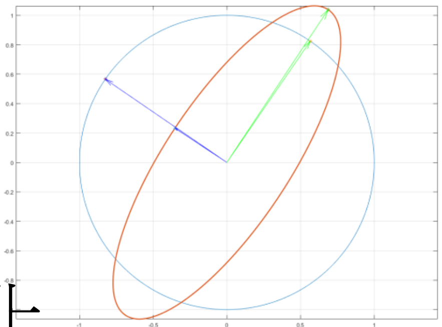
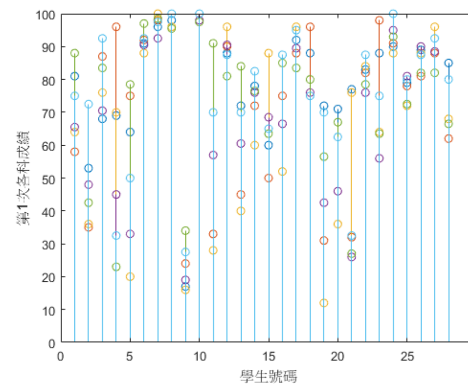
# 奇異值分解

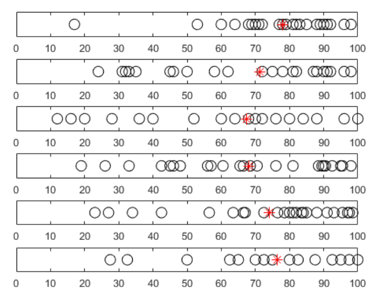
Singular Value Decomposition






- 最佳化問題
  單位圓(球)最佳化

- 應用
  學生成績分析



2018年

國立高雄大學　　應用數學系

數值方法專題　　指導教授：郭岳承　教授

# 最佳化問題

單位圓(球)最佳化

# 構想 Idea

- $2-norm(A)$與奇異值與$\mathrm{M}ax\|A\vec{x}\|$的關係？

  策略 Strategy：

  使用matlab畫出一單位圓，即 $\vec{v}=(sin\theta, cos\theta), \|\vec{v}\|_2=1$

  再經過線性轉換矩陣 A 的轉換後，求其 $\mathrm{M}ax\|A\vec{v}\|$ ，觀察之

- $A*A^T$ 的 eigenvector與$\|A\vec{v}\|$之$\vec{v}$向量之關係？

  策略 Strategy ：

  計算出 $\mathrm{M}ax\|A\vec{v}\|$時，記錄線性轉換前之向量$\vec{x}$，在計算$A*A^T$ 之

  eigenvector比較之

# 程式碼 Coding

```
format long e
%% Generate Unit circle
Num = 10^(6) ;
% linspace divide into (Num) interval
a = linspace (0 , 2*pi , Num);

% polar coordinates
x = cos(a);
y = sin(a);

%% A transpose unit circle
A = rand(2);

Ax = A(1,1) * x + A(1,2) * y;
Ay = A(2,1) * x + A(2,2) * y;

```

極座標系
Polar coordinate system
下討論曲線變化比較方便
令

$$\vec{v} = (sin\theta, cos\theta)$$

為一單位圓
切割其角度$\theta$為 Num 等分

隨機生成A矩陣
經過線性轉換後的

$$(a(1,1) * sin\theta, a(1,2) * cos\theta)$$
$$(a(2,1) * sin\theta, a(2,2) * cos\theta)$$

# 程式碼 Coding

```
%% Norm of Ax and A
Max = 0;
Min = 100;
for i = 1 : Num
    n = ( 2*pi / Num ) * i;
    q = cos(n);
    w = sin(n);

    q2 = A(1,1) * q + A(1,2) * w;
    w2 = A(2,1) * q + A(2,2) * w;

    Nv = ( q2 )*( q2 )' + ( w2 )*( w2 )' ;

    % 紀錄最大
    if(Nv > Max)
        Max = Nv ; % 最大||Ax||^2

        MaxX = cos(n);
        MaxY = sin(n);
        MaxV = [MaxX MaxY];

        AMaxX = A(1,1) * MaxX + A(1,2) * MaxY; % 線性轉換後 x 點座標
        AMaxY = A(2,1) * MaxX + A(2,2) * MaxY; % 線性轉換後 y 點座標

        AMaxV = [AMaxX AMaxY];
    end
```

