Mathematical modelling

Lecture 3, March 5, 2021

Faculty of Computer and Information Science
University of Ljubljana

2020/2021

Case 4: *A general matrix A. (using SVD)*

Theorem (Singular value decomposition - SVD)

Let $A \in \mathbb{R}^{n\times m}$ be a matrix. Then it can be expressed as a product

$$A = U\Sigma V^T,$$

where

- $U \in \mathbb{R}^{n\times n}$ *is an orthogonal matrix with left singular vectors* u_i *as its columns,*
- $V \in \mathbb{R}^{m\times m}$ *is an orthogonal matrix with right singular vectors* v_i *as its columns,*
- $\Sigma = \left[\begin{array}{ccc|c} \sigma_1 & & & 0 \\ & \ddots & & \vdots \\ & & \sigma_r & 0 \\ \hline & 0 & & 0 \end{array}\right] = \begin{bmatrix} S & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{n\times m}$ *is a diagonal matrix*

with singular values

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$$

on the diagonal.

Derivations for computing SVD

If $A = U\Sigma V^T$, then

$$A^TA = (V\Sigma^T U^T)(U\Sigma V^T) = V\Sigma^T\Sigma V^T = V\begin{bmatrix} S^2 & 0 \\ 0 & 0 \end{bmatrix} V^T \in \mathbb{R}^{m\times m},$$

$$AA^T = (U\Sigma V^T)(U\Sigma V^T)^T = U\Sigma\Sigma^T U^T = U\begin{bmatrix} S^2 & 0 \\ 0 & 0 \end{bmatrix} U^T \in \mathbb{R}^{n\times n}.$$

Let

$$V = \begin{bmatrix} v_1 & v_2 & \cdots & v_m \end{bmatrix} \quad \text{and} \quad U = \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix}$$

be the column decompositions of V and U.

Let $e_1, \ldots, e_m \in \mathbb{R}^m$ and $f_1, \ldots, f_n \in \mathbb{R}^n$ be the standard coordinate vectors of $\mathbb{R}^m$ and $\mathbb{R}^n$, i.e., the only nonzero component of e_i (resp. f_j) is the i-th one (resp. j-th one), which is 1. Then

$$A^TAv_i = V\Sigma^T\Sigma V^T v_i = V\Sigma^T\Sigma e_i = \begin{cases} \sigma_i^2 v_i, & \text{if } i \le r, \\ 0, & \text{if } i > r, \end{cases}$$

$$AA^Tu_j = U\Sigma\Sigma^T U^T u_j = U\Sigma\Sigma^T f_j = \begin{cases} \sigma_i^2 u_j, & \text{if } j \le r, \\ 0, & \text{if } j > r. \end{cases}$$

Further on,

$$(AA^T)(Av_i) = A(A^T A)v_i = \begin{cases} \sigma_i^2 Av_i, & \text{if } i \le r, \\ 0, & \text{if } i > r, \end{cases}$$

$$(A^T A)(A^T u_j) = A^T(AA^T)u_j = \begin{cases} \sigma_j^2 A^T u_j, & \text{if } j \le r, \\ 0, & \text{if } j > r. \end{cases}$$

It follows that:

- $\Sigma^T\Sigma = \begin{bmatrix} S^2 & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{m\times m}$ (resp. $\Sigma\Sigma^T = \begin{bmatrix} S^2 & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{n\times n}$) is the diagonal matrix with eigenvalues σ_i^2 of $A^T A$ (resp. AA^T) on its diagonal, so the singular values σ_i are their square roots.
- V has the corresponding eigenvectors (normalized and pairwise orthogonal) of $A^T A$ as its columns, so the right singular vectors are eigenvectors of $A^T A$.
- U has the corresponding eigenvectors (normalized and pairwise orthogonal) of AA^T as its columns, so the left singular vectors are eigenvectors of AA^T.

- Av_i is an eigenvector of AA^T corresponding to σ_i^2 and so

$$u_i = \frac{Av_i}{\|Av_i\|}$$

is a left singular vector corresponding to σ_i.

- $A^T u_j$ is an eigenvector of $A^T A$ corresponding to σ_j^2 and so

$$v_j = \frac{A^T u_j}{\|A^T u_j\|}$$

is a right singular vector corresponding to σ_j.

