人工智能数学基础作业题

1. 请将所有的向量范数和矩阵范数整理好，尤其是核范数以及矩阵诱导的范数等概念；

向量范数(Vector Norms)：

1) 0-范数：$\|\boldsymbol{x}\|_0 = \#\{\boldsymbol{i}|\boldsymbol{x}_i \neq \boldsymbol{0}\}$，向量中非零元素的数量。
2) 1-范数：$\|\boldsymbol{x}\|_1 = \sum_{i=1}^{N}|\boldsymbol{x}_i|$，向量所有元素的绝对值之和。
3) 2-范数（欧几里德范数）：$\|\boldsymbol{x}\|_2 = \left(\sum_{i=1}^{N}|\boldsymbol{x}_i|^2\right)^{\frac{1}{2}}$，欧式距离。
4) ∞-范数：$\|\boldsymbol{x}\|_\infty = \boldsymbol{max}_i|\boldsymbol{x}_i|$，所有向量元素中的最大值。
5) $-\infty$-范数：$\|\boldsymbol{x}\|_{-\infty} = \boldsymbol{min}_i|\boldsymbol{x}_i|$，所有向量元素中的最小值。
6) p-范数：$\|\boldsymbol{x}\|_p = \left(\sum_{i=1}^{N}|\boldsymbol{x}_i|^p\right)^{\frac{1}{p}}$（$p$取不同值或极限，可以推得上述几种特殊情况）。

矩阵范数(Matrix Norms)：

1) 1-范数（列和范数）：$\|\boldsymbol{A}\|_1 = \boldsymbol{max}_j\sum_{i=1}^{m}|\boldsymbol{a}_{i,j}|$，矩阵列向量中绝对值之和的最大值。
2) 2-范数（谱范数）：$\|\boldsymbol{A}\|_2 = \sqrt{\boldsymbol{\lambda}_1}$，$\boldsymbol{\lambda}_1$是$\boldsymbol{A}^T\boldsymbol{A}$的最大特征值。
3) ∞-范数（行和范数）：$\|\boldsymbol{A}\|_\infty = \boldsymbol{max}_i\sum_{j=1}^{n}|\boldsymbol{a}_{i,j}|$，矩阵行向量中绝对值之和的最大值。
4) F-范数（Frobenius 范数）：$\|\boldsymbol{A}\|_F = \left(\sum_{i=1}^{m}\sum_{j=1}^{n}|\boldsymbol{a}_{i,j}|^2\right)^{\frac{1}{2}}$，可以表示矩阵的能量。
$$\|\boldsymbol{A}\|_F^2 = \sum_{ij}|\boldsymbol{a}_{ij}|^2 = \sum_i\|\boldsymbol{A}_{i*}\|_2^2 = \sum_j\|\boldsymbol{A}_{*j}\|_2^2 = \boldsymbol{trace}(\boldsymbol{A}^T\boldsymbol{A}) = \boldsymbol{trace}(\boldsymbol{A}\boldsymbol{A}^T) = \boldsymbol{Energy}(\boldsymbol{A})$$
5) 核范数：$||A||_* = \sum_{i=1}^{r}\lambda_i, r = rank(A)$。

诱导矩阵范数(Induced Matrix Norms)： 将矩阵转换为向量形式计算，向量范数诱导出矩阵范数。

$\|\boldsymbol{A}\| = \boldsymbol{max}_{\|x\|=1}\|\boldsymbol{Ax}\|, \boldsymbol{A} \in \boldsymbol{C}^{m\times n}, \boldsymbol{x} \in \boldsymbol{C}^{n\times 1}$；若$A$非奇异，则$\frac{1}{\|A^{-1}\|} = \boldsymbol{min}_{\|x\|=1}\|\boldsymbol{Ax}\|$

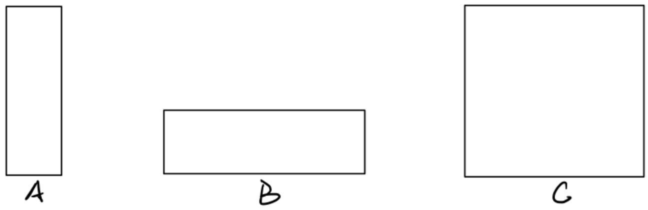
$\|\boldsymbol{A}\|_2 = \boldsymbol{max}_{\|x\|_2=1}\|\boldsymbol{Ax}\|_2 = \sqrt{\boldsymbol{\lambda}_{max}}$；若$A$非奇异，$\|\boldsymbol{A}^{-1}\|_2 = \frac{1}{\boldsymbol{min}_{\|x\|_2=1}\|\boldsymbol{Ax}\|_2} = \frac{1}{\sqrt{\boldsymbol{\lambda}_{min}}}$，$\lambda$分别为使得

$\boldsymbol{A}^T\boldsymbol{A} - \boldsymbol{\lambda I}$ 奇异的最大/小的特征值。

矩阵范数的运算：

- $\|\boldsymbol{A}\| \geq \boldsymbol{0}, \|\boldsymbol{A}\| = \boldsymbol{0} \Leftrightarrow \boldsymbol{A} = \boldsymbol{0}$
- $\|\boldsymbol{\alpha A}\| = |\boldsymbol{\alpha}|\|\boldsymbol{A}\|$
- $\|\boldsymbol{A} + \boldsymbol{B}\| \leq \|\boldsymbol{A}\| + \|\boldsymbol{B}\|$
- $\|\boldsymbol{AB}\| \leq \|\boldsymbol{A}\|\|\boldsymbol{B}\|$

2. 假设有三个矩阵A, B, C，其大小分别为$10 \times 2, 2 \times 10, 10 \times 10$：
   a) 假设需要计算矩阵乘积：ABC. 从有效性的观点来看，计算(AB)C还是计算A(BC)更有意义？
   b) 如果去计算矩阵乘积：CAB，则计算(CA)B和计算C(AB)哪个更有效？



设浮点乘时间耗费为a，浮点加时间耗费为b.

a) (AB)C : $\underbrace{(2a+b)*100}_{AB} + \underbrace{(10a+9b)*100}_{(AB)C} = 1200a+1000b.$

A(BC) : $\underbrace{(10a+9b)*20}_{BC} + \underbrace{(2a+b)*100}_{A(BC)} = 600a+280b.$

显然，A(BC)更有意义

b) C(AB) : AB为10*10矩阵，故乘法同于(AB)C.

