Numerical linear algebra Ⅱ

Woojoo Lee

The induced p-norms

$$
\begin{aligned}
||A||_p &= \max_{x \neq 0} \frac{||Ax||_p}{||x||_p} \\
&= \max_{||x||_p = 1} ||Ax||_p
\end{aligned}
$$

where

$$
||x||_p = (\sum_i |x_i|^p)^{1/p}
$$

Q) Is $||A||_2$ (spectral norm or 2-norm) the same as $||A||_F$?

Properties of the induced p-norms

1. $||\alpha A||_p = |\alpha|||A||_p$
2. $||A||_p \geq 0$ and $||A||_p = 0$ iff $A = 0$
3. $||A + B||_p \leq ||A||_p + ||B||_p$
4. $||Ax||_p \leq ||A||_p||x||_p$ if Ax is well defined
5. $||AB||_p \leq ||A||_p||B||_p$ if AB is well defined
6. $||A||_p||A^{-1}||_p \geq 1$ if A^{-1} exists
7. $1/||A^{-1}||_p = \min_{x \neq 0} \frac{||Ax||_p}{||x||_p}$

For example, if $p = 1$,

$$||A||_1 = \max_j \sum_i |a_{ij}|$$

If $p = 2$,

$$||A||_2 = \sqrt{\rho(A^T A)}$$

where $\rho(C)$ denotes the absolue value of the dominant eigenvalue of the matrix C.

Using matrix norms is useful for basic error analysis. Consider an example of $Ax = b$:

$$\begin{pmatrix} 1 & 1 \\ 1 & 1.0001 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$$

Q) What is the solution ?

$$\begin{pmatrix} 1 & 1 \\ 1 & 1.0001 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2 \\ 2.0001 \end{pmatrix}$$

Q) What is the solution ?

$\rightarrow$ This linear system is ill-posed (or ill-conditioned).

- The linear equation is well-conditioned if small Δb gives small Δx.
- The linear equation is ill-conditioned if small Δb can give large Δx.

A measure for the illposedness of the matrix

Consider $Ax = b$ where $b \neq 0$. Suppose that the solution is $x + \delta x$ for the small perturbation $b + \delta b$.

After simple algebraic manipulations, we get

$$\frac{||\delta x||_2}{||x||_2} \leq ||A^{-1}||_2||A||_2\frac{||\delta b||_2}{||b||_2}$$

Q) Interpret this inequality.

For a nonsingular A,

$$\kappa(A) = ||A^{-1}||_2 ||A||_2$$

is called the condition number of A.

Q) Compute the condition number when A is a (symmetric) PD matrix.

Some properties

1. $\kappa(A) \geq 1$ for all A
2. Small $\kappa(A)$ implies a well-conditioned matrix A
3. Large $\kappa(A)$ implies a ill-conditioned matrix A

There is an R-function to compute the condition number of a square matrix.

kappa {base}

Compute or Estimate the Condition Number of a Matrix

Description

The condition number of a regular (square) matrix is the product of the *norm* of the matrix and the norm of its inverse (or pseudo-inverse), and hence depends on the kind of matrix-norm.

kappa() computes by default (an estimate of) the 2-norm condition number of a matrix or of the *R* matrix of a *QR* decomposition, perhaps of a linear fit. The 2-norm condition number can be of the largest to the smallest *non-zero* singular value of the matrix.

rcond() computes an approximation of the **r**eciprocal **cond**ition number, see the details.

Usage

```
kappa(z, ...)
## Default S3 method:
kappa(z, exact = FALSE,
      norm = NULL, method = c("qr", "direct"), ...)
## S3 method for class 'lm'
kappa(z, ...)
## S3 method for class 'qr'
kappa(z, ...)

.kappa_tri(z, exact = FALSE, LINPACK = TRUE, norm = NULL, ...)

rcond(x, norm = c("O","I","1"), triangular = FALSE, ...)
```

Singular value decomposition (SVD)

Let A be an $m \times k$ matrix.
There exists an $m \times m$ orthogonal matrix U and a $k \times k$ orthogonal matrix V such that

$$A = U\Sigma V^T = \sum_i^{\min(m,k)} \sigma_i u_i v_i^T$$

where Σ has (i,i) entry $\sigma_i \neq 0$ for $i = 1, \cdots, \min(m,k)$.
σ_i is called singular values of A.

Remark: Rank of A = the number of non-zero singular values of A

SVD is closely-connected to a result concerning the approximation of a rectangular matrix by a matrix having lower rank.

