Spazi vettoriali e Trasformazioni Lineari

R. Basili

Corso di *Web Mining* e *Retrieval*
a.a. 2008-9

April 10, 2009

Outline

Outline

Linear Transformation

Linear Transformation:

Any transformation T of vectors $\underline{x}$ in the space is such that the transformed vector $T(\underline{x})$ lies in another (sometimes the same) vector space.
A transformatin $T : V_n \rightarrow W_m$ is said to be *linear* **iff**:

$$\begin{aligned} T(\alpha \underline{x}) &= \alpha T(\underline{x}) \\ T(\underline{x} + \underline{y}) &= T(\underline{x}) + T(\underline{y}) \\ T(\alpha \underline{x} + \beta \underline{y}) &= \alpha T(\underline{x}) + \beta T(\underline{y}) \end{aligned}$$

Linear Transformation

Linear Transformation and Basis:

The effect of linear transformation T from V_n to W_m is entirely determined by its effect on the basis $\{\underline{b}_1, .., \underline{b}_n\}$ of the originating space V_n, thus we need to know for a generic vector $\underline{x} = \alpha_1 \underline{b}_1 + ... + \alpha_n \underline{b}_n$, just the vectors:

$$\underline{b}'_i = T(\underline{b}_i) \qquad \forall i = 1, ...n \text{ that is}$$

$$\underline{b}'_k = T(\underline{b}_k) = \sum_{i=1}^{n} a_{ik} \underline{b}_i$$

as $T(\underline{x}) = T(\sum_i \alpha_i \underline{b}_i) = \sum_i \alpha_1 T(\underline{b}_i)$.
Notice that the coefficents $a_{ik} \quad \forall i, k = 1, ..., n$ form a square matrix $\mathbf{A}$.

Linear Transformation

Linear Transformation and Basis:

For a generic vector pair $\underline{x} = \sum_i x_i \underline{b}_i$ and $\underline{y} = \sum_i y_i \underline{b}_i$, such that $T(\underline{x}) = \underline{y}$, it follows.

$$T(\underline{x}) = \sum_{k=1}^{n} x_k T(\underline{b}_k) = \sum_{k=1}^{n} x_k \sum_{i=1}^{n} a_{ik} \underline{b}_i =$$

$$\sum_i \left(\sum_k a_{ik} x_k \right) \underline{b}_i \text{ (but also) } =$$

$$\sum_i y_i \underline{b}_i = \underline{y}$$

from which we deduce that

$$y_i - \sum_k a_{ik} x_k = 0 \qquad \forall i = 1, \ldots, n$$

as $\underline{b}_i$ are all linearly independent.

Linear Transformation and Matrices

The operation $y_i = \sum_k a_{ik} x_k \quad (\forall i = 1, \ldots, n)$ suggests a matrix representation with a specific vector by matrix (i.e. row by column) multiplication. First of all the a_{ik} coefficient define a square matrix **A**:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{pmatrix}$$

while $\underline{x}$ and $\underline{y}$ are as usual column vectors:

$$\underline{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \qquad \underline{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

Matrices-Vector multiplications

Moreover, we see that the trasnformation T over the vector $\underline{x}$, with $T(\underline{x}) = \underline{y}$, can be expressed as follows:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

where the ith component y_i of $\underline{y}$ corresponds to the component-wise multiplication between the ith row of $\mathbf{A}$ and the vector $\underline{x}$, i.e. $y_i = \sum_k a_{ik} x_k$. Notice that this also corresponds to an inner product between rows in $\mathbf{A}$ and $\underline{x}$.

This is also written with more synthesis:

$$\underline{y} = \mathbf{A}\underline{x}$$

Matrices and Linear Transformations

Matrices and Linear Transformations

Matrices $\mathbf{A}$ thus represent linear transformations between vectors in a space V_n.
Every T corresponds to a matrix $\mathbf{A}$ and viceversa.

Non singular transformations

We can ask if the inverse transformation exist for each T.
A linear transformation T is *non singular* when the inverse transformation T^{-1} exists such that whereas $\underline{y} = T(\underline{x})$ then $\underline{x} = T^{-1}(\underline{y})$.
The correspondng matrices $\mathbf{a}$ follow the same terminology, $\mathbf{A}^{-1}$ is called the inverse matrix of $\mathbf{A}$, and $\underline{x} = \mathbf{A}^{-1}\underline{y}$.
When $\mathbf{A}^{-1}$ exist for $\mathbf{A}$, then $\mathbf{A}$ is *non singular*, otherwise it is called *singular*.