(CA)B : $\underbrace{(10a+9b)*20}_{CA} + \underbrace{(2a+b)*100}_{(CA)B} = 600a+280b.$

显然，(CA)B更有效

3. 证明：如果矩阵满足：$A=-A^T$，则对任意列向量x, 都有：$x^TAx=0$.

证： 不妨设$A=\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}$，$x=[x_1,x_2,\dots,x_n]^T$，

则由$A=-A^T$有$a_{ij}=-a_{ji}(i,j\le n)$，显然A的主对角线上$a_{kk}=0(k\le n)$.

则对于任意列向量x，

$x^TAx=[\sum_{i=1}^n x_i a_{i1},\sum_{i=1}^n x_i a_{i2},\dots,\sum_{i=1}^n x_i a_{in}]\cdot x=\sum_{j=1}^n x_j\sum_{i=1}^n x_i a_{ij}=\sum_j\sum_i a_{ij}x_ix_j$.

而由上式显然可知$a_{ij}x_ix_j=-a_{ji}x_jx_i$，故可化简得$x^TAx=0$

4. 考虑二维空间中的两条直线：$y=3x+4$，和$y=5x+2$。将其写成矩阵形式，$A\begin{bmatrix} x \\ y \end{bmatrix}=b$，则$A,b$分别是什么？这两条直线的交点坐标如何求解？

解：$\begin{cases} y=3x+4 \\ y=5x+2 \end{cases} \Rightarrow \begin{cases} y-3x=4 \\ y-5x=2 \end{cases} \Rightarrow \begin{cases} -3x+y=4 \\ -5x+y=2 \end{cases}$

$\therefore \begin{bmatrix} -3 & 1 \\ -5 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}=\begin{bmatrix} 4 \\ 2 \end{bmatrix}$. 即$A=\begin{bmatrix} -3 & 1 \\ -5 & 1 \end{bmatrix}$，$b=\begin{bmatrix} 4 \\ 2 \end{bmatrix}$

交点坐标：高斯消元法

① 前向消元 $\begin{bmatrix} -3 & 1 \\ 0 & -\frac{2}{3} \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}=\begin{bmatrix} 4 \\ -\frac{14}{3} \end{bmatrix}$

② 后向替换 $x=1$，$y=7$ ⇒ 交点坐标(1,7)

5. 证明：任何正交的上三角矩阵是对角矩阵。

证明：不妨设A为正交的上三角矩阵，则A必然满足：

① A是上三角矩阵，主对角线以下元素为0

② A满足正交性 $A^TA=I\ (A^T=A^{-1})$

$A=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & a_{22} & & \vdots \\ & & \ddots & \\ 0 & \cdots & \cdots & a_{nn} \end{bmatrix}$

则 A^TA 矩阵第i行第j列元素 x_{ij}

$$x_{ij}=\sum_k a_{ki}\cdot a_{kj}$$

且当 $k>m$ 时，$a_{km}=0$（上三角） ∴ $k\le\min(i,j)$ 时，$a_{ki}\cdot a_{kj}$ 才不为0.

故当 $i=j$ 时，$x_{ij}=x_{ii}=\sum_{k=1}^{i}a_{ki}^2=1$

而 $i\neq j$ 时 $x_{ij}=\sum_{k=1}^{\min(i,j)}a_{ki}\cdot a_{kj}=0$

对于 $i=1$ 时，$x_{11}=a_{11}^2=0$，$x_{1m}=a_{11}a_{1m}=0$ $(m\neq 1)$ 显然 $a_{1m}=0\ (m\neq 1)$

同理即可递推得，当且仅当对角线上 $a_{ii}\neq 0$ 且 $a_{ii}^2=1$，其余位置均为0.

∴ A为对角矩阵

6. 计算对称矩阵$A=\begin{bmatrix}a&a&a&a\\a&b&b&b\\a&b&c&c\\a&b&c&d\end{bmatrix}$的LU分解，$L=?$, $U=?$若要求$A=LU$的四个主元（就是上三角矩阵中对角线元素）都不为0，则对a,b,c,d的四个条件分别是什么？

解：$A=\begin{bmatrix}a&a&a&a\\a&b&b&b\\a&b&c&c\\a&b&c&d\end{bmatrix}\Rightarrow\begin{bmatrix}a&a&a&a\\0&b-a&b-a&b-a\\0&b-a&c-a&c-a\\0&b-a&c-a&d-a\end{bmatrix}$

$$\underbrace{\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&-1&1\end{bmatrix}}_{L_3}\underbrace{\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&-1&1&0\\0&-1&0&1\end{bmatrix}}_{L_2}\underbrace{\begin{bmatrix}1&0&0&0\\-1&1&0&0\\-1&0&1&0\\-1&0&0&1\end{bmatrix}}_{L_1}\begin{bmatrix}a&a&a&a\\a&b&b&b\\a&b&c&c\\a&b&c&d\end{bmatrix}=\underbrace{\begin{bmatrix}a&a&a&a\\0&b-a&b-a&b-a\\0&0&c-b&c-b\\0&0&0&d-c\end{bmatrix}}_{U}$$

$$\begin{bmatrix}a&a&a&a\\0&b-a&b-a&b-a\\0&0&c-b&c-b\\0&0&c-b&d-b\end{bmatrix}$$

$$L=L_1^{-1}L_2^{-1}L_3^{-1}=\begin{bmatrix}1&0&0&0\\1&1&0&0\\1&0&1&0\\1&0&0&1\end{bmatrix}\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&1&1&0\\0&1&0&1\end{bmatrix}\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&1&1\end{bmatrix}=\begin{bmatrix}1&0&0&0\\1&1&0&0\\1&1&1&0\\1&1&1&1\end{bmatrix}$$

若四个主元都不为0，则满足 $a\neq b\neq c\neq d$

7. 三对角矩阵是指除主对角以及相邻的上下两个对角上的元素非零外，别的元素都为零的矩阵。请分解下列矩阵为：$A=LU$，如果矩阵对称则进一步有：$A=LDL^T$

a) $A=\begin{bmatrix}1&1&0\\1&2&1\\0&1&2\end{bmatrix}$;

解： 分解为LU形式，解得$L=\begin{bmatrix}1&0&0\\1&1&0\\0&1&1\end{bmatrix}$, $U=\begin{bmatrix}1&1&0\\0&1&1\\0&0&1\end{bmatrix}$;