Algorithm for SVD

- Compute the eigenvalues and an orthonormal basis consisting of eigenvectors of the symmetric matrix $A^T A$ or AA^T (depending on which is of them is of smaller size).
- The singular values of the matrix $A \in \mathbb{R}^{n \times m}$ are equal to $\sigma_i = \sqrt{\lambda_i}$, where λ_i are the nonzero eigenvalues of $A^T A$ (resp. AA^T).
- The left singular vectors are the corresponding orthonormal eigenvectors of AA^T.
- The right singular vector are the corresponding orthonormal eigenvectors of $A^T A$.
- If u (resp. v) is a left (resp. right) singular vector corresponding to the singular value σ_i, then $v = Au$ (resp. $u = A^T v$) is a right (resp. left) singular vector corresponding to the same singular value.
- The remaining columns of U (resp. V) consist of an orthonormal basis of the kernel (i.e., the eigenspace of $\lambda = 0$) of AA^T (resp. $A^T A$).

General algorithm for A^+ (long version)

1. For $A^T A$ compute its eigenvalues

$$\lambda_1 \geq \lambda_2 \geq \cdots, \geq \lambda_r > \lambda_{r+1} = \ldots = \lambda_m = 0$$

   and the corresponding orthonormal eigenvectors

$$v_1, \ldots, v_r, v_{r+1}, \ldots, v_m,$$

   and form the matrices

$$\Sigma = \mathrm{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_m}) \in \mathbb{R}^{n \times m},$$

$$V_1 = \begin{bmatrix} v_1 & \cdots & v_r \end{bmatrix}, \quad V_2 = \begin{bmatrix} v_{r+1} & \cdots & v_m \end{bmatrix} \quad \text{and} \quad V = \begin{bmatrix} V_1 & V_2 \end{bmatrix}.$$

2. Let

$$u_1 = \frac{Av_1}{\|Av_1\|}, \quad u_2 = \frac{Av_2}{\|Av_2\|}, \quad \ldots \quad , \quad u_r = \frac{Av_r}{\|Av_r\|},$$

   and $u_{r+1}, \ldots, u_n$ vectors, such that $\{u_1, \ldots, u_n\}$ is an ortonormal basis for $\mathbb{R}^n$. Form the matrices

$$U_1 = \begin{bmatrix} u_1 & \cdots & u_r \end{bmatrix}, \quad U_2 = \begin{bmatrix} u_{r+1} & \cdots & u_n \end{bmatrix} \quad \text{and} \quad U = \begin{bmatrix} U_1 & U_2 \end{bmatrix}.$$

3. Then

$$A^+ = V\Sigma^+ U^T.$$

General algorithm for A^+ (short version)

1. For $A^T A$ compute its **nonzero** eigenvalues
$$\lambda_1 \geq \lambda_2 \geq \cdots, \geq \lambda_r > 0$$
and the corresponding orthonormal eigenvectors
$$v_1, \ldots, v_r,$$
and form the matrices
$$S = \operatorname{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_r}) \in \mathbb{R}^{r \times r},$$
$$V_1 = \begin{bmatrix} v_1 & \cdots & v_r \end{bmatrix}.$$
2. Put the vectors
$$u_1 = \frac{Av_1}{\|Av_1\|}, \quad u_2 = \frac{Av_2}{\|Av_2\|}, \quad \ldots \quad , \quad u_r = \frac{Av_r}{\|Av_r\|}$$
in the matrix
$$U_1 = \begin{bmatrix} u_1 & \cdots & u_r \end{bmatrix}.$$
3. Then
$$A^+ = V_1 \Sigma^+ U_1^T.$$

Correctness of the computation of A^+

Step 1. $V\Sigma^+U^T$ is equal to A^+.

(i) $AA^+A = A$:

$$\begin{aligned} AA^+A &= (U\Sigma V^T)(V\Sigma^+U^T)(U\Sigma V^T) = U\Sigma(V^TV)\Sigma^+(U^TU)\Sigma V^T \\ &= U\Sigma\Sigma^+\Sigma V^T = U\Sigma V^T = A. \end{aligned}$$

(ii) $A^+AA^+ = A^+$: Analoguous to (i).

(iii) $(AA^+)^T = AA^+$:

$$\begin{aligned} (AA^+)^T &= \Big((U\Sigma V^T)(V\Sigma^+U^T)\Big)^T = \Big(U\Sigma\Sigma^+U^T\Big)^T \\ &= \left(U \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} U^T\right)^T = U \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} U^T \\ &= (U\Sigma V^T)(V\Sigma^+U^T) = A^+. \end{aligned}$$

(iv) $(A^+A)^T = A^+A$: Analoguous to (iii).