Matrix operations

Matrix multiplication by a scalar and sum

$\alpha\mathbf{A} = (\alpha a_{ik})$
$\mathbf{A} + \mathbf{B} = (a_{ik}) + (b_{ik}) = (a_{ik} + b_{ik})$

Matrix multiplication: $\mathbf{C} = \mathbf{AB}$

$$\begin{pmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & c_{22} & \dots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \dots & c_{nn} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \dots & b_{nn} \end{pmatrix}$$

Matrix operations and transformations

Matrix multiplication and transformations

Matrix multiplications are the counterpart of the compositions between linear transformations, i.e.

$$\mathbf{AB}\underline{x} = \underline{y} \text{ when } T_A T_B(\underline{x}) = \underline{y}$$

Symmetry

Matrix multiplications are clearly non symmetric, i.e.

$$\underline{y} = \mathbf{AB}\underline{x} \neq \mathbf{BA}\underline{x} = \underline{y}'$$

and correspondingly

$$\underline{y} = T_A T_B(\underline{x}) \neq T_B T_A(\underline{x}) = \underline{y}'$$

Matrix operations and transformations

Zero Matrix

The zero matrix $\underline{\mathbf{0}}$ is the neutral elements with respect to the matrix sums, i.e.

$$\forall \mathbf{A}, \quad \mathbf{A} + \underline{\mathbf{0}} = \underline{\mathbf{0}} + \mathbf{A} = \mathbf{A}$$

It corresponds to the unique matrix $\mathbf{A}$ whereas $\forall i, k = 1, ..., n \quad a_{ik} = 0$.
For $n = 3$, $\underline{\mathbf{0}}$ is as follows:

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Matrix operations: Identity

Identity Matrix

The identity matrix $\underline{\mathbf{I}}$ is the neutral elements with respect to the matrix multiplication, i.e.

$$\forall \mathbf{A}, \quad \mathbf{AI} = \mathbf{IA} = \mathbf{A}$$

It corresponds to the matrix with all elements in the main diagonal equal to 1, e 0 elsewehere, i.e.:

$$\mathbf{I} = (a_{ik}) = \delta_{ik} = \begin{cases} 1 & i = k \\ 0 & i \neq k \end{cases}$$

For $n = 3$, $\mathbf{I}$ is as follows:

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

Change of Basis

Change of Basis

Given two alternative basis $B = \{\underline{b}_1, ..., \underline{b}_n\}$ and $B' = \{\underline{b}'_1, ..., \underline{b}'_n\}$, such that the square matrix $\mathbf{C} = (c_i k)$ describe the change of the basis, i.e.

$$\underline{b}'_k = c_{1k}\underline{b}_1 + c_{2k}\underline{b}_2 + ...c_{nk}\underline{b}_n \qquad \forall k = 1, ..., n$$

Matrix and Change of Basis

Matrix and Change of Basis

The effect of the matrix $\mathbf{C}$ on a generic vector $\underline{x}$ allows to compute the change of basis according only to the involved basis B and B'. For every $\underline{x} = \sum_{k=1}^{n} x_k \underline{b}_k$ such that in the new basis B', $\underline{x}$ can be expressed by $\underline{x} = \sum_{k=1}^{n} x'_k \underline{b}'_k$, then it follows that:

$$\underline{x} = \sum_{k=1}^{n} x'_k \underline{b}'_k = \sum_{k} x'_k \left(\sum_{i} c_{ik} \underline{b}_i \right) = \sum_{i,k=1}^{n} x'_k c_{ik} \underline{b}_i$$

from which it follows that:

$$x_i = \sum_{k=1}^{n} x'_k c_{ik} \qquad \forall i = 1, \ldots, n$$

The above condition suggests that $\mathbf{C}$ is sufficient to describe any change of basis through the matrix vector mutliplication operations:

$$\underline{x} = \mathbf{C} \underline{x}'$$

Matrix and Change of Basis

Matrix and Change of Basis

The effect of the matrix $\mathbf{C}$ on a matrix $\mathbf{A}$ can be seen by studying the case where $\underline{x}, \underline{y}$ are the expression of two vectors in a base B while their counterpart on B' are $\underline{x}', \underline{y}'$, respectively. Now if $\mathbf{A}$ and $\mathbf{B}$ are such that $\underline{y} = \mathbf{A}\underline{x}$ and $\underline{y}' = \mathbf{B}\underline{x}'$, then it follows that:

$$\begin{aligned} \underline{y} = \mathbf{C}\underline{y}' = \mathbf{A}\underline{x} = \mathbf{A}(\mathbf{C}\underline{x}') &= \mathbf{AC}\underline{x}' \\ &\text{(this means that)} \\ \underline{y}' &= \mathbf{C}^{-1}\mathbf{AC}\underline{x}' \end{aligned}$$

from which it follows that:

$$\mathbf{B} = \mathbf{C}^{-1}\mathbf{AC}$$

The transformation of basis $\mathbf{C}$ is a *similarity transformation* and matrices $\mathbf{A}$ and $\mathbf{C}$ are said *similar*.

