Ex 18.1 What is the rank of 3×3 diagonal matrix below ?

$$A = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 2 & 1 \end{pmatrix}$$

Solut^n:- Determinant of $A = \begin{vmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 2 & 1 \end{vmatrix}$

$= 1(1\times1 - 2\times1) - 1(0\times1 - 1\times1) + 0(0-1)$

$= 1(1-2) - 1(0-1) + 0(-1)$

$= 1(-1) + 1 + 0(-1)$

$= -1 + 1$

$= 0.$

Matrix rank :-

You can think of a $r \times c$ matrix as a set of r row vectors, each having c elements, or you can think of it as a set of c column vectors each having r elements.

The rank of a matrix is defined as :-

(a) maximum number of linearly independent column vectors in the matrix or

(b) maximum number of linearly independent row vectors in the matrix.

Both definitions are equivalent.

Criterion for linear independence:

Let $a_1, a_2, \ldots, a_n$ be n vectors and let A be the $n \times n$ matrix with these vectors as columns. Then $\{a_1, a_2, \ldots, a_n\}$ are linearly independent if and only if

$$\det(A) = |A| = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} \neq$$

Det of our matrix is 0.
so, all the vectors are not linearly independ

The maximum number of linearly independent ~~rows~~/vectors in a matrix is equal to the number of non zero rows in its row echelon form matrix.

Therefore to find the rank of a matrix we simply transform the matrix to its row echelon form and count the numbe of rows which are non zero.

Row echelon form.

A matrix is in row echelon form (ref) when it satisfies the follow conditions :-

I. The first non zero element in each row called the leading entry is 1

- Each leading entry is in a column to the right of leading entry in ~~row~~ the previous row.

- Rows with all zero elements, if any, are below rows having non zero elements.

Our matrix A is :-

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

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

Aref.

Count the number of non zero rows in Aref = 2.

$\therefore$ Rank of matrix A is 2.

Page No.
Date / /

Ex 18.2 Show that $\lambda = 2$ is an eigenvalue of

$$C = \begin{pmatrix} 6 & -2 \\ 4 & 0 \end{pmatrix}$$

Find the corresponding eigenvector.

Solution:- An eigenvector is a non zero vector that satisfies the equation

$$A\vec{v} = \lambda\vec{v}$$

where A is a square matrix
λ is a scalar
and $\vec{v}$ is the eigenvector.

λ is called eigenvalue. Eigen values and eigen vectors are also known as respectively characteristic roots and characteristic vectors or latent roots and latent vectors.

You can find eigenvalues and eigenvectors by treating a matrix as a system of linear equations and solving for the values of the variables that make up the compone of the eigenvector.

Finding the eigenvalues and corresponding eigenvectors of the matrix

$$A = \begin{bmatrix} 6 & -2 \\ 4 & 0 \end{bmatrix}$$

means applying the above formula to get

$$A\vec{v} = \lambda\vec{v}$$

$$\begin{bmatrix} 6 & -2 \\ 4 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \lambda \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

in order to solve for λ, x_1 and x_2.

This statement is equivalent to the system of equation

$$\left.\begin{array}{l} 6x_1 - 2x_2 = \lambda x_1 \\ 4x_1 = \lambda x_2 . \end{array}\right\} \begin{array}{l} \text{—— (1)} \\ \text{—— (2)} \end{array}$$

$$\left.\begin{array}{l} (6-\lambda)x_1 - 2x_2 = 0 \\ 4x_1 - \lambda x_2 = 0 \end{array}\right\} \text{by rearranging (1) \& (2).}$$

A necessary and sufficient condition for this system to have a non zero vector $[x_1, x_2]$ is that the determinant of coefficient matrix

$$\left| \begin{bmatrix} 6-\lambda & -2 \\ 4 & -\lambda \end{bmatrix} \right| = 0 .$$

Accordingly,

$$\begin{vmatrix} 6-\lambda & -2 \\ 4 & -\lambda \end{vmatrix} = 0$$

$$\therefore (6-\lambda)(-\lambda) - (-2)(4) = 0$$

$$-6\lambda + \lambda^2 + 8 = 0$$

$$\lambda^2 - 6\lambda + 8 = 0$$

$$\lambda^2 - 4\lambda - 2\lambda + 8 = 0$$
$$\lambda(\lambda-4) - 2(\lambda-4) = 0$$
$$(\lambda-2)(\lambda-4) = 0 \quad \text{——— (3)}$$

$\therefore$ $\left.\begin{array}{l}\lambda = 2 \\ \lambda = 4\end{array}\right\}$ are eigenvalues.