Step 2. $V\Sigma^+U^T$ is equal to $V_1\Sigma^+U_1^T$.

$$V\Sigma U^T = \begin{bmatrix} V_1 & V_2 \end{bmatrix} \begin{bmatrix} S & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} U_1^T \\ U_2^T \end{bmatrix} = \begin{bmatrix} V_1 S & 0 \end{bmatrix} \begin{bmatrix} U_1^T \\ U_2^T \end{bmatrix} = V_1 S U_1^T.$$

Example

Compute the SVD and A^+ of the matrix $A = \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{bmatrix}$.

- $AA^T = \begin{bmatrix} 17 & 8 \\ 8 & 17 \end{bmatrix}$ has eigenvalues 25 and 9. The eigenvectors of A^TA corresponding to the eigenvalues 25, 9, 0 are

$$v_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \end{bmatrix}^T, \; v_2 = \begin{bmatrix} \frac{1}{\sqrt{18}} & -\frac{1}{\sqrt{18}} & \frac{4}{\sqrt{18}} \end{bmatrix}^T, \; v_3 = \begin{bmatrix} \frac{2}{\sqrt{3}} & -\frac{2}{3} & -\frac{1}{3} \end{bmatrix}^T.$$

- $u_1 = \frac{Av_1}{\|Av_1\|} = \frac{Av_1}{\sigma_1} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}^T$ and $u_2 = \frac{Av_2}{\|Av_2\|} = \frac{Av_2}{\sigma_2} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}^T$.

- $A = U\Sigma V^T = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{18}} & -\frac{1}{\sqrt{18}} & \frac{4}{\sqrt{18}} \\ \frac{2}{\sqrt{3}} & -\frac{2}{3} & -\frac{1}{3} \end{bmatrix}.$
-

$$A^+ = V\Sigma^+ U^T = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{18}} & \frac{2}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{18}} & -\frac{2}{3} \\ 0 & \frac{4}{\sqrt{18}} & -\frac{1}{3} \end{bmatrix} \begin{bmatrix} \frac{1}{5} & 0 \\ 0 & \frac{1}{3} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{7}{45} & \frac{2}{45} \\ \frac{2}{45} & \frac{7}{45} \\ \frac{2}{9} & -\frac{2}{9} \end{bmatrix}.$$

An application of SVD: principal component analysis or PCA

PCA is a very well-known and efficient method for data compression, dimension reduction, ...

Due to its importance in different fields, it has many other names: discrete Karhunen-Loève transform (KLT), Hotelling transform, empirical orthogonal functions (EOF), ...

Let $\{X_1, \ldots, X_m\}$ be a sample of vectors from $\mathbb{R}^n$.

In applications, often $m << n$, where n is very large, for example, $X_1, \ldots, X_m$ can be

- vectors of gene expressions in m tissue samples or
- vectors of grayscale in images
- bag of words vectors, with components corresponding to the numbers of certain words from some dictionary in specific texts, ...,

or $n << m$ for example if the data represents a point cloud in a low dimensional space $\mathbb{R}^n$ (for example in the plane).

We will assume that $m << n$.

Also assume that the data is centralized, i.e., the centeroid is in the origin

$$\mu = \frac{1}{m} \sum_{i=1}^{m} X_i = 0 \in \mathbb{R}^n.$$

If not, we substract μ from all vectors in the data set.

Let

$$X = \begin{bmatrix} X_1 & X_2 & \cdots & X_m \end{bmatrix}^T$$

be the matrix of dimension $m \times n$ with data in the rows.

Let $X^T X \in \mathbb{R}^{m \times m}$ and $XX^T \in \mathbb{R}^{n \times n}$ be the covariance matrices of the data.

- The principal values of the data set $\{X_1, \ldots, X_r\}$ are the nonzero eigenvalues $\lambda_i = \sigma_i^2$ of the covariance matrices (where σ_i are the singular values of X).
- The principal directions in $\mathbb{R}^n$ are corresponding eigenvectors $v_1, \ldots, v_r$, i.e. the columns of the matrix V from the SVD of X. The remaining clolumns of V (i.e. the eigenvectors correspondong to 0) form a basis of the null space of X.
- The first column v_1, the first principal direction, corresponds to the direction in $\mathbb{R}^n$ with the largest variance in the data X_i, that is, the most informative direction for the data set, the second the second most important, ...
- The principal directions in $\mathbb{R}^m$ are the columns $u_1, \ldots, u_r$ of the matrix U and represent the coefficients in the linear decomposition of the vectors $X_1, \ldots, X_m$ along the orthonormal basis $v_1, \ldots v_n$ of $\mathbb{R}^n$.