尋訪變化角度$\theta$

存在一角度$\vartheta$ s.t

$$\vec{v} = (sin\vartheta, cos\vartheta)$$

$$\mathrm{A}\vec{v} = \begin{pmatrix} a(1,1) * \sin\vartheta + a(1,2) * cos\vartheta, \\ a(2,1) * sin\vartheta + a(2,2) * cos\vartheta \end{pmatrix}$$

$\|A\vec{v}\|$為最大

# 程式碼 Coding

```
% 記錄最小
if(Nv < Min )
    Min = Nv ;

    MinX = cos(n);
    MinY = sin(n);
    MinV = [MinX MinY];

    AMinX = A(1,1) * MinX + A(1,2) * MinY;
    AMinY = A(2,1) * MinX + A(2,2) * MinY;

    AMinV = [AMinX AMinY];
end
```

存在一角度ω s. t

$$\vec{u} = (sin\omega, cos\omega)$$

$$A\vec{u} = \begin{pmatrix} a(1,1) * \sin\omega + a(1,2) * cos\omega, \\ a(2,1) * sin\omega + a(2,2) * cos\omega \end{pmatrix}$$

$\|A\vec{u}\|$為最小

# 程式碼 Coding

比較 compare

```
%% Max||Ax|| 與 ||A|| 的關係
%% Q1 : Max|| Ax || = norm(A) ?
NA = norm( A ) ;
SMax = sqrt( Max );
sigma = svd (A) ;

[SMax , NA , sigma(1)]  % sigma(1) = norm( A ) = sqrt( Max )

```

$Norm(A)$、$\mathrm{Max}\|A\vec{v}\|$、$\sigma 1$ 之關係

```
>> [SMax , NA , sigma(1)]

ans =

     1.245223370350562e+00     1.245223370353542e+00     1.245223370353542e+00
```