Here, the objective criterion is based on Frobenius norm:

$$tr((A-B)(A-B)^T) = \sum_{i=1}^{m}\sum_{j=1}^{k}(a_{ij}-b_{ij})^2$$

$\rightarrow$

$$\begin{aligned}
tr((A-B)(A-B)^T) &= tr(UU^T(A-B)VV^T(A-B)^T) \\
&= tr(U^T(A-B)VV^T(A-B)^TU) \\
&= tr((\Sigma-C)(\Sigma-C)^T) \\
&= \sum_{i=1}^{m}\sum_{j=1}^{k}(\sigma_{ij}-c_{ij})^2 \\
&= \sum_{i=1}^{m}(\sigma_i-c_{ii})^2 + \sum_{i\neq j}(c_{ij})^2
\end{aligned}$$

R-session

$$A = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}$$

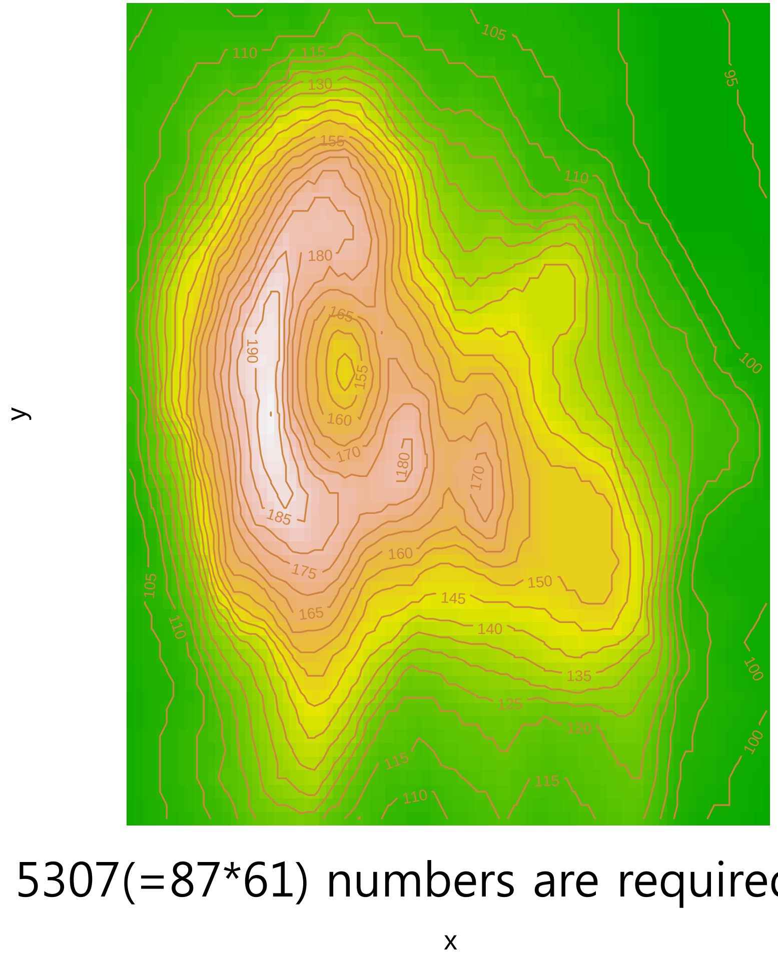
R command: svd(A)

The SVD is related to the spectral decomposition.