PCA provides a linear dimension reduction method based on a projection of the data from the space $\mathbb{R}^n$ into a lower dimensional subspace spanned by the first few principal vectors $v_1, \ldots, v_k$ in $\mathbb{R}^n$.

The idea is to approximate

$$X_i = u_{1,i} v_1 + \cdots + u_{k,i} v_k + \cdots + u_{m,i} v_m \cong u_{1,i} v_1 + \cdots + u_{k,i} v_k$$

with the first k most informative directions in $\mathbb{R}^n$ and supress the last $m - k$.

PCA has the following amazing property:

Theorem

Among all possible projections of $p\colon \mathbb{R}^n \to \mathbb{R}^k$ onto a k-dimensional subspace, PCA provides the best in the sense that the error

$$\sum_i \|X_i - p(X_i)\|^2$$

is the smallest possible.

1.3 The MP inverse and systems of linear equations

Let $A \in \mathbb{R}^{n\times m}$, where $m > n$. A system of equations $Ax = b$ that has more variables than constraints. Typically such system has infinitely many solutions, but it may happen that it has no solutions. We call such system an underdetermined system.

Theorem

1. *An underdetermined system of linear equations*

$$Ax = b \tag{1}$$

is solvable if and only if $AA^+b = b$.

2. *If there are infinitely many solutions, the solution* A^+b *is the one with the smallest norm, i.e.,*

$$\|A^+b\| = \min\{\|x\| : Ax = b\}.$$

Moreover, it is the unique solution of smallest norm.

Example

- The solutions of the underdetermined system $x + y = 1$ geometrically represent an affine line. Matricially, $A = \begin{bmatrix} 1 & 1 \end{bmatrix}$, $b = 1$. Hence, $A^+ b = A^+ 1$ is the point on the line, which is the nearest to the origin. Thus, the vector of this point is perpendicular to the line.
- The solutions of the underdetermined system $x + 2y + 3z = 5$ geometrically represent an affine hyperplane. Matricially, $A = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$, $b = 5$. Hence, $A^+ b = A^+ 5$ is the point on the hyperplane, which is the nearest to the origin. Thus, the vector of this point is normal to the hyperplane.
- The solutions of the underdetermined system $x + y + z = 1$ and $x + 2y + 3z = 5$ geometrically represent an affine line in $\mathbb{R}^3$. Matricially, $A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix}$, $b = \begin{bmatrix} 1 \\ 5 \end{bmatrix}$. Hence, $A^+ b$ is the point on the line, which is the nearest to the origin. Thus, the vector of this point is perpendicular to the line.

Proof of Theorem.

We already know that $Ax = b$ is solvable iff Gb is a solution, where G is any generalized inverse of A. Since A^+ is one of the generalized inverses, this proves the first part of the theorem.

To prove the second part of the theorem, first recall that all the solutions of the system are precisely the set

$$\{A^+b + (A^+A - I)z \colon z \in \mathbb{R}^m\}.$$

So we have to prove that for every $z \in \mathbb{R}^m$,

$$\|A^+b\| \leq \|A^+b + (A^+A - I)z\|.$$

We have that:

$$\|A^+b + (A^+A - I)z\|^2 =$$
$$= \left(A^+b + (A^+A - I)z\right)^T \left(A^+b + (A^+A - I)z\right)$$
$$= \left(A^+b\right)^T \left(A^+b\right) + 2\left(A^+b\right)^T (A^+A - I)z + \left((A^+A - I)z\right)^T \left((A^+A - I)z\right)$$
$$= \|A^+b\|^2 + 2\left(A^+b\right)^T (A^+A - I)z + \|(A^+A - I)z\|^2$$

Now,

$$\begin{aligned}\left(A^+b\right)^T(A^+A-I)z &= b^T(A^+)^T(A^+A-I)z\\ &= b^T(A^+)^T(A^+A)^Tz - b^T(A^+)^Tz\\ &= b^T\left((A^+A)A^+\right)^Tz - b^T(A^+)^Tz\\ &= b^T\left(A^+AA^+\right)^Tz - b^T(A^+)^Tz\\ &= b^T(A^+)^Tz - b^T(A^+)^Tz = 0,\end{aligned}$$

where we used the fact $(A^+A)^T = A^+A$ in the second equality.