分解为LDL^T形式，则$L=\begin{bmatrix}1&0&0\\1&1&0\\0&1&1\end{bmatrix}$, $D=\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}$

b) $A=\begin{bmatrix} a & a & 0 \\ a & a+b & b \\ 0 & b & b+c \end{bmatrix}$;

解： 分解为LU形式，解得$L=\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}, U=\begin{bmatrix} a & a & 0 \\ 0 & b & b \\ 0 & 0 & c \end{bmatrix}$;

分解为LDL^T形式，则$L=\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}, D=\begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix}$

8. 矩阵A的大小为$m\times n$, 假设矩阵的秩为r, 则矩阵A的行空间的维数为多少？其零空间的维数为多少？矩阵A的列空间的维数为多少？A^T的零空间的维数为多少？

1) A行空间的维数：矩阵A的行空间由A的行向量张成。矩阵的秩为r，表示矩阵A的行空间的线性无关的行向量的数量为r，因而矩阵A的行空间的维数为r。
2) A列空间的维数：同理，矩阵A的列空间的维数为r。
3) A零空间的维数：矩阵A的大小为$m\times n$，秩为r, 则零空间维数为$n-r$。
4) A^T零空间的维数：同理，矩阵A^T的大小为$n\times m$，秩为r, 则零空间维数为$m-r$。

9. 如果u, v是正交的单位向量（长度为 1），证明：$u+v$与$u-v$正交，这两个向量的长度是多少？

证明：由题可知 $<u,v>=0$ 且 $||u||_2=1$, $||v||_2=1$.

则 $<u+v, u-v> = <u,u> - <u,v> + <v,u> - <v,v>$

由内积的性质，原式 $= ||u||^2 - ||v||^2 = 0$

$\therefore$ $u+v$ 与 $u-v$ 正交

长度由勾股定理计算得 $\sqrt{2}$. $||u+v||^2 = ||u-v||^2 = ||u||^2 + ||v||^2$.

10. 正交矩阵Q满足：$||Qx||^2=||x||^2$。试证明：$(Qx)^T(Qy)=x^Ty$, 对任意向量x, y都成立，即由Q变换后不改变其长度和角度。

证明：因为Q为正交矩阵，故 $Q^TQ=I$，对 $\forall x,y\in R^n$，有.

$(Qx)^T(Qy) = x^TQ^TQy = x^TIy = x^Ty$

且 $||Qx||^2 = (Qx)^T(Qx) = x^Tx = ||x||^2$.

$\therefore$ 由Q变换不改变其长度和角度.

11. 置换矩阵是单位矩阵交换行列后形成。试解释为什么置换矩阵与任意别的置换矩阵都正交？假设$P=\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$, 则$P^TP=$ ___I_4___，$P^{-1}=$ ___P^T___.

解释： 对于一个置换矩阵P，每一行和每一列恰好有一个元素等于1，从而每一行和每一列都是正交的。并且，其转置矩阵等于其逆矩阵，原因是记P中为1为a_{ij}, 则P^T中a'_{ji}为1，所以PP^T中第i行第i列的值为1，并且此位置的十字上再无为1的值。

解得$P^TP=\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix}, P^{-1}=\begin{bmatrix}0&0&0&1\\1&0&0&0\\0&1&0&0\\0&0&1&0\end{bmatrix}$

12. 置换矩阵P(同上题)的特征向量分别为

$$x_1=(1,1,1,1)^T, x_2=(1,i,i^2,i^3), x_3=(1,i^2,i^4,i^6)^T, x_4=(1,i^3,i^6,i^9)^T.$$

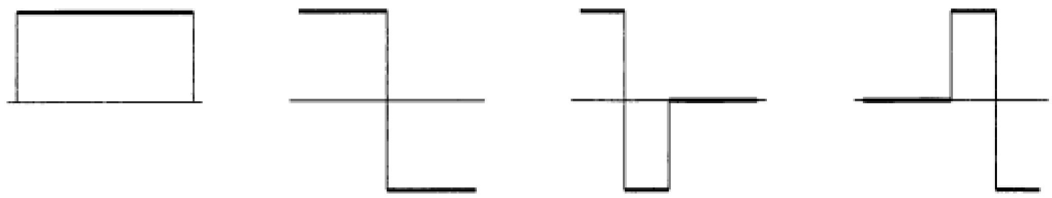
试求对应的四个特征值$\lambda_1,\lambda_2,\lambda_3,\lambda_4$。$4\times4$傅立叶变换矩阵F的列正好是这些特征向量，试证明：$Q=\frac{F}{2}=\frac{1}{2}\begin{bmatrix}1&1&1&1\\1&i&-1&-i\\1&-1&1&-1\\1&-i&-1&i\end{bmatrix}$时，有$Q^*Q=I$,（$Q^*$表示$Q$的复共轭转置）

解： 由$Ax=\lambda x$可知，$\lambda_1=1,\lambda_2=i,\lambda_3=i^2,\lambda_4=i^3$

则$Q^*=\frac{1}{2}\begin{bmatrix}1&1&1&1\\1&-i&-1&i\\1&-1&1&-1\\1&i&-1&-i\end{bmatrix}$，$Q^*Q=\frac{1}{4}\begin{bmatrix}1&1&1&1\\1&-i&-1&i\\1&-1&1&-1\\1&i&-1&-i\end{bmatrix}\begin{bmatrix}1&1&1&1\\1&i&-1&-i\\1&-1&1&-1\\1&-i&-1&i\end{bmatrix}$

故$Q^*Q=\frac{1}{4}\begin{bmatrix}1+1+1+1&1+i-1-i&1-1+1-1&1-i-1+i\\1-i-1+i&1-i^2+1-i^2&1+i-1-i&1+i^2+1+i^2\\1-1+1-1&1-i-1+i&1+1+1+1&1+i-1-i\\1+i-1-i&1+i^2++1+i^2&1-i-1+i&1-i^2+1-i^2\end{bmatrix}=I_4$

13. 小波变换也是一种常用的数学变换，其变换矩阵W只包含 1，-1 和 0 值。对于4×4的小波变换，变换矩阵$W=\begin{bmatrix}1&1&1&0\\1&1&-1&0\\1&-1&0&1\\1&-1&0&-1\end{bmatrix}$. 试计算：$W^TW,W^{-1}$。注意，该变换矩阵的每一列的示意图如下图所示。