Adjont Matrix

Adjoint (Transpose) of a matrix

The adjoint $\mathbf{A}^*$ of a matrix $\mathbf{A}$ is the unique matrix such that

$$(\mathbf{A}^T \underline{x}, \underline{y}) = (\underline{x}, \mathbf{A}\underline{y})$$

In case $\mathbf{A}$ has real values (as always in this course) the adjoint $\mathbf{A}^*$ is noted as $\mathbf{A}^T$ and it is called *transpose* of matrix $\mathbf{A}$. $\mathbf{A}^T$ is obtained from $\mathbf{A}$ by exchanging rows and columns, i.e.

$$\mathbf{A} = (a_{ij}) \Longrightarrow \mathbf{A}^T = (a_{ji})$$

Self-adjointness and Idempotence

Self-Adjoint matrices

A matrix $\mathbf{A}$ is *self-adjoint* **iff** the following holds:

$$(\mathbf{A}\underline{x}, \underline{y}) = (\underline{x}, \mathbf{A}\underline{y})$$

Note that the above means that when $\mathbf{A}$ takes only real values, then $\mathbf{A}$ is *symmetric*, i.e. $\mathbf{A} = \mathbf{A}^T$. Diagonal matrices are always self-adjoint.

Idempotence

A matrix $\mathbf{A}$ is idempotent **iff** the following holds:

$$\mathbf{E}\mathbf{E}\underline{x} = \mathbf{E}\underline{x} \qquad \forall \underline{x}$$

Projectors

Projectors

Linear transformations that are

- Idempotent (i.e. $\mathbf{E}\mathbf{E}\underline{x} = \mathbf{E}\underline{x} \qquad \forall \underline{x}$)
- Self-Adjoint: (i.e. $(\mathbf{A}\underline{x}, \underline{y}) = (\underline{x}, \mathbf{A}\underline{y})$

are called *projectors*.

Examples

Some noticeable examples of projectors are alrady known:

- (Null Matrix) The operator $\underline{\mathbf{O}}$ is a projector: it maps every vector $\underline{x}$ in the null vector $\underline{0}$.
- (Idenity) The operator $\underline{\mathbf{I}}$ is a projector: it maps every vector $\underline{x}$ into itself.

Projectors are applications between a vector space V_n and one of its subspaces: as $\underline{0}$ and $\underline{1}$ are part of this subspace it has still the properties of being a vector space.

Projectors

1-dimensional projections

Given a basis $B = \{\underline{b}_i\}$, for every $\underline{b}_i$ a projector $\mathbf{P}_i$ can be built, that maps any $\underline{x} = \sum_i x_i \underline{b}_i$ in the subspace generated (or spanned) by $\underline{b}_i$, i.e.

$$\mathbf{P}_i \underline{x} = x_i \underline{b}_i$$

If B is an orthonormal basis, $\mathbf{P}_i$ are a collection of orthogonal projectors:

- every vector in the space spanned by $\underline{b}_i$ will be projected into itself
- every vector orthogonal to $\underline{b}_i$ will be projected into the null vector, $\underline{0}$.
- Every vector $\underline{x}$ is the sum of a vector $\underline{x}_i$ in the subspace spanned by $\underline{b}_i$ and a vector $\underline{x}^{\perp}$ in the subspace orthogonal to $\underline{b}_i$, i.e. $\underline{x} = \underline{x}_i + \underline{x}^{\perp}$.

Projectors

1-dimensional projections: example

Let $B = \{\underline{b}_i, \quad i = 1, ..., n\}$ be an orthonormal basis, and $\underline{x} = \sum_i x_i \underline{b}_i$ be a normalized vector, i.e. $\|\underline{x}\| = 1$ (or $\sum_i |x_i|^2 = 1$). It is of course true that, when $\mathbf{P}_i$ is the projector relative to $\underline{b}_i$, then $\mathbf{P}_i \underline{x} = x_i$. As

$$\begin{aligned}
(\underline{x}, \mathbf{P}_i \underline{x}) &= (\underline{x}, \mathbf{P}_i \mathbf{P}_i \underline{x}) && \text{(Idempotence)} \\
&= (\mathbf{P}_i \underline{x}, \mathbf{P}_i \underline{x}) && \text{(Self-adjointness)} \\
&= (x_i \underline{b}_i, x_i \underline{b}_i) \\
&= x_i^2 (\underline{b}_i, \underline{b}_i) = |x_i|^2
\end{aligned}$$

then the base B establishes through projectors $\mathbf{P}_i$, a probability distribution in the individual spaces spanned by $\mathbf{P}_i$.
Selecting a base B is like *deciding about a specific point of view on the space*, and its ability to represent vectors (as representations for objects) $\underline{x}$.