得知：$2-norm(A)=\mathrm{Max}\|A\vec{v}\|=\sigma 1$

# 程式碼 Coding

比較 compare

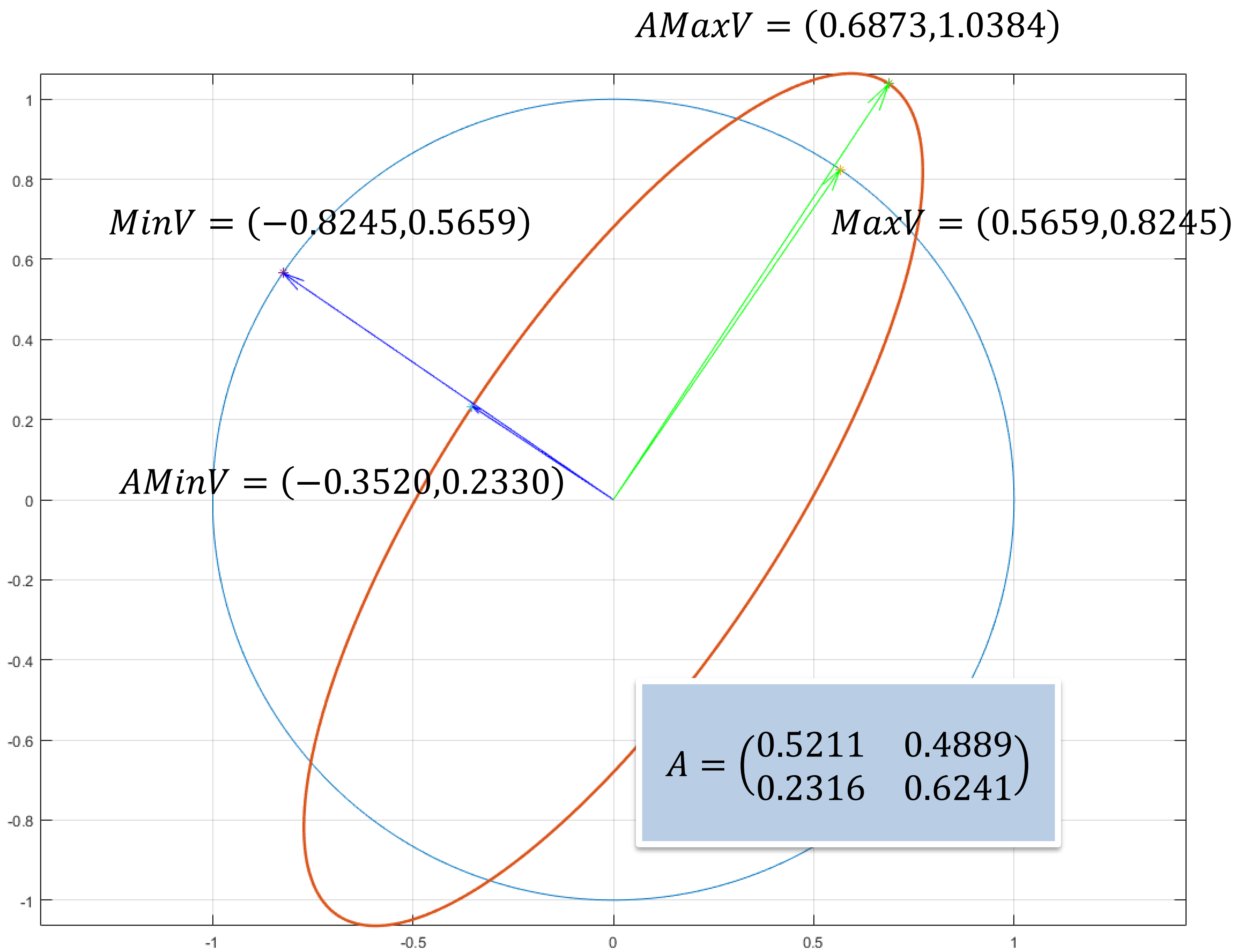
```
%% Q2 : eigvetor(A'*A) 與 x 在最大值時之關係 ?

[AV , AD] = eig( A'*A ) ;

AV
[MinV' , MaxV']

```

```
[MinV' , MaxV']

AV =

    -8.244888898089262e-01     5.658781411060552e-01
     5.658781411060552e-01     8.244888898089262e-01



ans =

    -8.244902057245385e-01     5.658762238019091e-01
     5.658762238019091e-01     8.244902057245385e-01
```

$A * A^T$ 的 eigenvector 與

$\mathrm{M}ax\|A\vec{v}\|$和$\mathrm{M}in\|A\vec{u}\|$的向量$\vec{v}$ 、$\vec{u}$相同

# 單位圓在線性轉換下的最佳化




#圖為15位浮點數計算下，而途中數值為四捨五入制小數後4位

# 推估

- 矩陣$\boldsymbol{A_{nm}}$而 $\|\boldsymbol{A}\|_{\boldsymbol{2}} = \underset{\|\vec{\boldsymbol{x}}\|=\boldsymbol{1}}{\boldsymbol{max}}\|\boldsymbol{A\vec{x}}\|_{\boldsymbol{2}}$

  證明方向：

  使用Lagrange Multipliers

  constraints $\|\vec{x}\|_2 = 1$

證明

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{pmatrix}$$ is an n-by-m matrix

$$2 - norm(A) = \max_{x \neq 0} \frac{\|A\vec{x}\|_2}{\|\vec{x}\|_2}$$

Let $\|\vec{x}\|_2 = 1$ Then $\|A\|_2 = \max_{\|\vec{x}\|=1} \|A\vec{x}\|_2$

$\|A\vec{x}\|_2 = (A\vec{x})^t A\vec{x} = \vec{x}^t(A^t A)\vec{x}$ $\|\vec{x}\|_2 = \vec{x}^t\vec{x}$