第 1 列　　第 2 列　　第 3 列　　第 4 列

图 1.13 题中4×4小波变换列的几何意义

解： 代入计算得$W^TW=\begin{bmatrix}1&1&1&1\\1&1&-1&-1\\1&-1&0&0\\0&0&1&-1\end{bmatrix}\begin{bmatrix}1&1&1&0\\1&1&-1&0\\1&-1&0&1\\1&-1&0&-1\end{bmatrix}=\begin{bmatrix}4&0&0&0\\0&4&0&0\\0&0&2&0\\0&0&0&2\end{bmatrix}$

同时$W^{-1}=\begin{bmatrix}1/4&1/4&1/4&1/4\\1/4&1/4&-1/4&-1/4\\1/2&-1/2&0&0\\0&0&1/2&-1/2\end{bmatrix}$

14. 马尔可夫矩阵$A=\begin{bmatrix}0.6 & 0.2\\ 0.4 & 0.8\end{bmatrix}$, $A^{\infty}=\begin{bmatrix}\frac{1}{3} & \frac{1}{3}\\ \frac{2}{3} & \frac{2}{3}\end{bmatrix}$, 试求其特征值和特征向量。并解释，为何$A^{100}$近似于$A^{\infty}$。在马尔科夫蒙特卡洛采样方法中，这种概率转移方式可以用来生成服从某种分布的序列，为什么？

解： 由特征值定义，$|A-\lambda E|=0 \Rightarrow \begin{vmatrix}0.6-\lambda & 0.2\\ 0.4 & 0.8-\lambda\end{vmatrix}=(0.6-\lambda)(0.8-\lambda)-0.08=0$

解得$\lambda_1=1, \lambda_2=0.4$,代回原式可以解得对应的特征向量分别为$[1,2]^T, [1,-1]^T$

A^{100}近似A^{∞}：显然，由矩阵A引出的马尔可夫链状态数有限且不可约，并且满足非周期性。因而$\lim_{n\to\infty} A_{ij}^{n}$存在且与$i$无关，达到平稳分布。这种状态一般在有限次转移达到，这里达到的次数小于 100，因而可以用A^{100}近似A^{∞}。

马尔科夫蒙特卡洛采样方法： MCMC 方法的基本思想是通过在马尔可夫链上进行随机游走，使得经过足够长的时间后，样本的分布趋于我们所期望的目标分布。而无论初始给定何种概率分布，经过足够长的马尔科夫游走之后，总会得到一个由且仅由转移矩阵A决定的概率分布。此算法的关键在于构造适当的链（状态转移矩阵），可以根据细致平稳条件引入接收率α构造。

15. 请计算下面凸函数在指定点的次微分：

a) $f(x_1,x_2,x_3)=max\{|x_1|,|x_2|,|x_3|\}$, 在点$(x_1,x_2,x_3)=(0,0,0)$

b) $f(x)=e|x|$, 在点$x=0$ 处（这是标量）

c) $f(x1,x2)=max\{x_1+x_2-1, x_1-x_2+1\}$在点$(x_1,x_2)=(1,1)$

解： ①对于$f(x_1,x_2,x_3)$在点$(0,0,0)$的次微分为$\{[-1,1],[-1,1],[-1,1]\}$；

②对于$f(x)=e|x|$在点$x=0$处，次微分为$[-e,e]$；

③对于$f(x1,x2)$在点$(1,1)$的次微分为$\{1,[-1,1]\}$。

16. 用最速下降法求解$\min f(x)=x_1-x_2+2x_1^2+2x_1x_2+x_2^2$, 取$x_0=(0,0)^T$要求迭代两次。

解： ①首先对$f(x)$求梯度$\nabla f(x)=(1+4x_1+2x_2,-1+2x_1+2x_2)^T$，则在$x_0=(0,0)^T$处代入得$\nabla f(x_0)=(1,-1)^T$，设下降率为$\lambda_1$，则由最速下降法$\min\left\|f\left(x_0-\lambda_1\nabla f(x_0)\right)\right\|$，即$\min(-\lambda_1-\lambda_1+2\lambda_1^2-2\lambda_1^2+\lambda_1^2)=\min(\lambda_1^2-2\lambda_1)$, 则解得$\lambda_1=1$，此时$x_1=(-1,1)^T$；

②代入$\nabla f(x_1)=(-1,-1)^T$，设下降率为λ_2，则由最速下降法$\min\left\|\nabla f\left(x_1-\lambda_2\nabla f(x_1)\right)\right\|$，同理解得$\lambda_2=0.2$，此时$x_2=(-0.8,1.2)^T$；代入得$f(x_2)=-1.2$。

17. 设$f_1(x),f_2(x),f_3(x),\cdots,f_m(x)$均为凸函数，讨论下列函数是否是凸函数。若是则给出证明，否则举一反例。

a) $g(x)=max\{f_1(x),f_2(x),\cdots,f_m(x)\}, x\in R$

b) $g(x)=min\{f_1(x),f_2(x),\cdots,f_m(x)\}, x\in R$

c) $g(x)=\sum_{i=1}^{n}x_i\log(x_i)$, $x\in R^n, x_i>0, \sum_{i=1}^{n}x_i=1$

d) $g(x)=\lfloor x\rfloor=\max\{k\le x, k$是整数$\}$

e) $g(x)=\sum_{i=1}^{k}x_{(i)}$，$x\in R^n, k\in Z, 1\le k\le n$，$x_{(i)}$表示向量$x$的第$i$个最大元素

证明：显然，函数f的定义域是凸集，故g的定义域也是凸集。

①只需证明max函数是凸的，对于$h_1(x)=\max\big(u(x),v(x)\big)$，$u,v$是凸函数。则对于$0<\lambda<1$，

$h_1(\lambda x+(1-\lambda)y)=\max\big(u(\lambda x+(1-\lambda)y),v(\lambda x+(1-\lambda)y)\big)$

$\le\max\big(\lambda u(x)+(1-\lambda)u(y),\lambda v(x)+(1-\lambda)v(y)\big)$

$\le\lambda\max\big(u(x),v(x)\big)+(1-\lambda)\max\big(u(y),v(y)\big)=\lambda h_1(x)+(1-\lambda)h_1(y)$