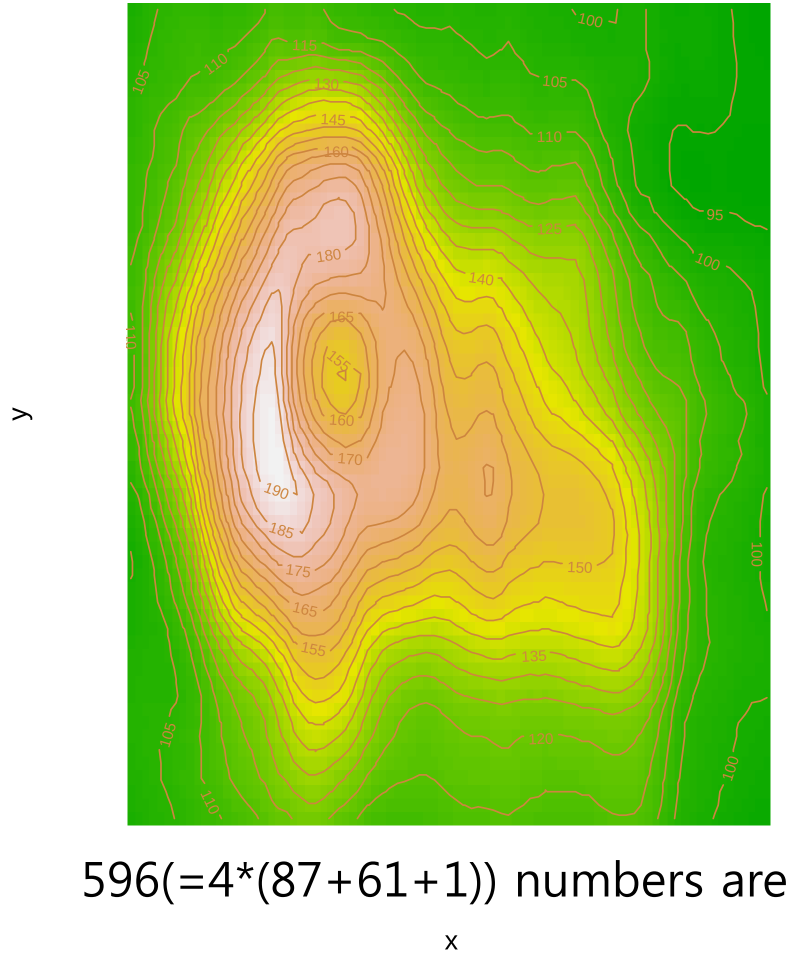
Let A be a $m \times n$ matrix.
Consider AA^T and A^TA.