Projectors and probabilty distributions

1-dimensional projections and probabilities

Notice how it is also true that a given normalized vector $\underline{x} \in V_n$ determines a probability distibution in different subspaces generated by the $\mathbf{P}_i$.
This function depends on $\underline{x} \in V_n$ and ranges in the set of spaces L_i of V_n, as follows:

$$\mu_{\underline{x}}(L_i) = (\mathbf{P}_i\underline{x}, \mathbf{P}_i\underline{x}) = |\mathbf{P}_i\underline{x}|^2$$

Properties of $\mu_{\underline{x}}$

- $\mu_{\underline{x}}(\underline{0}) = 0$
- $\mu_{\underline{x}}(V_n) = 1$
- $\mu_{\underline{x}}(L_i \oplus L_j) = \mu_{\underline{x}}(L_i) + \mu_{\underline{x}}(L_j)$, whenever $L_i \cap L_j = \emptyset$. $L_i \oplus L_j$ is the smallest subspace of V_n that contains both L_i and L_j.

Eigenvalues and eigenvectors

Eigenvectors

An eigenvector $\underline{x}$ for a matrix $\mathbf{A}$ is a non-zero vector for which a scalar $\lambda \in \Re$ exists such that

$$\mathbf{A}\underline{x} = \lambda \underline{x}$$

The value of the scalar λ is called eigenvalue of $\mathbf{A}$ associated to $\underline{x}$, and correspond to the scaling factor along the direction of $\underline{x}$.

Example

$$\mathbf{A} = \begin{pmatrix} 0 & 2 \\ 2 & 0 \end{pmatrix} \quad \text{and} \quad \underline{x} = \begin{pmatrix} 3 \\ 3 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 2 \\ 2 & 0 \end{pmatrix} \begin{pmatrix} 3 \\ 3 \end{pmatrix} = \begin{pmatrix} 6 \\ 6 \end{pmatrix} = 2 \begin{pmatrix} 3 \\ 3 \end{pmatrix}$$

$\underline{x}$ is an eigenvector of $\mathbf{A}$ and $\lambda = 2$ is its eigenvalue.

Eigenvalues, eigenvectors and some properties

Eigenvalues, eigenvectors: Some Consequences

When a matrix $\mathbf{A}$ has an eigenvector $\underline{x}$ it must satisfy the following condition:

$$\mathbf{A}\underline{x} = \lambda\underline{x}$$

We can rewrite the condition $\mathbf{A}\underline{x} = \lambda\underline{x}$ as

$$(\mathbf{A} - \lambda\mathbf{I}\underline{x}) = \underline{0}$$

where $\mathbf{I}$ is the Identity matrix.

In order for a non-zero vector $\underline{x}$ to satisfy this equation, $\mathbf{A} - \lambda\mathbf{I}$ *must not be invertible*(see next slide).
The consequence is that the determinant of $\mathbf{A} - \lambda\mathbf{I}$ must equal 0. This function is $p(\lambda) = det(\mathbf{A} - \lambda\mathbf{I})$, called the *characteristic polynomial* of $\mathbf{A}$.
The eigenvalues of $\mathbf{A}$ are simply the roots of the characteristic polynomial of $\mathbf{A}$.

Eigenvalues, eigenvectors and some properties: Proof

$\mathbf{A} - \lambda\mathbf{I}$ *must not be invertible: Why?*

$\mathbf{A} - \lambda\mathbf{I}$ *must not be invertible*, as otherwise, if $\mathbf{A} - \lambda\mathbf{I}$ has an inverse, and

$$\begin{aligned}(\mathbf{A} - \lambda\mathbf{I})^{-1}(\mathbf{A} - \lambda\mathbf{I})\underline{x} &= (\mathbf{A} - \lambda\mathbf{I})^{-1}\underline{0x} \\ \mathbf{I}\underline{x} &= 0.\end{aligned}$$

the zero vector is derived. This is not admissibile as, by definition, $\underline{x} \neq \underline{0}$.