Since there are two values of λ that satisfy the equation (3). thus there are two eigenvalues of the original matrix and these are $\lambda = 2$, $\lambda = 4$.

Hence proved / shown that $\lambda = 2$ is an eigenvalue of $C = \begin{pmatrix} 6 & -2 \\ 4 & 0 \end{pmatrix}$.

We can find eign vectors which correspon to the eigenvalues by plugging λ back into the equations and (1),(2) solving for x_1 & x_2.

To find eigenvector corresponding to $\lambda = 2$ start with.

$$(6-\lambda)x_1 - 2x_2 = 0$$
$$\& \ 4x_1 - \lambda x_2 = 0$$

Substitute $\lambda = 2$,

$$(6-2)x_1 - 2x_2 = 0$$
$$4x_1 - 2x_2 = 0$$

$$\therefore 4x_1 = 2x_2.$$

$$\therefore 2x_1 = x_2.$$

There are infinite number of values for x, which satisfy this equation, the only restriction is that not all the components of an eigenvector can equal zero. So, if $x_1 = 1$ then $x_2 = 2$ and an eigenvector corresponding to $\lambda = 2$ is $[1, 2]$.

Ex 18.3. Compute the unique eigen decomposition of the 2×2 matrix in (18.4).

Solution

$$\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

$$A\vec{v} = \lambda\vec{v} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \lambda\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$2x_1 + x_2 = \lambda x_1$$
$$x_1 + 2x_2 = \lambda x_2$$

$\therefore$ $\cancel{(2-\lambda)x_2}$

$$(2-\lambda)x_1 + x_2 = 0$$
$$x_1 + (2-\lambda)x_2 = 0$$

$$\begin{bmatrix} (2-\lambda) & 1 \\ 1 & (2-\lambda) \end{bmatrix}$$

$$\begin{vmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{vmatrix} = 0$$

$$(2-\lambda)(2-\lambda) - 1 \cdot 1 = 0$$
$$\lambda^2 - 4\lambda + 3 = 0$$

$$(\lambda-3)(\lambda-1) = 0.$$

$\therefore$ Eigenvalues are $\lambda = 1$, $\lambda = 3$

To find eigenvectors :-

$\underline{\lambda = 3}$ $\quad (2-\lambda)x_1 + x_2 = 0$

$(2-3)x_1 + x_2 = 0$

$x_1 = x_2$

$\therefore$ If $x_1 = 1$, $x_2 = 1$ $\therefore$ eigenvector $[1, 1]$

$\underline{\lambda = 1}$ $\quad (2-1)x_1 + x_2 = 0$

$x_1 = -x_2$.

$\therefore$ eigenvector $[1, -1]$.

SVD (Example).

Q $\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$.

$A\vec{v} = \lambda\vec{v}$

$$\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \lambda\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

$$2x_1 + x_2 = \lambda x_1$$
$$x_1 + 2x_2 = \lambda x_2$$

$$(2-\lambda)x_1 + x_2 = 0$$
$$x_1 + (2-\lambda)x_2 = 0$$

$$\begin{vmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{vmatrix} = 0$$

$$(2-\lambda)(2-\lambda) - 1\cdot 1 = 0$$
$$\lambda^2 - 4\lambda + 3 = 0$$
$$(\lambda-3)(\lambda-1) = 0.$$

$\therefore$ Eigenvalues $\lambda = 1$, $\lambda = 3$.