$\vec{x}^t = (x_1, x_2, \dots \dots, x_m)$

f$(x_1, x_2, \dots \dots, x_m) = \vec{x}^t(A^t A)\vec{x}$
g$(x_1, x_2, \dots \dots, x_m) = {x_1}^2 + {x_2}^2 + \cdots \dots + {x_m}^2$

Let $A^t A = \begin{pmatrix} a'_{11} & \cdots & a'_{1m} \\ \vdots & \ddots & \vdots \\ a'_{m1} & \cdots & a'_{mm} \end{pmatrix}$

證明

$$f(x_1, x_2, \ldots\ldots, x_m)$$

$$= (x_1, x_2, \ldots\ldots, x_m)\begin{pmatrix} a'_{11} & \cdots & a'_{1m} \\ \vdots & \ddots & \vdots \\ a'_{m1} & \cdots & a'_{mm} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ . \\ . \\ . \\ x_m \end{pmatrix}$$

$$= \left(\sum_{i=1}^{m} a'_{i1} * x_i \,, \ldots\ldots, \sum_{i=1}^{m} a'_{im} * x_i\right)\begin{pmatrix} x_1 \\ x_2 \\ . \\ . \\ . \\ x_m \end{pmatrix}$$

$$= \sum_{i=1}^{m} a'_{i1} * x_i * x_1 + \cdots + \sum_{i=1}^{m} a'_{im} * x_i * x_m$$

證明 $\nabla \mathrm{f}(x_1, x_2, \dots\dots, x_m)$

$$= \begin{pmatrix} \partial f/\partial x_1 \\ \partial f/\partial x_2 \\ \vdots \\ \partial f/\partial x_m \end{pmatrix}$$

$$= \begin{pmatrix} (2a'_{11}x_1 + a'_{12}x_2 + \cdots + a'_{1n}x_n) + (a'_{21}x_2 + \cdots + a'_{n1}x_n) \\ (a'_{21}x_1 + 2a'_{22}x_2 + \cdots + a'_{2m}x_m) + (a'_{12}x_1 + a'_{32}x_3 \dots + a'_{m2}x_m) \\ \vdots \\ (a'_{m1}x_1 + a'_{m2}x_2 + \cdots + 2a'_{mm}x_m) + (a'_{1m}x_m + \cdots + a'_{m-1m-1}x_{m-1m-1} \end{pmatrix}$$

$$= \begin{pmatrix} (2a'_{11})x_1 + (a'_{12} + a'_{21})x_2 + \cdots + (a'_{1n} + a'_{n1})x_n \\ (a'_{21} + a'_{12})x_1 + (2a'_{22})x_2 + \cdots + (a'_{2m} + a'_{m2})x_m \\ \vdots \\ (a'_{m1} + a'_{1m})x_1 + (a'_{m2} + a'_{2m})x_2 + \cdots + (2a'_{mm})x_m \end{pmatrix}$$

$$= \begin{pmatrix} a'_{11} & \cdots & a'_{1m} \\ \vdots & \ddots & \vdots \\ a'_{m1} & \cdots & a'_{mm} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} + \begin{pmatrix} a'_{11} & \cdots & a'_{m1} \\ \vdots & \ddots & \vdots \\ a'_{1m} & \cdots & a'_{mm} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix}$$

證明

Since $A^t A$ is symmetric matrix ( $(A^t A)^t = AA^t$ )

$$\nabla \mathrm{f}(x_1, x_2, \dots \dots, x_m) = (A^t A + AA^t)\vec{x} = 2(A^t A)\vec{x}$$

$$\nabla \mathrm{g}(x_1, x_2, \dots \dots, x_m) = \begin{pmatrix} \partial g/\partial x_1 \\ \partial g/\partial x_2 \\ \vdots \\ \partial g/\partial x_m \end{pmatrix} = 2\vec{x}$$