$g(x)$为是多个凸函数$f_1(x),f_2(x),f_3(x),\cdots,f_m(x)$的max操作，因而$g(x)$也是凸的。

②min函数是非凸的，对于$h_2(x)=\min(x,-x)=-|x|$，显然$f_1(x)=x, f_2(x)=-x$是凸函数。但是不满足对任意$0<\lambda<1$，都有$h_2(\lambda x+(1-\lambda)y)\le\lambda h_2(x)+(1-\lambda)h_2(y)$，如$\lambda x+(1-\lambda)y=0$，但$\lambda x\ne0$，$(1-\lambda)f(y)\ne0$，显然左侧等于0，右侧为一个负数，等式不成立。

③首先证明$\sum_{i=1}^{n}p_i\log\left(\frac{1}{p_i}\right)\le\sum_{i=1}^{n}p_i\log\left(\frac{1}{q_i}\right)$，对于$\forall x>0, logx\le x-1$，等号仅当$x=1$成立。

则$\sum_{i=1}^{n}p_i\log\left(\frac{1}{p_i}\right)-\sum_{i=1}^{n}p_i\log\left(\frac{1}{q_i}\right)=\sum_{i=1}^{n}p_i\log\left(\frac{q_i}{p_i}\right)\le\sum_{i=1}^{n}p_i\left(\frac{q_i}{p_i}-1\right)=\sum_{i=1}^{n}(q_i-p_i)=0$，

即$\sum_{i=1}^{n}p_i\log(p_i)\ge\sum_{i=1}^{n}p_i\log(q_i)$，当且仅当$p_i=q_i$等号成立。

设P_1,P_2是两个概率分布，下证对任意$0<\lambda<1$，$g(\lambda P_1+(1-\lambda)P_2)\le\lambda g(P_1)+(1-\lambda)g(P_2)$

根据引理，令$q_i=\lambda p_{1i}+(1-\lambda)p_{2i}$，因此$p_{1i}\ne q_i$，不满足引理等号成立的条件。

由$\sum_{i=1}^{n}p_{1i}\log\left(\frac{1}{p_{1i}}\right)\le\sum_{i=1}^{n}p_{1i}\log\left(\frac{1}{q_i}\right)$，因此$\sum_{i=1}^{n}p_{1i}\log\left(\frac{q_i}{p_{1i}}\right)<0$，$\sum_{i=1}^{n}p_{2i}\log\left(\frac{q_i}{p_{2i}}\right)<0$

则$g(\lambda P_1+(1-\lambda)P_2)-\lambda g(P_1)-(1-\lambda)g(P_2)$

$=\sum_{i=1}^{n}(\lambda p_{1i}+(1-\lambda)p_{2i})\log(\lambda p_{1i}+(1-\lambda)p_{2i})-\lambda\sum_{i=1}^{n}p_{1i}logp_{1i}-(1-\lambda)\sum_{i=1}^{n}p_{2i}\log p_{2i}$

$=\lambda\sum_{i=1}^{n}p_{1i}\log\left(\frac{\alpha p_{1i}+(1-\alpha)p_{2i}}{p_{1i}}\right)+(1-\lambda)\sum_{i=1}^{n}p_{2i}\log\left(\frac{\alpha p_{1i}+(1-\alpha)p_{2i}}{p_{2i}}\right)$

$=\lambda\sum_{i=1}^{n}p_{1i}\log\left(\frac{q_i}{p_{1i}}\right)+(1-\lambda)\sum_{i=1}^{n}p_{2i}\log\left(\frac{q_i}{p_{2i}}\right)<0$，等式成立，所以$g(x)$为凸函数。

④$\lfloor x\rfloor$是非凸的，对于任意$0<\lambda<1$

$g(\lambda x+(1-\lambda)y)=\lfloor\lambda x+(1-\lambda)y\rfloor\le\lambda\lfloor x\rfloor+(1-\lambda)\lfloor y\rfloor$并不一定成立，如$x=0.9, y=3, \lambda=0.4$，等式左边等于$\lfloor0.36+1.8\rfloor=2$，等式右边等于$0.6\times3=1.8$。等式不成立。

⑤对于任意$0<\lambda<1$，$g(x)=\sum_{i=1}^{k}x_{(i)}$

$$g(\lambda x+(1-\lambda)y)=\sum_{i=1}^{k}(\lambda x+(1-\lambda)y)_{(i)}\le\lambda\sum_{i=1}^{k}x_{(i)}+(1-\lambda)\sum_{i=1}^{k}y_{(i)}=\lambda g(x)+(1-\lambda)g(y)$$

所以$g(x)$是凸函数。

18. 将下列线性规划问题化成标准型，并采用代数法，求解其所有的基本解，验证其最优解。

a) $Min\, z = 2x_1 - x_2 + 2x_3$，$s.t.\begin{cases} -x_1 + x_2 + x_3 = 4 \\ -x_1 + x_2 - x_3 \le 6 \\ x_1 \le 0, x_2 \ge 0, x_3\text{无约束} \end{cases}$

解： $Max\, z' = 2x_7 + x_2 - 2(x_4 - x_5) + 0x_6$

$s.t.\begin{cases} x_7 + x_2 + (x_4 - x_5) = 4 \\ x_7 + x_2 - (x_4 - x_5) + x_6 = 6 \\ x_2, x_4, x_5, x_6, x_7 \ge 0 \end{cases}, A = \begin{bmatrix} 1 & 1 & -1 & 0 & 1 \\ 1 & -1 & 1 & 1 & 1 \end{bmatrix}$，此时$n = 5$，$m = 2$

所以基本解为 $(0,0,0,2,4)^T, (0,0,1,0,5)^T, (0,0,-4,10,0)^T, (0,-1,0,0,5)^T, (0,4,0,10,0)^T,$ $(5,0,1,0,0)^T, (4,0,0,2,0)^T, (5,-1,0,0,0)^T$

其中最优解为12，当$x = (0,0,1,0,5)^T$时取得。

b) $Min\, z = 2x_1 + x_2 + 3x_3 + x_4$，$s.t.\begin{cases} x_1 + x_2 + x_3 + x_4 \le 7 \\ 2x_1 - 3x_2 + 5x_3 = -8 \\ x_1 - 2x_3 + 2x_4 \ge 1 \\ x_1, x_3 \ge 0, x_2 \le 0, x_4\text{无约束} \end{cases}$

