**数据包络（DEA）分析法（备用）**

数据包络分析是评价多输入指标和多输出指标的较为有效的方法，将投入与产出进行比较。它的结果包含的意思有：

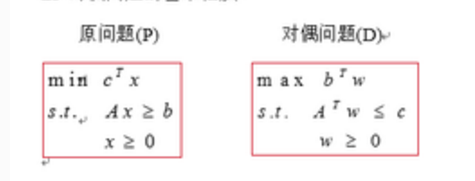
①θ=1，DEA有效，表示投入与产出比达到最优

②θ<1，非DEA有效，表示投入与产出比没有达到最优，一般来说，θ越大说明效果越好。

数据包络分析是通过对投入的指标和产出的指标做了一个线性规划，并且进行变换后，然后根据其线性规划的对偶问题（线性规划对偶问题具有经济学意义），求解这个对偶问题的最值就是θ。

---------------------------------------------------------------------

PS什么是线性规划的对偶形式：数学专业会学



就是min与max互换，目标函数的系数C与约束条件中的b互换，约束条件的大于号与小于号互换。

---------------------------------------------------------------------

注意：看这里时最好打开司老师的书，381页，381的推导过程不是求所有的决策单元的效率的数学模型，而是求某个j0评价决策单元（就是某个个体，在司老师中的的例子是某个学校）的效率的数学模型。最后的θ也是针对这个学校的θ值，但是求解的时候用到了所有学校的数据。

数据包络分析在数学建模比赛中用的很少，出现的比较多的就是2016年美赛C题，很多队伍用，当时题目背景和司老师的书的题目背景一样。但是数据包络分析的思想还是很好的，把评价问题分为投入与产出两部分，这样的思想如果做某些评价问题还是很好的，给大家推荐一个这样的例子（没有用数据包络分析，但是用了投入与产出对比的思想：《美国大学生数学建模竞赛题解析与研究》第一辑第三章医疗评价模型二：输入输出系统评估模型），下载链接见回归分析的那本。

--------------------同学整理的数据包络分析介绍----------------------

数据包络分析(data envelopment analysis,DEA)是一个对多投入\多产出的多个决策单元的效率评价方法。是1978年由CHARNES和COOPER创建的。可广泛使用于业绩评价。

DEA特别适用于具有多输入多输出的复杂系统，这主要体现在以下几点：

1. DEA以决策单位各输入/输出的权重为变量，从最有利于决策单元的角度进行评价，从而避免了确定各指标在优先意义下的权重；
2. 假定每个输入都关联到一个或者多个输出，而且输入/输出之间确实存在某种关系，使用DEA方法则不必确定这种关系的显示表达式。

-------------------------------------------------------------------------------------------------------

例 利用DEA方法对天津市的可持续发展进行评价。在这里选取较具代表性的指标作为输入变量和输出变量，见表1。

表1 各决策单元输入、输出指标值

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 序号 | 决策  单元 | 政府财政收入占  GDP的比例/% | 环保投资占GDP  的比例/% | 每千人科技  人员数/人 | 人均GDP  /元 | 城市环境  质量指数 |
| 1 | 1990 | 14.40 | 0.65 | 31.30 | 3621.00 | 0.00 |
| 2 | 1991 | 16.90 | 0.72 | 32.20 | 3943.00 | 0.09 |
| 3 | 1992 | 15.53 | 0.72 | 31.87 | 4086.67 | 0.07 |
| 4 | 1993 | 15.40 | 0.76 | 32.23 | 4904.67 | 0.13 |
| 5 | 1994 | 14.17 | 0.76 | 32.40 | 6311.67 | 0.37 |
| 6 | 1995 | 13.33 | 0.69 | 30.77 | 8173.33 | 0.59 |
| 7 | 1996 | 12.83 | 0.61 | 29.23 | 10236.00 | 0.51 |
| 8 | 1997 | 13.00 | 0.63 | 28.20 | 12094.33 | 0.44 |
| 9 | 1998 | 13.40 | 0.75 | 28.80 | 13603.33 | 0.58 |
| 10 | 1999 | 14.00 | 0.84 | 29.10 | 14841.00 | 1.00 |

输入变量：政府财政收入占GDP的比例、环保投资占GDP的比例、每千人科技人员数。输出变量：经济发展（用人均GDP表示）、环境发展（用城市环境质量指数表示；计算过程中，城市环境指数的数值作了归一化处理）。

MATLAB程序为：

clc,clear

format long

load('data.txt');%把原始数据保存在纯文本文件data.txt中

X=data(:,[1:3]);%X为输入变量，3为输入变量的个数

X=X';

Y=data(:,[4:5]);%Y为输出变量，5（3+2），2为输出变量的个数

Y=Y';

n=size(X',1);m=size(X,1);s=size(Y,1);

A=[-X' Y'];

b=zeros(n,1);

LB=zeros(m+s,1);UB=[];

for i=1:n;

f=[zeros(1,m) -Y(:,i)'];

Aeq=[X(:,i)',zeros(1,s)];beq=1;

w(:,i)=linprog(f,A,b,Aeq,beq,LB,UB);

E(i,i)=Y(:,i)'\*w(m+1:m+s,i);

end

theta=diag(E)';

fprintf('用DEA方法对此的相对评价结果为：\n');

disp(theta);

omega=w(1:m,:)

mu=w(m+1:m+s,:)

——————————————————————————————————

补充：

定义1 若线性规划问题的最优目标值theta=1，则称决策单元是弱DEA有效地。

定义2 若线性规划问题存在最优解omega>0，mu>0，并且其最优目标值theta=1，则称决策单元是DEA有效的。

从上述定义可以看出，所谓DEA有效，就是指那些决策单元，它们的投入产出比达到最大。因此，可以用DEA来对决策变量进行评价。

——————————————————————————————————

计算结果见表2，最优目标值用表示。显而易见，该市在20世纪90年代的发展是朝着可持续方向前进的。

表2 用DEA方法对天津市可持续发展的相对评价结果

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 年份 |  | 结论 | 年份 |  | 结论 |
| 1990 | 0.2901843 | 非DEA有效 | 1995 | 0.7182609 | 非DEA有效，规模收益递增 |
| 1991 | 0.2853571 | 非DEA有效，规模收益递减 | 1996 | 0.9069108 | 非DEA有效，规模收益递增 |
| 1992 | 0.2968261 | 非DEA有效，规模收益递增 | 1997 | 1 | DEA有效，规模收益递增 |
| 1992 | 0.3425151 | 非DEA有效，规模收益递增 | 1998 | 1 | DEA有效，规模收益不变 |
| 1994 | 0.4594712 | 非DEA有效，规模收益递增 | 1999 | 1 | DEA有效，规模收益不变 |

后面三个不是1，只是非常接进1，实际运算结果见下图：