Use **Lagrange Multipliers**

$$\nabla \mathrm{f}(x_1, x_2, \dots \dots, x_m) = \lambda \nabla \mathrm{g}(x_1, x_2, \dots \dots, x_m)$$

Then $(A^t A)\vec{x} = \lambda \vec{x}$

We can conclude

If $\|\vec{x}\|_2 = 1$ $\|A\|_2 = \max_{\|\vec{x}\|=1} \|A\vec{x}\|_2$

The eigenvector of $(A^t A)$ satisfy $\max_{\|\vec{x}\|=1} \|A\vec{x}\|_2 = \|A\|_2$

# 最佳化

近似矩陣

# 證明

$$\|A\|_f = \sum_{ij} {a^2}_{ij} = (tr(A^t A))^{1/2}$$

$SVD$

$\mathrm{A} = \mathrm{USV}^t \quad rank(A) = r$

$S = diag(\sigma_1, \dots, \sigma_r) \quad \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$

求近似矩陣M，最接近矩陣A

$rank(M) = k < r$

$$\min_{rank(M) \leq r} \|A - M\|_f$$

$$= \min_{rank(M) \leq r} \|U^t(A - M)\|_f = \min_{rank(M) \leq r} \|U^t(A - M)V\|_f$$

$$= \min_{rank(M) \leq r} \|U^t AV - U^t MV\|_f = \min_{rank(M) \leq r} \|S - U^t MV\|_f$$

$D = diag(d_1, \dots, d_k)$

# 證明

$$\min_{rank(M)\le r} \|A - M\|_f$$

$$= \min_D \|S - D\|_f = \min_D (\sum_{i=1}^{r} (\sigma_i - d_i))^{\frac{1}{2}}$$

最佳矩陣M

$D = diag(d_1, \dots, d_k, 0, \dots, 0)$

$d_1 \ge d_2 \ge \cdots \ge d_k > 0 = d_{k+1} = \cdots = d_r$

最小誤差

$\min \|A - M\|_f = ({\sigma_{k+1}}^2 + \cdots + {\sigma_r}^2)^{1/2}$

# 應用

學生成績分析

# 構想

# Idea

- 藉由SVD，分解原班級成績形成的A矩陣，取得一近似矩陣M，分析其班級成績

  策略 Strategy：

  分解A矩陣，分析奇異值，取得最近似矩陣M

  運用最近似矩陣的左右奇異向量，分析學生成績能與班級強項組成

  再輔以基礎統計，佐以分析班級組成

# 程式碼 Coding

```
%% Select the N-th test (1 to 5)
N = 1 ;

%% Sort the data
for j = 1 : 5
    eval(['B',int2str(j),' = A',int2str(j),''';']);
    eval(['B',int2str(j),'=sort(B',int2str(j),',2);']);
end