Thus,

$$\|A^+b + (A^+A-I)z\|^2 = \|A^+b\|^2 + \|(A^+A-I)z\|^2 \geq \|A^+b\|^2,$$

with the equality iff $(A^+A-I)z = 0$. This proves the second part of the theorem. □

Example

Find the point on the plane $3x + y + z = 2$ closest to the origin.

- In this case,
$$A = \begin{bmatrix} 3 & 1 & 1 \end{bmatrix} \quad \text{and} \quad b = [2].$$
- We have that $AA^T = [11]$ and hence its only eigenvalue is $\lambda = 11$ with eigenvector $u = [1]$, implying that
$$U = [1] \quad \text{and} \quad \Sigma = \begin{bmatrix} \sqrt{11} & 0 & 0 \end{bmatrix}.$$
- Hence,
$$v_1 = \frac{A^T u}{\|A^T u\|} = \frac{A^T u}{\sigma_1} = \frac{1}{\sqrt{11}} \begin{bmatrix} 3 & 1 & 1 \end{bmatrix}^T.$$
-
$$A^+ = V\Sigma^+ U^T = \frac{1}{\sqrt{11}} \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix} \frac{1}{\sqrt{11}} [1] = \begin{bmatrix} \frac{3}{11} \\ \frac{1}{11} \\ \frac{1}{11} \end{bmatrix}.$$
-
$$x^+ = A^+ b = \begin{bmatrix} \frac{6}{11} & \frac{2}{11} & \frac{2}{11} \end{bmatrix}^T.$$

Overdetermined systems

Let $A \in \mathbb{R}^{n \times m}$, where $n > m$. This system is called overdetermined, since here are more constraints than variables. Such a system typically has no solutions, but it might have one or even infinitely many solutions.

Least squares approximation problem: if the system $Ax = b$ has no solutions, then a best fit for the solution is a vector x such that the error $||Ax - b||$ or, equivalently in the row decomposition

$$A = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix},$$

its square

$$||Ax - b||^2 = \sum_{i=1}^{n} (\alpha_i x - b_i)^2,$$

is the smallest possible.

Theorem

If the system $Ax = b$ has no solutions, then $x^+ = A^+ b$ is the unique solution to the least squares approximation problem:

$$||Ax^+ - b|| = \min\{\|Ax - b\| \colon x \in \mathbb{R}^n\}.$$

Proof.

Let $A = U\Sigma V^T$ be the SVD decomposition of A. We have that

$$\|Ax - b\| = \|U\Sigma V^T - b\| = \|\Sigma V^T - U^T b\|,$$

where we used that

$$\|U^T v\| = \|v\|$$

in the second equality (which holds since U^T is an orthogonal matrix). Let

$$\Sigma = \begin{bmatrix} S & 0 \\ 0 & 0 \end{bmatrix}, \quad U = \begin{bmatrix} U_1 & U_2 \end{bmatrix}, \quad V = \begin{bmatrix} V_1 & V_2 \end{bmatrix}, \quad \text{where}$$

$S \in \mathbb{R}^{r\times r}$, $U_1 \in \mathbb{R}^{n\times r}$, $U_2 \in \mathbb{R}^{n\times(n-r)}$, $V_1 \in \mathbb{R}^{m\times r}$, $V_2 \in \mathbb{R}^{m\times(m-r)}$.

Thus,

$$\|\Sigma V^T - U^T b\| = \left\| \begin{bmatrix} S & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} x - \begin{bmatrix} U_1^T \\ U_2^T \end{bmatrix} b \right\|$$
$$= \left\| \begin{bmatrix} S V_1^T x - U_1^T b \\ U_2^T b \end{bmatrix} \right\|.$$

But this norm is minimal iff

$$S V_1^T x - U_1^T b = 0$$

or equivalently

$$x = V_1 S^{-1} U_1^T b = A^+ b.$$

□

Remark

The closest vector to b in the column space $C(A) = \{Ax : x \in \mathbb{R}^m\}$ of A is the orthogonal projection of b onto $C(A)$. It follows that $A^+ b$ is this projection. Equivalently, $b - (A^+ b)$ is orthogonal to any vector Ax, $x \in \mathbb{R}^m$, which can be proved also directly.

Example

Given points $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ in the plane, we are looking for the line $ax + b = y$ which is the least squares best fit.

If $n > 2$, we obtain an overdetermined system

$$\begin{bmatrix} x_1 & 1 \\ \vdots & \\ x_n & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}.$$

The solution of the least squares approximation problem is given by

$$\begin{bmatrix} a \\ b \end{bmatrix} = A^+ \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}.$$

The line $y = ax + b$ in the regression line.