To find eigenvectors :-

$$(2-\lambda)x_1 + x_2 = 0$$
$$(2-3)x_1 + x_2 = 0$$

$$x_1 = x_2.$$

If $x_1 = 1$; $x_2 = 1$

$\therefore$ Eigenvector $(1, 1)$

$\underline{\underline{\lambda = 1}}$

$(2-1)\,x_1 + x_2 = 0$

$x_1 = -x_2$

If $x_1 = 1$, $x_2 = -1$

$\therefore$ eigenvector $[1\ -1]$

Now

$$\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

$\vec{v_1} = [1\ 1]$

$u_1 = \dfrac{\vec{v_1}}{|\vec{v_1}|} = \left(\dfrac{1}{\sqrt{2}}, \dfrac{1}{\sqrt{2}}\right)$.

$\vec{w_2} = \vec{v_2} - \vec{u_1} \cdot \vec{v_2} * \vec{u_1}$

$= (1, -1) - \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) \cdot (1, -1) * \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$

$= (1, -1) - \left(\frac{1}{\sqrt{2}} + \frac{-1}{\sqrt{2}}\right) * \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$

$= (1, -1) - 0 = (1, -1)$

$$u_2 = \frac{\vec{w_2}}{|\vec{w_2}|}$$

$$= \left(\frac{1}{\sqrt{2}} \quad \frac{-1}{\sqrt{2}} \right).$$

$$\therefore \quad U = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{bmatrix}.$$

Now for V

$$A^T A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

$$= \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix}$$

$$\begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \lambda \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$5x_1 + 4x_2 = \lambda x_1$$

$$4x_1 + 5x_2 = \lambda x_2.$$

$$(5-\lambda)x_1 + 4x_2 = 0$$

$$4x_1 + (5-\lambda)x_2 = 0$$

$$\begin{vmatrix} (5-\lambda) & 4 \\ 4 & (5-\lambda) \end{vmatrix} = 0$$

$$(5-\lambda)(5-\lambda) - 16 = 0$$

$$25 - 10\lambda + \lambda^2 - 16 = 0$$

$$\lambda^2 - 10\lambda + 9 = 0$$

$$\lambda^2 - 9\lambda - 1\lambda + 9 = 0$$

$$\lambda(\lambda-9) - 1(\lambda-9) = 0$$

$$(\lambda-1)(\lambda-9) = 0$$

$$\lambda = 1, \quad \lambda = 9$$

$\underline{\lambda = 1}$

$$(5-\lambda)x_1 + 4x_2 = 0$$

$$4x_1 + 4x_2 = 0$$

$$\therefore \; x_1 = -x_2.$$

if $x_1 = 1$; $x_2 = -1$.

$\underline{\lambda = 9}$

$$-4x_1 + 4x_2 = 0$$

$$x_1 = x_2$$

if $x_1 = 1$ $x_2 = 1$.

$\therefore$ Eigenvectors $\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$

$v_1 = [1, 1]$

$\vec{u_1} = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$.

$\vec{w_2} = \vec{v_2} - \vec{u_1} \cdot \vec{v_2} * \vec{u_1}$

$= (1, -1) - \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) \cdot (1, -1) * \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$

$= (1, -1) - (0)\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$

$= (1, -1)$.

$\vec{u_2} = \left(\frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}\right)$.

$\therefore V = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{+1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{pmatrix}$.

$\therefore U = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{bmatrix}$ $\quad V^T = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$

$$S = \begin{bmatrix} \sqrt{9} & 0 \\ 0 & \sqrt{1} \end{bmatrix}$$

$$= \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}.$$

$$\therefore \; USV^T = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{bmatrix}$$

SVD (LSI).

Latent Semantic indexing has two advantages over the vector space model :-

1) Synonymy : Synonymy refers to the case where two different words have the same meaning. (say car and automobile)

2) Polysemy : Polysemy refers to the case where term has multiple meanings.

Synonymy and polysemy are handled in LSI but not in vector space model.

Even for a collection of modest size, the term document matrix C is likely to have several tens of thousands of rows and columns, and a rank in tens of thousands as well.

In LSI, we use the SVD, to construct a low rank approximation C_K to the term-document matrix, for a value of K far smaller than the original rank of C.