Application I – image compression

Volcano data example

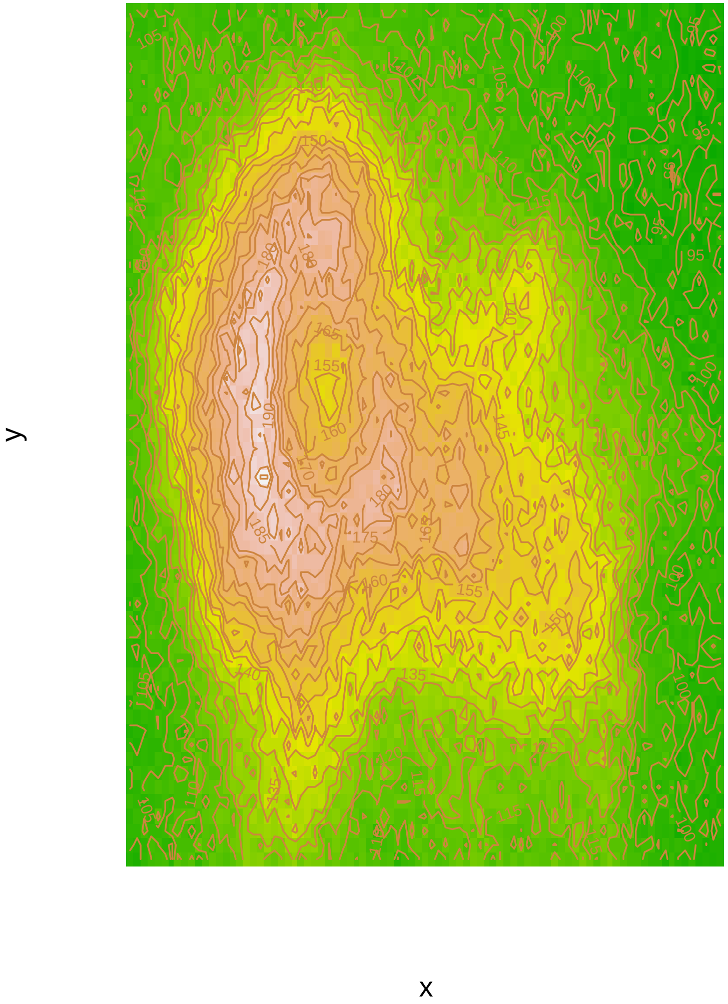


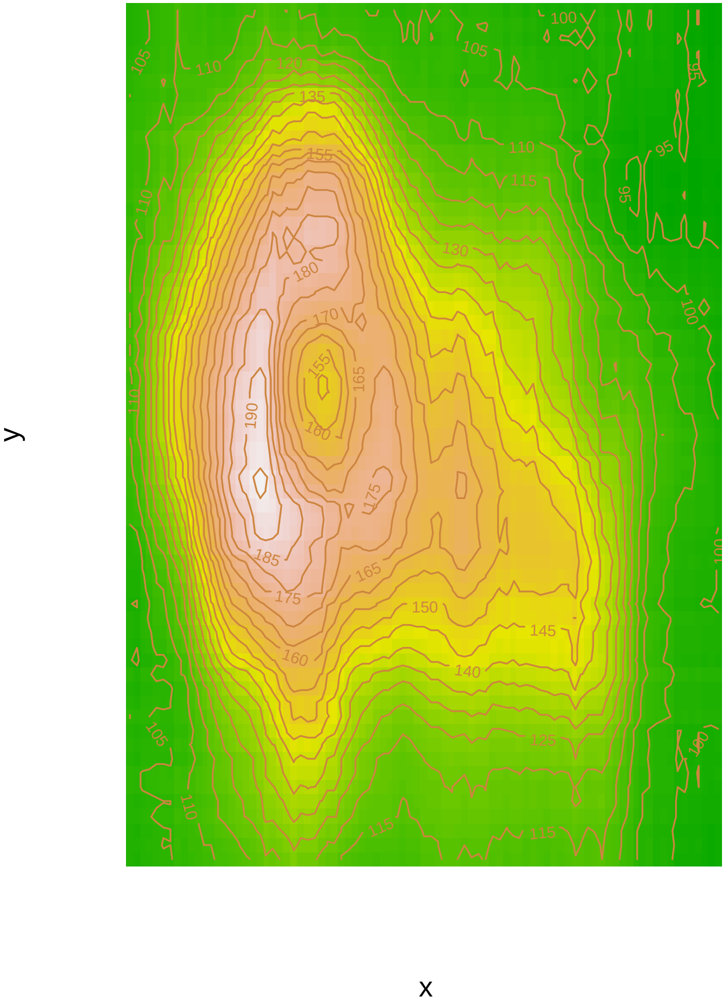
5307(=87*61) numbers are required



596(=4*(87+61+1)) numbers are required

Application I I– denoising (image compression)