```

因為有五筆考試資料

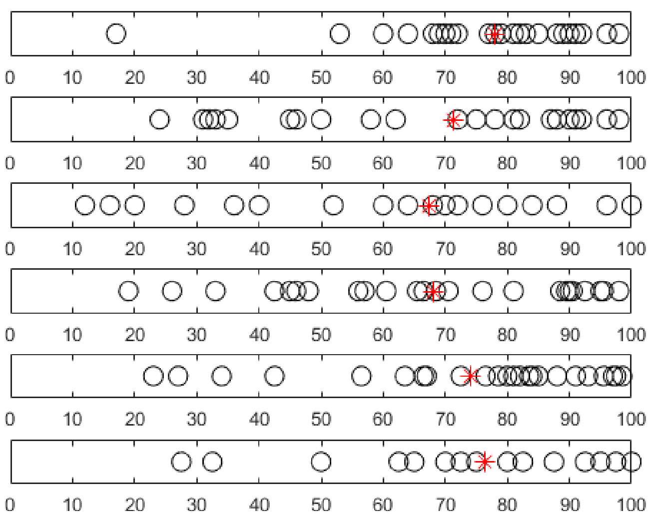
為方便調閱與觀察五次考試

故控制參數　N　以表示第N次考試

# 程式碼 Coding

## 圖形化 Graph

```
%% Graph the N-test Subject Grade in class
clf
for k = 1 : 6

    eval(['data1 = B',int2str(N),'(',int2str(k),',:);']);      %# Sample data set 1
    eval(['data2 = sum(B',int2str(N),'(',int2str(k),',:))/28;']);      %# Sample data set 2
    eval(['subplot( 6,1,',int2str(k),' );plot(data1,0,''ko'',''MarkerSize'',10);']);hold on   %# Plot data set 1
    eval(['subplot( 6,1,',int2str(k),' );plot(data2,0,''r*'',''MarkerSize'',10);']);hold on   %# Plot data set 2
    xlim([0 100]);
    set(gca,'ytick',[])
end
```




接著印出六科成績

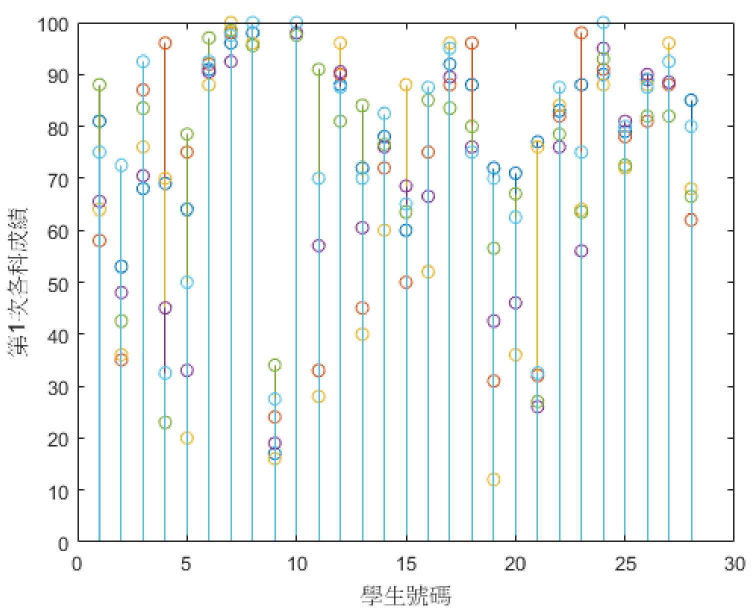
並在各科的平均位置坐上標記

觀察整體成績走向

# 程式碼 Coding

圖形化 Graph

```
%% Graph the N-test Students' Grade
figure ;
eval(['stem(A',int2str(N),');']);
xlabel('學生號碼');
ylabel(['第',int2str(N),'次各科成績']);
```




接著印出各個學生各科成績的分布

直接觀察出極端值的存在

成績最糟的9號與成績最好的10號

還有

各科成績差距非常大的4號

# 程式碼 Coding

奇異值分解 SVD

```
%% SVD
for i = 1:5
    eval(['[U',int2str(i),',S',int2str(i),',V',int2str(i),']','= svd(A',int2str(i),');']);
    eval(['U',int2str(i),'= -U',int2str(i),';']);
    eval(['V',int2str(i),'= -V',int2str(i),';']);
end
```

針對各次考試的SVD分解

為了方便觀察左右奇異向量變號

# 程式碼 Coding

近似矩陣M

```
%% Rank(1) matrix M
eval(['sigma = diag(S',int2str(N),') ;']);
tau = cumsum( sigma.^2 ) / sum ( sigma.^2 ) ;
figure;
plot ( tau( 1 : 6 ), '.' );
```