Thus each row/column is mapped to K-dimensional space, this space is defined by the K principal eigen vectors (corresponding to the largest eigen values) of C^TC and CC^T

Note that matrix C_k is itself still an $m \times n$ matrix irrespective of k.

A query vector is mapped to its representation in the LSI space by the transformation

$$\vec{q_k} = \Sigma_k^{-1} U_k^T \vec{q}$$

Now, we may use cosine similarities to compute the similarity between a query and a document, between two documents or between two terms.

If a query is close to a document in the original space, it remains relatively close in the k-dimensional space.

Consider the term document matrix

	d1	d2	d3	d4	d5	d6
ship	1	0	1	0	0	0
boat	0	1	0	0	0	0
ocean	1	1	0	0	0	0
voyage	1	0	0	1	1	0
trip	0	0	0	1	0	1

Matrix A

Its SVD is,

U as below

$$\begin{matrix} -0.44 & -0.30 & +0.57 & 0.58 & -0.25 \\ -0.13 & -0.33 & -0.59 & 0.00 & 0.73 \\ -0.48 & -0.51 & -0.37 & 0 & -0.61 \\ -0.70 & 0.35 & 0.15 & -0.58 & 0.16 \\ -0.26 & 0.65 & -0.41 & 0.58 & -0.09 \end{matrix}$$

Σ as below,

$$\begin{matrix} 2.16 & 0 & 0 & 0 & 0 \\ 0 & 1.59 & 0 & 0 & 0 \\ 0 & 0 & 1.28 & 0 & 0 \\ 0 & 0 & 0 & 1.00 & 0 \\ 0 & 0 & 0 & 0 & 0.39 \end{matrix}$$

V^T

$$\begin{matrix} -0.75 & -0.28 & -0.20 & -0.45 & -0.33 & -0.12 \\ -0.29 & -0.53 & -0.19 & 0.63 & 0.22 & 0.41 \\ 0.28 & -0.75 & 0.45 & -0.20 & 0.12 & -0.33 \\ 0.00 & 0.00 & 0.58 & 0.00 & -0.58 & 0.58 \\ -0.53 & 0.29 & 0.63 & 0.19 & 0.41 & -0.22 \end{matrix}$$

FOR K= 2,

Σ_2 is
$$\begin{matrix} 2.16 & 0 & 0 & 0 & 0 \\ 0 & 1.59 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{matrix}$$

From this we compute C_2.

$$\begin{matrix} -1.62 & -0.60 & -0.44 & -0.97 & -0.70 & -0.26 \\ -0.46 & -0.84 & -0.30 & 1.00 & 0.35 & 0.65 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{matrix}$$

Truncated $(V)^T$ is

$$\begin{matrix} -1.62 & -0.60 & -0.44 & -0.97 & -0.70 & -0.26 \\ -0.46 & -0.84 & -0.30 & 1.00 & 0.35 & 0.65 \end{matrix}$$

The retrieval quality may improve by the dimensionality reduction.

Curse of dimensionality

The curse of dimensionality refers to various phenomena that arise when analyzing and organizing data in high dimensional spaces (often with hundreds or thousands of dimensions) that do not occur in low dimensional settings. ~~such~~.

The common theme of these problems is that when the dimensionality increases, the volume of the space increases so fast that the available data becomes sparse.

This sparsity is problematic for any method that requires statistical significance. Organizing and searching data often relies on detecting areas where objects form groups with similar properties; in high dimensional data however objects appear to be sparse and dissimilar in many ways which prevents common data organization strategies from being efficient. Hence we need dimensionality reduction.

Application of LSI.

The new low dimensional space can be typically used to :

1. Compare the documents in low dimensional space (data clustering, document classification).

2. Find similar documents across languages, after analyzing a base set of translated documents (cross language retrieval).

3. Find relations between terms (synonymy and polysemy).

4. Given a query of terms, translate it into low dimensional space and find matching documents (information retrieval).

5. Find the best similarity between small groups of terms in a semantic way (ie in a context of knowledge corpus). eg: in a MCQ multiple choice questions answering model.