解： $Max\, z' = -2x_1 + x_5 - 3x_3 - (x_6 - x_7) + 0x_8 + 0x_9$

$s.t.\begin{cases} x_1 - x_5 + x_3 + (x_6 - x_7) + x_8 = 7 \\ -2x_1 - 3x_5 - 5x_3 = 8 \\ x_1 - 2x_3 + 2(x_6 - x_7) - x_9 = 1 \\ x_1, x_3, x_5, x_6, x_7, x_8, x_9 \ge 0 \end{cases}, A = \begin{bmatrix} 1 & 1 & -1 & 1 & -1 & 1 & 0 \\ -2 & -5 & -3 & 0 & 0 & 0 & 0 \\ 1 & -2 & 0 & 2 & -2 & 0 & -1 \end{bmatrix}$

因为第二个式子恒为负，故问题无可行解。

19. 证明：线性规划(LP)问题可行解x为基本可行解的充要条件是x中正分量对应矩阵A的系数列向量线性无关。

证： 不妨设x中正分量为$x_1, x_2, x_3, \dots, x_k (k < n)$，$x$中正分量$x_B$对应矩阵$A$的系数列向量$\alpha_1, \alpha_2, \alpha_3, \dots, \alpha_k$。设约束方程数目为$m (m < n)$

首先证明<u>必要性</u>。若x为基本可行解，则$k = m$，且满足$Ax = x_1\alpha_1 + x_2\alpha_2 + x_3\alpha_3 + \cdots + x_m\alpha_m = b$。假设对应列向量线性相关，即存在一组非零系数$c_1, c_2, c_3, \dots, c_m$，使得$c_1\alpha_1 + c_2\alpha_2 + c_3\alpha_3 + \cdots + c_m\alpha_m = \mathbf{0}$。这可以转化为$A(x - y) = c_1\alpha_1 + c_2\alpha_2 + c_3\alpha_3 + \cdots + c_m\alpha_m = \mathbf{0}$，所以存在另一个$y$与$x$满足同样的线性约束，且它们之间的差异可以表示为矩阵A中对应的系数列向量的线性组合，但是x为基本可行解，不能被表示为其他基本可行解的线性组合，因此假设不成立，x中正分量对应矩阵A的系数列向量线性无关。

而后证明<u>充分性</u>。若x中正分量对应矩阵A的系数列向量线性无关，而x是可行解。从而其非零分量必为正分量，不妨设$x = [x_1 \cdots x_k 0 \cdots 0]^T$，由对应列向量线性无关可知$A$的秩为$k$，因此一定有$k = m$，于是$x$为基本可行解。

20. 讨论线性规划问题与其对偶问题的联系与区别。

答：联系：

1. 对于$LP:\begin{cases} max\ z = c^T x \\ s.t.\ Ax \le b \\ x \ge 0 \end{cases} \Rightarrow DP:\begin{cases} min\ f = b^T y \\ A^T y \ge c \\ y \ge 0 \end{cases}$，对偶问题的对偶是原问题（对称性）。
2. 若x, y分别是 LP,DP 问题的可行解，则$c^T x \le b^T y$（弱对偶定理），即可推出若原规划 LP 问题有最优解，则对偶规划问题 DP 也有最优解，反之亦然，并且两者的目标函数值相等。（主对偶定理）

区别：

1. 原问题与对偶问题目标函数相反，且约束矩阵互为转置。原问题变元由问题本身决定，而对偶问题中的变元有原问题中的约束条件数决定。
2. 原问题直观描述问题本身的建模，而对偶问题通常用于评估原问题的最优解的可行性、灵敏度分析等。

21. 判断下列说法是否正确

a) 考虑具有**有界**可行集的线性规划问题，如果x是一个最优解，则其必定是一个最优基本可行解。

b) 如果线性规划问题有多个解，则其必定有无穷多个解。

c) 考虑标准形式的线性规划问题：$\min c^T x$，$\mathrm{s.t.} Ax = b, x \ge 0$，这里矩阵$A$的维数为$m \times n$，且行满秩。每个最优解上最多有$m$个变量值为正。

d) 对于(c)中的标准线性规划问题，可行域的每个顶点至多有$n - m$个相邻顶点。

答：①错误。若存在无穷多最优解，则最优解不一定为基本可行解。

②正确。考虑若可行域非空，则可行域是个凸集，若有多个解，则一定有无穷多个解。

③错误。由①知最优解不一定为基本可行解，而只有基本解需要满足最多有m个变量值为正的限制，可行解无此限制。

④正确。可行域可视为带约束的多面体，顶点为(LP)的基本可行解，包含m个非零元素，要使顶点相邻，则要有一条边（或面）连接，则显然相邻顶点数最多对应与此向量处于一个平面的基本可行解，即非零元素对应A的列只相差1列，故有$n - m$个相邻顶点。

注：只考虑有界最优解的情况，则可行解是有限的。因而①和③正确，错误仅当存在无穷多最优解时。

22. 生成式与判别式模型的异同是什么？

	生成式模型	判别式模型
特点	对后验概率建模，从统计的角度表示数据的分布情况，能够反映同类数据本身的相似度	寻找不同类别之间的最优分类面，反映的是异类数据之间的差异

区别	估计的是联合概率分布：$p(x,y)$	估计的是条件概率分布:$p(y\|x)$
联系	由产生式模型可以得到判别式模型， 但由判别式模型得不到产生式模型。	
常见模型	朴素贝叶斯，马尔可夫随机场等	逻辑回归，支持向量机（SVM）， 神经网络，k近邻算法等
优点	①实际上带的信息更丰富 ②研究单类问题灵活性强 ③模型可以通过增量学习得到，能用于数据不完整情况。	①分类边界更灵活，能清晰的分辨出类别间差异特征； ②适用于较多类别的识别； ③模型简单，比较容易学习。
缺点	学习和计算过程比较复杂，且容易产生错误信息	不能反映训练数据本身的特性，变量间的关系不可视。
性能	较差	较好

23. 朴素贝叶斯分类器的基本原理是什么？

答：基于属性条件独立假设，假设一组特征$x_1, x_2, x_3, \dots, x_m$相互独立，则由贝叶斯公式，

$$P(y|x_1, x_2, x_3, \dots, x_m) = \frac{P(x_1|y)P(x_2|y)\dots P(x_m|y)P(y)}{P(x_1)P(x_2)\dots P(x_m)} = \frac{P(y)\prod_{i=1}^{m} P(x_i|y)}{\prod_{i=1}^{m} P(x_i)}$$