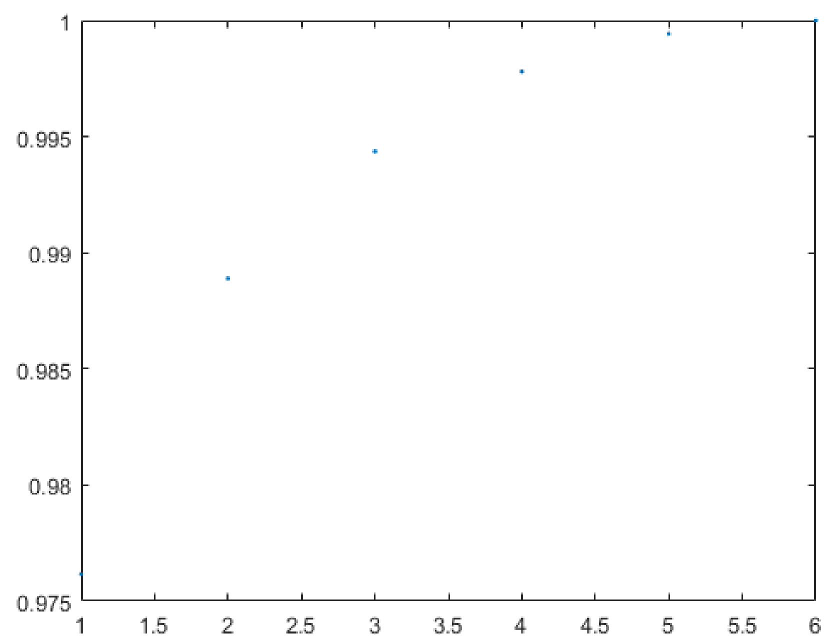
求近似矩陣A最近似矩陣M

分析奇異值

發現第一個奇異值下的近似矩陣
便有97.5%以上近似

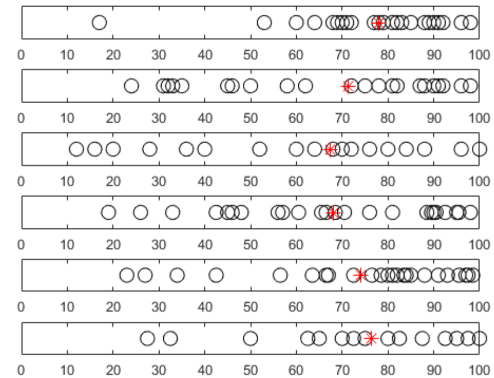
代表rank(1)下的近似矩陣M能

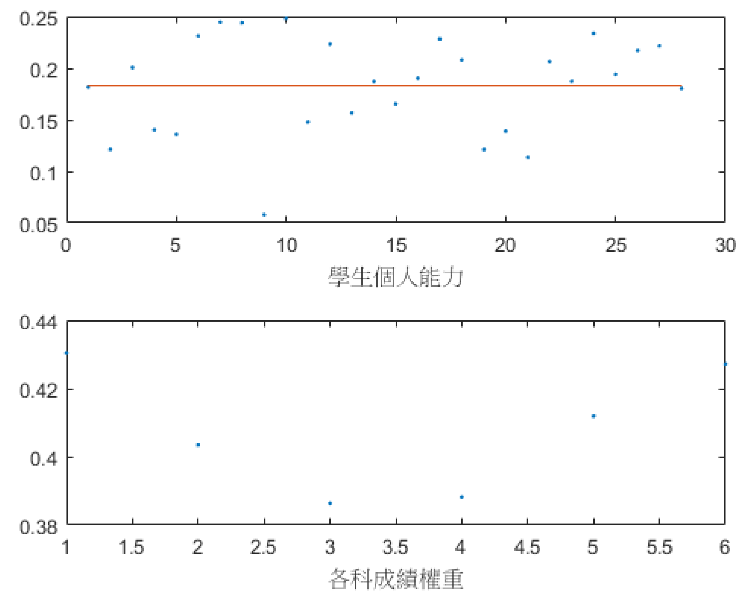
最代表原矩陣A



# 程式碼 Coding

```
eval(['U',int2str(N),'A = sum(U',int2str(N),'(:,1))/28 ;']); % average ability

figure;
eval(['subplot( 2,1,1 ) ,plot( U',int2str(N),'( : ,1 ),''.'');']);hold on
eval(['subplot( 2,1,1 ) ,plot( [1,28],[U',int2str(N),'A,U',int2str(N),'A] ,''-'');
xlabel('學生個人能力');
eval(['subplot( 2,1,2 ) ,plot( V',int2str(N),'( : ,1 ),''.'');']);hold off
xlabel('各科成績權重');

eval(['Z = V',int2str(N),''';']);
eval(['M',int2str(N),'=U',int2str(N),'(:,1)*sigma(1)*Z(1,:);']);