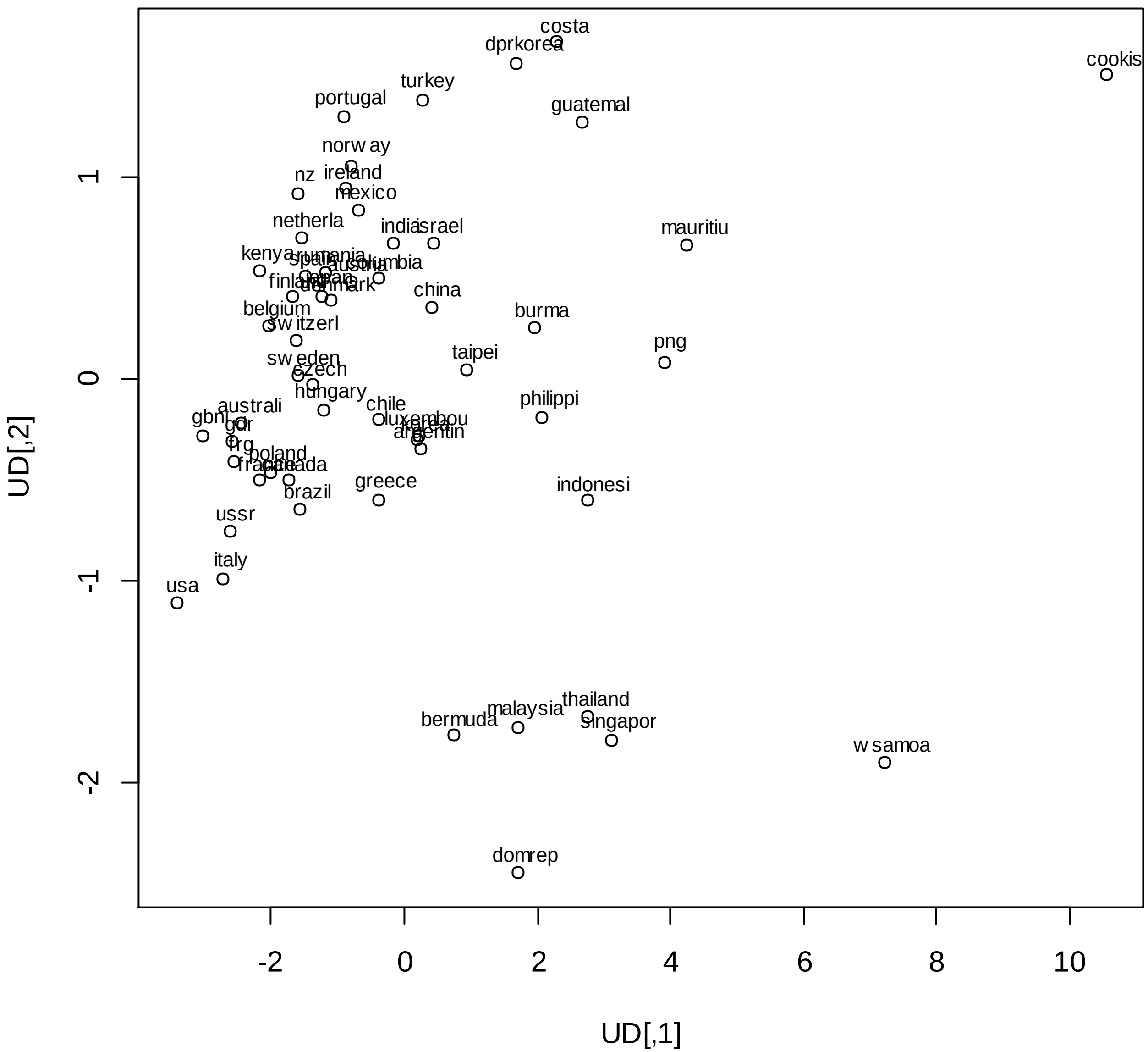


Application III – dimension reduction

Table 8.6 National Track Records for Men

Country	100 m (s)	200 m (s)	400 m (s)	800 m (min)	1500 m (min)	5000 m (min)	10,000 m (min)	Marathon (min)
Argentina	10.23	20.37	46.18	1.77	3.68	13.33	27.65	129.57
Australia	9.93	20.06	44.38	1.74	3.53	12.93	27.53	127.51
Austria	10.15	20.45	45.80	1.77	3.58	13.26	27.72	132.22
Belgium	10.14	20.19	45.02	1.73	3.57	12.83	26.87	127.20
Bermuda	10.27	20.30	45.26	1.79	3.70	14.64	30.49	146.37
Brazil	10.00	19.89	44.29	1.70	3.57	13.48	28.13	126.05
Canada	9.84	20.17	44.72	1.75	3.53	13.23	27.60	130.09
Chile	10.10	20.15	45.92	1.76	3.65	13.39	28.09	132.19
China	10.17	20.42	45.25	1.77	3.61	13.42	28.17	129.18
Columbia	10.29	20.85	45.84	1.80	3.72	13.49	27.88	131.17
Cook Islands	10.97	22.46	51.40	1.94	4.24	16.70	35.38	171.26
Costa Rica	10.32	20.96	46.42	1.87	3.84	13.75	28.81	133.23
Czech Republic	10.24	20.61	45.77	1.75	3.58	13.42	27.80	131.57
Denmark	10.29	20.52	45.89	1.69	3.52	13.42	27.91	129.43
DominicanRepublic	10.16	20.65	44.90	1.81	3.73	14.31	30.43	146.00
Finland	10.21	20.47	45.49	1.74	3.61	13.27	27.52	131.15
France	10.02	20.16	44.64	1.72	3.48	12.98	27.38	126.36
Germany	10.06	20.23	44.33	1.73	3.53	12.91	27.36	128.47
Great Britain	9.87	19.94	44.36	1.70	3.49	13.01	27.30	127.13
Greece	10.11	19.85	45.57	1.75	3.61	13.48	28.12	132.04
Guatemala	10.32	21.09	48.44	1.82	3.74	13.98	29.34	132.53
Hungary	10.08	20.11	45.43	1.76	3.59	13.45	28.03	132.10