$P(x_i)$是已知的先验概率，故$P(y|x_1, x_2, x_3, \dots, x_m) \propto P(y)\prod_{i=1}^{m} P(x_i|y)$，建立一个分类模型，我们用已知的类变量的所有可能的值计算概率，并选择输出概率是最大的结果。

$$\hat{y} = arg\max_{y} P(y)\prod_{i=1}^{m} P(x_i|y)$$

24. 贝叶斯准则、最大后验概率准则、最大似然准则有何联系？

答：当知道先验概率$p(x_0)$和$p(x_1)$，则发送一个信号的平均损失C_M可以表示为：

$$C_M = p(x_0)C_0 + p(x_1)C_1 = p(x_0)[(1-\beta)C_{00} + \beta C_{10}] + p(x_1)[(1-\alpha)C_{11} + \alpha C_{01}]$$

$$= p(x_0)C_{00} + p(x_1)C_{11} + p(x_0)\cdot\beta(C_{10} - C_{00}) + p(x_1)\cdot\alpha(C_{01} - C_{11})$$

$$= p(x_0)C_{10} + p(x_1)C_{11} + \int_{s_{y_0}} [(C_{01} - C_{11})\, p(x_1)p(y|x_1) - (C_{10} - C_{00})p(x_0)p(y|x_0)]\, dy$$

找判决阈值等价于最小化分险，即：$\min\left\{\int_{s_{y_0}} [B(y) - A(y)]\, dy\right\}$，等价于$max\left\{\int_{-\infty}^{y} [A(y) - B(y)]\, dy\right\}$，**显然**，当$A(y) = B(y)$是满足条件，即

$$(C_{01} - C_{11})\, p(x_1)p(y|x_1) = (C_{10} - C_{00})p(x_0)p(y|x_0)$$

①**贝叶斯准则**即为<u>损失函数、转移概率分布和事件的先验概率都已知</u>时，

$$y = \begin{cases} 1, \dfrac{p(y|x_1)}{p(y|x_0)} > \dfrac{(C_{10} - C_{00})p(x_0)}{(C_{01} - C_{11})p(x_1)} \\ 0, \dfrac{p(y|x_1)}{p(y|x_0)} \le \dfrac{(C_{10} - C_{00})p(x_0)}{(C_{01} - C_{11})p(x_1)} \end{cases}$$

令$L(y)=\frac{p(y|x_1)}{p(y|x_0)}$为似然比，$\lambda=\frac{(C_{10}-C_{00})p(x_0)}{(C_{01}-C_{11})p(x_1)}$为检测阈值，则条件转化为$L(y)\gtreqless\lambda$来判断接收的类别。

②而**最大后验概率准则**为损失函数未知时，简单假设错误判决的损失为1而正确判决的损失为0，即$C_{00}=C_{11}=0$，$C_{01}=C_{10}=1$，此时$\lambda=\frac{(C_{10}-C_{00})p(x_0)}{(C_{01}-C_{11})p(x_1)}=\frac{p(x_0)}{p(x_1)}$，而后条件转化为$\frac{p(y|x_1)}{p(y|x_0)}\gtreqless\frac{p(x_0)}{p(x_1)}\Leftrightarrow\frac{p(y|x_1)}{p(y|x_0)}=\frac{p(x_1|y)p(y)}{p(x_1)}\cdot\frac{p(x_0)}{p(x_0|y)p(y)}\gtreqless\frac{p(x_0)}{p(x_1)}\Leftrightarrow\frac{p(x_1|y)}{p(x_0|y)}\gtreqless 1$来判断接收的类别。

③**最大似然准则**当损失函数和先验概率均未知时，假定两种判决出错的概率和代价基本相同，假定常数$\lambda=\frac{(C_{10}-C_{00})p(x_0)}{(C_{01}-C_{11})p(x_1)}=1$，此时条件转化为$L(y)\gtreqless 1$来判断接收的类别。

25. 通信系统的三个基本特性是为了确保通信的__有效__性、__可靠__性与__安全__性。

26. 无噪信道输入为X,输出为Y，则其互信息$I(X;Y)=$?

答：$I(X;Y)=H(X)-H(X|Y)$

$I(X;Y)=H(Y)-H(Y|X)$

$I(X;Y)=H(X)+H(Y)-H(XY)$

27. 二分类（0/1）问题的交叉熵如何定义？

答：$H(p,q)=-p(x)\log\big(q(x)\big)-p(1-x)\log\big(q(1-x)\big)$

28. KL 散度非负，是否正确？

答：正确。$\log\left(\frac{q_i}{p_i}\right)=\log e\cdot\ln\left(\frac{q_i}{p_i}\right)\le\log e\cdot\left(\frac{q_i}{p_i}-1\right)$, 从而：

$$\sum_i p_i\log\left(\frac{q_i}{p_i}\right)\le\sum_i p_i\log e\cdot\left(\frac{q_i}{p_i}-1\right)=\log e\cdot\sum_i(q_i-p_i)=0$$

故 KL 散度$D(p||q)=\sum_{x\in X}p(x)\log\left(\frac{p(x)}{q(x)}\right)\ge 0$恒成立，当且仅当$\frac{p_i}{q_i}=1$等号成立。

29. 有约束问题的拉格朗日函数是什么？

$$L(x,u,v)=f(x)+\Sigma_{i=1}^{m}u_i g_i(x)+\Sigma_{i=1}^{l}v_i h_i(x)$$

其中，u,v分别为对应不等式约束和等式约束的拉格朗日乘子向量或者对偶变量。

30. 梯度下降和牛顿法的基本原理是什么？

答：根据函数的逼近思想

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

则可以导出$x_{k+1}=x_k-\lambda_k(B^k)^{-1}\nabla f(x_k)$，当$B^k$为常数倍单位矩阵时可以导出梯度下降

法；B^k为海瑟矩阵时可以导出牛顿法。

①梯度下降法(Gradient Descent) 的基本思想是通过迭代地更新模型参数，沿着目标函数的负梯度方向，按照一定步长逐步逼近最优解。

②而牛顿法(Newton's method)则是利用目标函数的一阶导数（梯度）和二阶导数（海森矩阵）信息在每一次迭代中近似目标函数的局部曲线，来进行参数更新，从而逐步逼近最优解。