```




根據最大值10號小孩
最小值 9號小孩
推估**左奇異向量U(Column1)**代表**學生能力指標**

根據班級整體科目平均
推估**右奇異向量V(Column1)**代表**班級成績指標**

# 分析

# analysis

- 用SVD跟基礎統計量分析

  發現左奇異向量U跟個人成績會有相關，代表著學生能力指標

  發現右奇異向量V跟班級各科平均會有相關，代表著班級成績指標

- 誤差

  但是經由這種降維的近似矩陣M，是不是有些學生的成績可能會有所誤差呢？

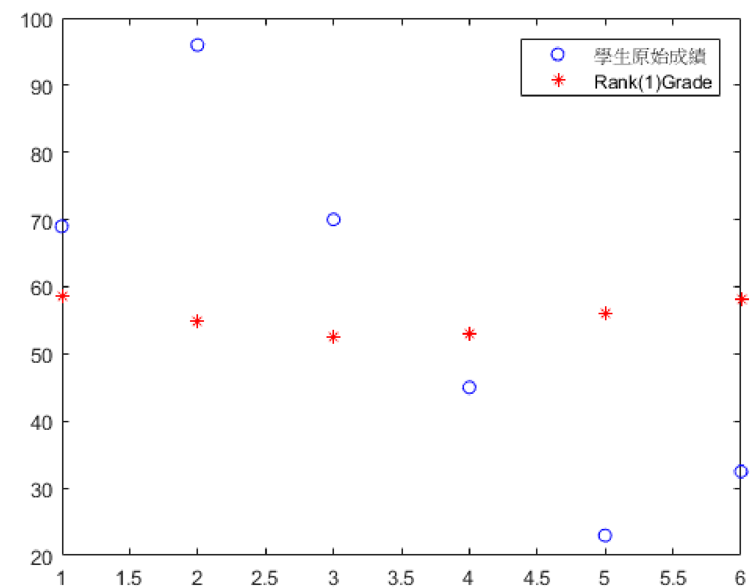
  下面的CODE會舉一個例子解釋

# 程式碼 Coding

誤差

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 4 | 69 | 96 | 70 | 45 | 23 | 32.5 | 335.5 | 55.92 | 24.92 | 22 |

```
%% Deviation about rank(1) matrix M
% Standard Deviation Of Students' Grade over 20
% 4th student
eval(['A',int2str(N),'(4,:);']);
eval(['M',int2str(N),'(4,:);']);

eval(['plot (A',int2str(N),'(4,:),''bo'');']); hold on
eval(['plot (M',int2str(N),'(4,:),''r*'');']); hold off
legend('學生原始成績','Rank(1)Grade');
```




藉由變異數(藍色)24.92

可以看出4號小孩的各科離散程度蠻大的

o代表原始成績
*代表近似矩陣M的4號小孩成績

# 結論 conclusion

- 與傳統統計方式相比

  傳統統計會需要龐大的資料量做分析

  但是經由SVD後可以針對其奇異值分析近似矩陣M

  直觀的觀察左右奇異向量，做一個推估

  可對於一些特例(誤差大)或是觀察詳細資料
  並不適用SVD分析

- 設計課程與教學

  導師可以根據SVD分解

  左奇異向量U(學生個人能力)
  右奇異向量V(班級整體科目)

  去設計課程或是個別輔導學生

  再針對特例再去特別輔導各科