Johnson and Wichern, 6ed, Table 8-6



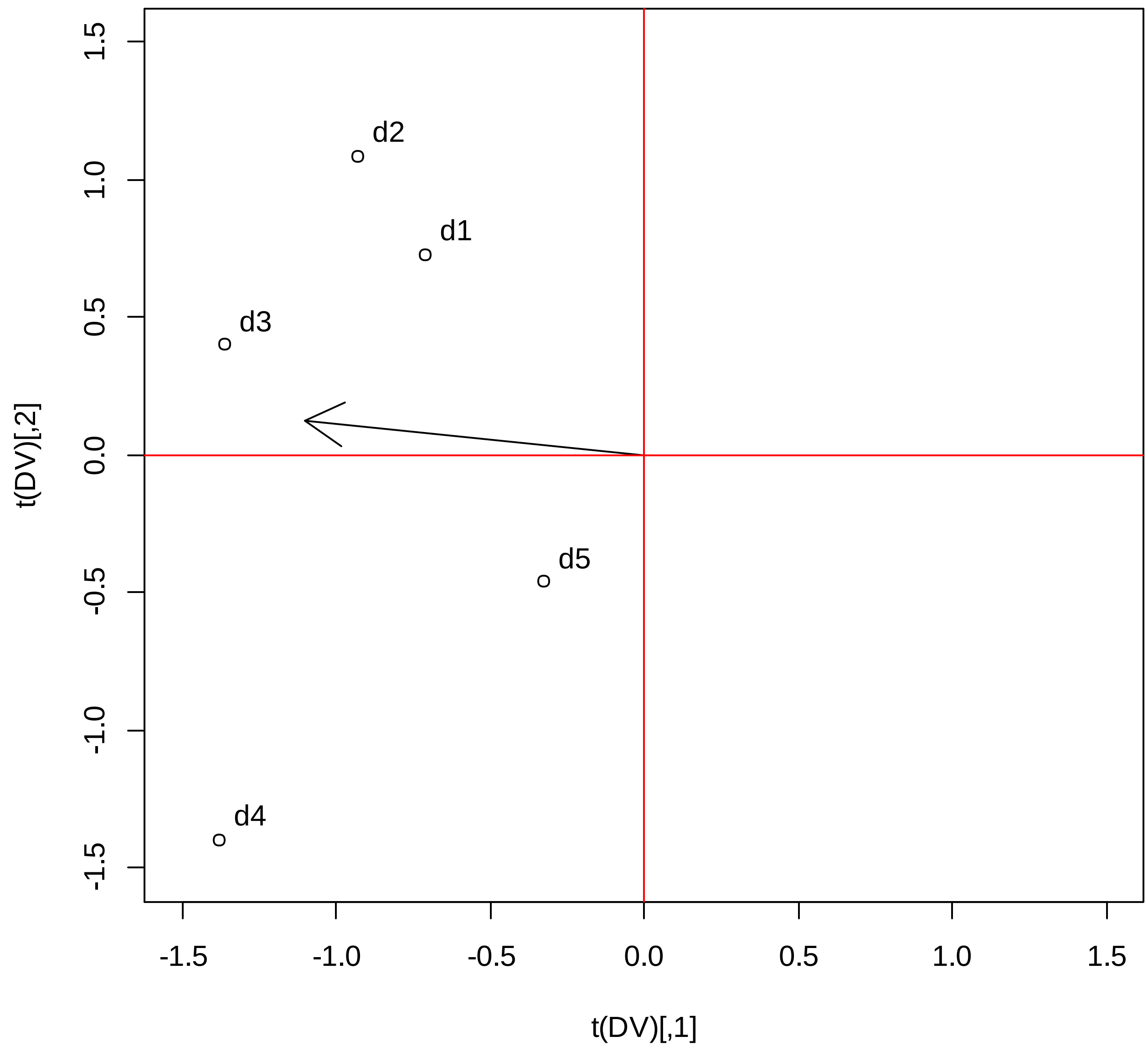

```
              [,1]        [,2]
[1,] 6.005059  3.90246427
[2,] 6.372346  3.17789739
[3,] 6.725328  1.70914832
[4,] 6.971893  0.08556937
[5,] 7.049911 -0.96237977
[6,] 6.890390 -2.14806141
[7,] 6.935745 -2.11246626
[8,] 6.465892 -3.02188058
```

Application IV – latent semantic indexing

```
> tdm
               d1 d2 d3 d4 d5
romeo           1  0  1  0  0
juliet          1  1  0  0  0
happy           0  1  0  0  0
dagger          0  1  1  0  0
live            0  0  0  1  0
die             0  0  1  1  0
free            0  0  0  1  0
new-hampshire   0  0  0  1  1
```

A given query : die , dagger

Q) What is the most relevant document to the query ?



d1
d2
d3
d4
d5
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
t(DV)[,1]
t(DV)[,2]

Some references for LSI

1) https://en.wikipedia.org/wiki/Latent_semantic_indexing

2) https://en.wikipedia.org/wiki/Latent_semantic_analysis

3) www.engr.uvic.ca/~seng474/**svd**.pdf