31. PCA 的基本原理？

答： PCA 的主要思想是从原始的空间中顺序地找一组相互正交的坐标轴，将n维特征映射到全新的k维正交特征向量上。而对于k维正交特征向量的选择，一般选择协方差矩阵对应特征值最大的k个特征向量，这些特征向量所形成的矩阵，可以看成高维空间中的刚性旋转，将矩阵应用到数据的时候，对应主特征向量的方向就是数据方差最大的方向。

32. SVD 的基本原理和计算方法？

答： 基本原理：对于矩阵$B(m \times n)$，利用矩阵B^TB与BB^T具有相同特征值的性质。并且BB^T与B^TB的特征向量(u_i, v_i)之间满足关系：$u_i\sqrt{\lambda_i} = Bv_i$，因而有$[u_1, \cdots, u_m]\Sigma = \mathrm{B}[v_1, \cdots, v_n] \Rightarrow U\Sigma = BV$,且$U, V$都是单位正交矩阵,$UU^T = VV^T = I$, 从而有$B = U\Sigma V^T$。

计算方法：对于方阵$(n \times n)$的情况下。

①由B^TB的特征分解获得特征向量及对应特征值$\lambda_1, \cdots, \lambda_n$，则

$$\Sigma = \begin{bmatrix} \sqrt{\lambda_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sqrt{\lambda_n} \end{bmatrix}, V = [v_1, v_2, \ldots, v_n]$$

②计算$U = BV\Sigma^{-1}$，即可将B分解为U, Σ, V。

而对于非方阵的情况下，假设矩阵D是由矩阵$B_{m \times n}$填充0而成的$d \times d$的方阵($d = \max\{m, n\}$)，这时按照分块矩阵的处理可以获得非方阵的 SVD 分解。

$$D = [B\ 0] = U\Sigma V^T = U\begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} V_1 & 0 \\ 0 & V_2 \end{bmatrix}^T, m > n$$

$$D = \begin{bmatrix} B \\ 0 \end{bmatrix} = U\Sigma V^T = \begin{bmatrix} U_1 & 0 \\ 0 & U_2 \end{bmatrix}\begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix}V^T, m < n$$

33. RPCA 的基本原理？

答： 鲁棒主成分分析(Robust Principal Component Analysis,RPCA)可以将观测的高维数据分解为低秩成分和稀疏成分，能够处理含有异常值或噪声的数据的降维。

若N为满足稀疏约束的噪声矩阵，问题可以形式化为：$\min\limits_{L,N} rank(L) + \lambda\|N\|_0, s.t.\ M = L + N$，而在稀疏建模中，$l_1$范数是$l_0$范数的最佳凸松弛，而矩阵核范数是$rank(\cdot)$函数

的最佳凸松弛，从而问题可以变为$\min_{L,N}||L||_*+\lambda||N||_1, s.t.\ M=L+N$，从而可以通过拉格朗日乘子法、矩阵奇异值分解(SVD)以及软阈值函数$T_\varepsilon(M)$进行迭代求解。

34. 二维情况下的L_1，L_0和L_2的几何解释是什么？

答： 对于二维坐标(x,y)，

1） $L_1=|x|+|y|$，对于某一个给定的L_1，其几何解释是一个以原点为中心的正方形，顶点在坐标轴上，对角线长度为$2L_1$。

2） L_0表示x和y不为0的个数，即(x,y)在原点$L_0=0$,在坐标轴上$L_0=1$，其余位置$L_0=2$。

3） $L_2=\sqrt{x^2+y^2}$，对于某一个给定的L_2，其几何解释是一个以原点为中心的圆形，半径长等于L_2。

35. 矩阵的 spark 是指什么？

答： 给定矩阵A的spark 是指矩阵A中线性相关的最小列的数目。Spark 可以用来使用l_0范数来刻画矩阵的零空间，如果向量在$Ax=0$ 的零空间中，必须满足$||x||_0\geq spark(A)$。

36. 稀疏表示中的 OMP 算法基本流程是什么？

答： 任务：获取(P_0): $\min_x ||x||_0, s.t.\ Ax=b$的近似解

ALGORITHM Orthogonal Matching Pursuit (OMP)

1：**input** A, B, ε
2：$k=0, x^0=\mathbf{0}, r^0=b-Ax^0=b, S^0=Support\{x^0\}=\emptyset$;
3：**while** $\|r^k\|_2>\varepsilon$ **do**
4： $k=k+1$;
5： ***Sweep***:计算误差$\epsilon(j)=\min_{z_j}||a_jz_j-r^{k-1}||_2^2$,其中$z_j$用$z_j^*=\frac{a_j^Tr^{k-1}}{||a_j||_2^2}$代入；
6： **更新支撑**：*求*$\epsilon(j)$*的最小指标*j_0: $\forall j\notin S^{k-1}, \epsilon(j_0)\leq\epsilon(j)$，更新$S^k=S^{k-1}\cup\{j_0\}$
7： **更新临时解**：计算$x^k=minize_x||Ax-b||_2^2, s.t.\ Support\{x\}=S^k$；
8： **更新残差**:计算$r^k=b-Ax^k$;
9：**end while**
10：**return** x^k

37. 稀疏表示中的测不准原理是指什么？

答： ①一般的不确定原理是指由傅立叶变换耦合的一对共轭变量不能都具有任意高精度，（不能同时在时域和频域都能高精度确定信号），即对函数$f(x)$，以及其傅立叶变换$F(\omega)$必定满足不等式：

$$\int_{-\infty}^{\infty}x^2|f(x)|^2dx\cdot\int_{-\infty}^{\infty}\omega^2|F(\omega)|^2d\omega\geq\frac{1}{2}, 假设\int_{-\infty}^{\infty}|f(x)|^2dx=1$$

②互相关性$\mu(A)$的一对正交基Ψ,Φ，任意非零向量$b\in\mathbb{R}^n$在基上的表示分别为α,β,则有：

$$||\alpha_0||+||\beta||_0\geq\frac{2}{\mu(A)}$$

③线性系统$[\Psi,\Phi]x=b$的任意两个不同解不能都是稀疏的，有以下的测不准原理：

$$||x_1||_0+||x_2||_0\geq\frac{2}{\mu(A)}$$

则线性系统$[\Psi,\Phi]x=b$的一个候选解有小于$\frac{1}{\mu(A)}$个非零值，则该解就是最稀疏的解，并且任何其他解都必定是稠密的。