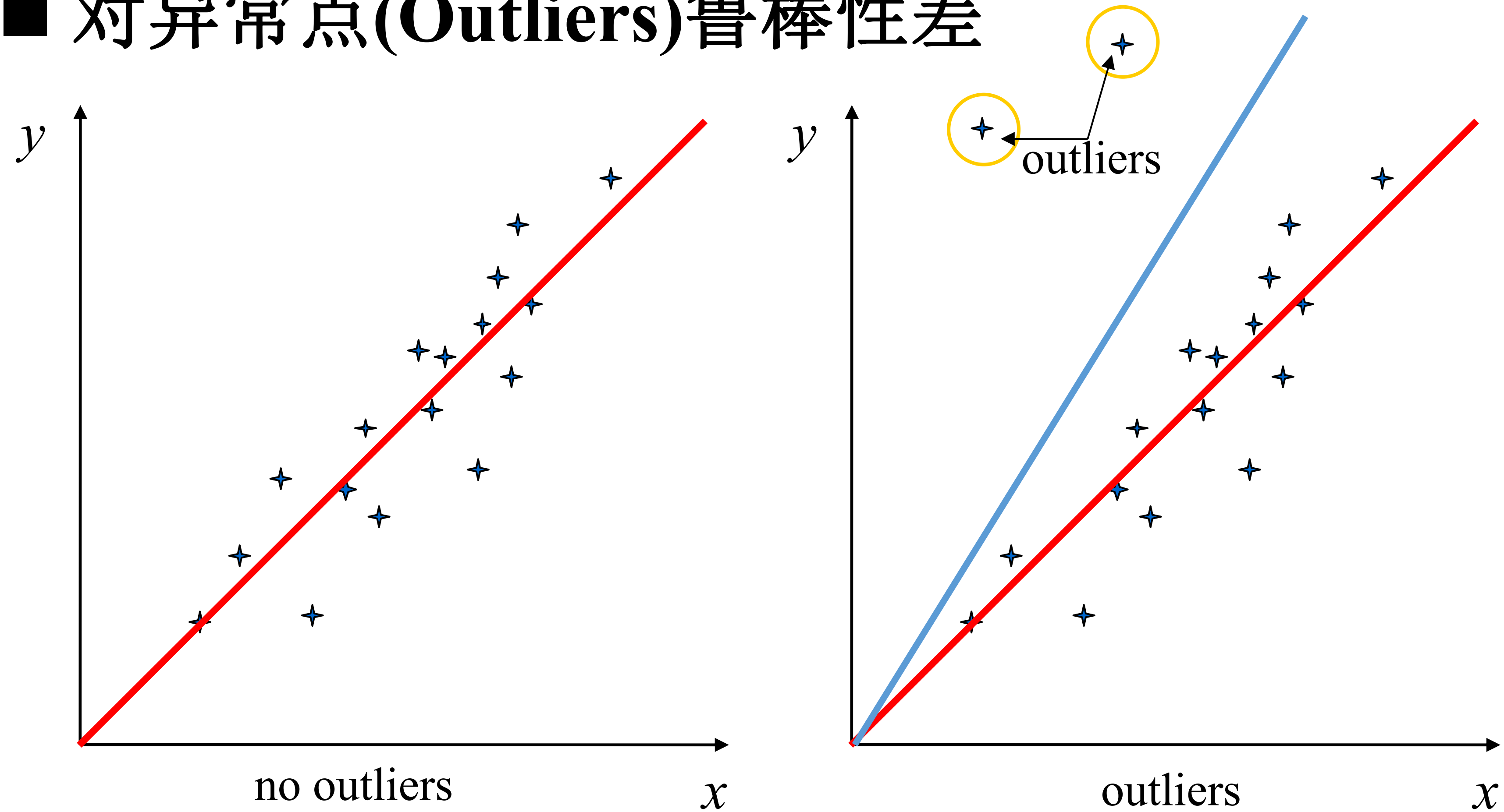
- 最大似然估计

$$p\left(e_1,e_2,\ldots,e_N\right) = \prod_{n=1}^{N} p\left(e_n\right) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^N \exp\left(-\sum_{n=1}^{N} e_n^2 \Big/ 2\sigma^2\right)$$

- 概率解释

$$\max_{\mathbf{w}} p\left(e_1,e_2,\ldots,e_N;\mathbf{w}\right) \Leftrightarrow \min_{\mathbf{w}} \sum_{n=1}^{N} e_n^2 \Leftrightarrow \boxed{\min_{\mathbf{w}} \left\|\mathbf{y}-\mathbf{X}\mathbf{w}\right\|_2^2}$$

■ **对异常点(Outliers)鲁棒性差**



- ➢ **随机取样一致 (Random Sample Consensus, RANSAC)**
- ➢ **鲁棒回归 (Robust Regression)**

线性回归-->建模非线性

■ 线性的含义

$$y = wx + b$$ 线性？

$$y = w_1 x + w_2 x^2 + w_3 x^3 + b$$ 线性？

$$x_1 = x, x_2 = x^2, x_3 = x^3$$
$$y = w_1 x_1 + w_2 x_2 + w_3 x_3 + b$$ 线性？

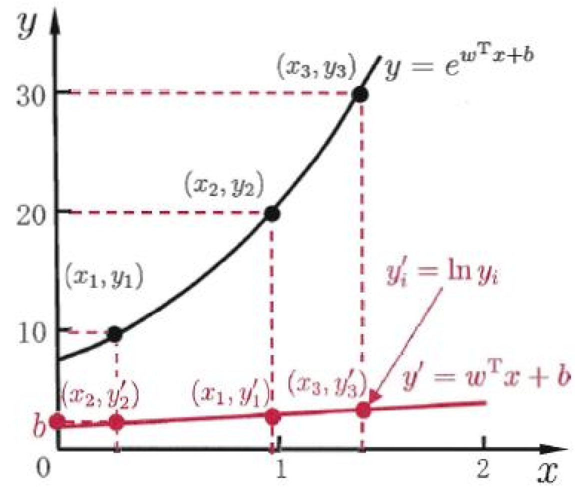
- 线性并不指对输入变量的线性，而是指对参数空间的线性。也就说对于输入来说，完全可以对先对其进行非线性变换，再进行线性组合。从这个角度来说，线性模型完全具有描述非线性的能力。
- 通用非线性化方法：<u>核学习方法 **(Kernel-based Learning Algorithms)**</u>

■ 广义线性模型 (Generalized Linear Model)

➢ 对数线性回归 (Log-Linear Regression)：

$$y = \exp(wx + b) \qquad \log y = wx + b$$

形式上仍是线性回归，实质上已经是求取
输入空间到输出空间的非线性函数映射



摘自《机器学习》，清华大学出版社，周志华著，P56图3.1

➢ 广义线性模型 (Generalized Linear Model)：

$$y = g^{-1}\left(\mathbf{w}^T \mathbf{x}\right) \qquad y' = g(y) = \mathbf{w}^T \mathbf{x}$$

单调可微函数

小结

- 使用导数基本概念
- 最速下降法
- 牛顿法
- 共轭梯度法
- 拟牛顿法
- 信赖域方法

■ 作业

- ➢ 习题 3、4、5、10、14、17、18、19