Eigenvalues and eigenvectors

An example: computing eigenvalues

Let $\mathbf{A} = \begin{pmatrix} 2 & -4 \\ -1 & -1 \end{pmatrix}$. Then
$p(\lambda) = (2-\lambda)(-1-\lambda) - (-4)(-1) = \lambda^2 - \lambda - 6 = (\lambda-3)(\lambda+2)$
The eigenvectors are then the solution of the linear equation system given by $(\mathbf{A} - \lambda\mathbf{I})\underline{x} = \underline{0}$.
Given the first eigenvalue $\lambda_1 = 3$, $(\mathbf{A} - 3\mathbf{I})\underline{x} = \underline{0}$ gives the following system:

$$\begin{cases} -x_1 - 4x_2 & = & 0 \\ -x_1 - 4x_2 & = & 0 \end{cases}$$

This suggests that all vectors of the form $\alpha\underline{x}_1$ are eigenvectors with $\underline{x}_1^T = (-4, 1)$. The span of the vector $(-4, 1)^T$ is the eigenspace corresponding to $\lambda_1 = 3$. Correspondingly, the span of the vector $\underline{x}_2 = (1, 1)^T$ corresponds to the eigenspace of $\lambda_2 = -2$.
Notice that $\underline{x}_1$ and $\underline{x}_2$ are linearly independent, so they can form a basis.

Eigenvalues and eigenvectors

Eigenvectors of Symmetric matrices

A symmetric non singular real-valued matrix $\mathbf{A}$ is such that $\mathbf{A} = \mathbf{A}^T$, and on two dimensions, this means that :

$$\begin{array}{ll} i) & a_{11}, a_{22} \neq 0 \\ ii) & a_{12} = a_{21} = a \end{array}$$

In order for $\mathbf{A}$ to have two real eigenvalues the following must hold:

$$\begin{aligned} p(\lambda) &= (a_{11} - \lambda)(a_{22} - \lambda) - a^2 = \\ &= \lambda^2 - (a_{11} + a_{22})\lambda + a_{11}a_{22} - a^2 = 0 \end{aligned}$$

from which eigenvalues are distinct **iff**:

$$(a_{11} - a_{22})^2 + 4a^2 \geq 0$$

The above inequality is always satisfied, with the 0 case only when $\mathbf{A} = \mathbf{I}$.

Eigenvalues and eigenvectors

Eigenvectors and orthogonality

Whenever a matrix $\mathbf{A}$ has n distinct eigenvectors $\underline{x}_i$ with all real-valued and distinct eigenvalues λ_i, it is called *non-degenerate*.
A non degenerate matrix $\mathbf{A}$ has all the eigenvectors mutually orthogonal.

In fact, given two any eigenvectors $\underline{x}_1 \neq \underline{x}_2$, with $\mathbf{A}\underline{x}_i = \lambda_i \underline{x}_i \qquad (i = 1,2)$, it follows that

$$\lambda_1(\underline{x}_1, \underline{x}2) = (\mathbf{A}\underline{x}_1, \underline{x}_2) = (\underline{x}_1, \mathbf{A}\underline{x}2) = \lambda_2(\underline{x}_1, \underline{x}_2)$$

$$\text{from which it follows that} \qquad (\lambda_1 - \lambda_2)(\underline{x}_1, \underline{x}_2) = 0$$

However as $\lambda_1 \neq \lambda_2$, and $\underline{x}_1$, $\underline{x}_2$ were arbitrarily chosen, the result is that

$$\forall i,j = 1,...,n \qquad (\underline{x}_i, \underline{x}_j) = \begin{cases} \|\underline{x}_i\|^2 & i = j \\ 0 & i \neq j \end{cases}$$

Spectral Theorem

Spectral theorem

For every self-adjoint matrix $\mathbf{A}$ on a finite dimensional inner product space V_n, there correspond real valued numbers $\alpha_1, ..., \alpha_r$, and orthonormal projections $\mathbf{E}_1, ..., \mathbf{E}_r$,with $r \leq n$, such that:

- (1) all α_l are *pairwise distinct*
- (2) all $\mathbf{E}_j$ are not null (i.e. $\forall j, \mathbf{E}_j \neq \mathbf{0}$
- (3) $\sum_{j=1}^{r} \mathbf{E}_j = \mathbf{I}$
- (4) $\mathbf{A} = \sum_{j=1}^{r} \alpha_j \mathbf{E}_j$

Notice that the set of self-adjoint matrices whenever the underlying field is the set of real numbers consists of the set of symmetric matrices. the spectral theorem suggests that a possible basis where to diagonalize them is always available through their eigenvectors.
Applications: document similarity matrices where $a_{ij} = sim(d_i, d_j)$.

References

Vectors, Operations, Norms and Distances

K. Van Rijesbergen, The Geometry of Information Retrieval, CUP Press, 2004. Chapter 4.