COMP229: Introduction to Data Science
Lecture 24: Singular Value Decomposition

Vitaliy Kurlin, vitaliy.kurlin@liverpool.ac.uk
Autumn 2018, Computer Science department
University of Liverpool, United Kingdom

Sample high-dimensional data

The aim is to look at PCA (Principal Component Analysis) from a different point of view introducing the more general Singular Value Decomposition.

Recall that sample data can be represented by a $k \times n$ matrix S, where s_{ij} is the j–th sample value of the i–th feature, i.e. rows of S correspond to measurement types, while columns reflect trials.

If rows of S have zero means, consider the $n \times k$ matrix $W = \dfrac{S^T}{\sqrt{n-1}}$ whose columns have means 0.

The covariance matrix of data S

Then $W^T W = \dfrac{(S^T)^T S^T}{n-1} = \dfrac{SS^T}{n-1}$ is the covariance matrix of the data sample S.

Definition 24.1. A *Singular Value Decomposition* of any $n \times k$ matrix W is $U\Sigma V^T$, where U, V are orthogonal matrices (high-dimensional rotations) and Σ is a diagonal (scaling) matrix with ordered *singular* values $\sigma_1 \geq \sigma_2 \geq \dots 0$ on the diagonal.

Informal interpretation of SVD in Definition 24.1:

any linear map is a rotation $\times$ scaling $\times$ rotation.

Two rotation matrices in the SVD

In $W = U \cdot \Sigma \cdot V^T$ the first $n \times n$ matrix U is a rotation in the space $\mathbb{R}^n$ of trials, the last $k \times k$ matrix V is a rotation in the space $\mathbb{R}^k$ of features.

Revise Definition 19.6 of an orthogonal matrix: $V^{-1} = V^T$. Let $\vec{u}_1, \dots, \vec{u}_n$ be the columns of U, and $\vec{v}_1, \dots, \vec{v}_k$ be the columns of V.

Claim 23.1 (for columns instead of rows): orthogonality $U^T U = I = V^T V$ means that

$\vec{u}_1, \dots, \vec{u}_n$ is an orthonormal basis in $\mathbb{R}^n$: $\vec{u}_i \cdot \vec{u}_j = \delta_{ij}$.

$\vec{v}_1, \dots, \vec{v}_k$ is an orthonormal basis in $\mathbb{R}^k$: $\vec{v}_i \cdot \vec{v}_j = \delta_{ij}$.

The basis maps to the scaled basis

The $n \times k$ matrix Σ in $W = U \cdot \Sigma \cdot V^T$ has diagonal elements $\Sigma_{ii} = \sigma_i$ for $i \leq \min\{k, n\}$, 0 otherwise.

Claim 24.2. $f : \vec{v} \mapsto W\vec{v}$ satisfies $f(\vec{v}_i) = \sigma_i \vec{u}_i$.

Proof. The SVD formula $W = U \cdot \Sigma \cdot V^T$ implies that $W \cdot V = (U\Sigma)(V^T V) = U \cdot \Sigma$. Split the matrix identity into the identities for columns:
the i-th column of $W \cdot V$ is the vector $W\vec{v}_i$,
the i-th column of $U \cdot \Sigma$ is $\sigma_i \vec{u}_i$.

So f maps basis vectors $\vec{v}_i$ to the scaled vectors $\sigma_i \vec{u}_i$ for $i \leq \min\{k, n\}$, otherwise to $\vec{0}$. □

The covariance matrix M of S

The original data sample $k \times n$ matrix S has the following covariance matrix (from Lecture 23):

$M = W^T W = (U \cdot \Sigma \cdot V^T)^T (U \cdot \Sigma \cdot V^T) = (V \Sigma^T U^T) U \Sigma V^T = V(\Sigma^T \Sigma) V^{-1}$, where $\Sigma^T \Sigma$ is the $k \times k$ matrix with the diagonal elements σ_i^2.

Revise PCA in Lecture 23: $M = CDC^{-1}$, where $C = V$ is an orthogonal matrix with columns equal to eigenvectors of M, D is the diagonal matrix consisting of the eigenvalues of M.

Columns of the matrices U, V

Claim 24.3. In $W = U \cdot \Sigma \cdot V^T$ the columns of V (*right-singular* vectors of W) are the eigenvectors of $W^T W$. The columns of U (called *left-singular* vectors of W) are the eigenvectors of WW^T.

Proof of the 2nd part is similar to the 1st above: $WW^T = U \cdot \Sigma \cdot V^T (U \cdot \Sigma \cdot V^T)^T =$ $U \cdot \Sigma \cdot V^T (V \Sigma^T U^T) = U(\Sigma\Sigma^T)U^{-1}$. The final expression with the diagonal $n \times n$ matrix $\Sigma\Sigma^T$ consisting of σ_i^2 on the diagonal means that σ_i^2 are the eigenvalues of WW^T (or $W^T W$). $\square$

An example of the SVD

Data S: $k = 2$ marks, $n = 5$ students, 0 means.

Subjects	1	2	3	4	5
Maths	1	0	0	–1	0
English	0	1	0	0	–1

Then $W = \dfrac{S^T}{\sqrt{n-1}} = \dfrac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ -1 & 0 \\ 0 & -1 \end{pmatrix}$.

Find the covariance matrix $M = W^T W$.

Eigenvalues and singular values

The covariance $M = W^T W = \begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \end{pmatrix}$ has the eigenvalues $\lambda_1 = \lambda_2 = 0.5$ and singular values $\sigma_1 = \sigma_2 = \frac{1}{\sqrt{2}}$ with the orthonormal eigenvectors $\vec{v}_1 = (1, 0)^T$ and $\vec{v}_2 = (0, 1)^T$. Then $V = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

and $\Sigma = \dfrac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$. Write the SVD for W.

The new 5×5 matrix WW^T

The columns of U in the SVD formula $W = U\Sigma V^T$ are orthonormal eigenvectors of WW^T.

$$W = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ -1 & 0 \\ 0 & -1 \end{pmatrix}, \frac{1}{4}\begin{pmatrix} 1 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 & 1 \end{pmatrix}$$

$= WW^T$ has eigenvectors equal to the columns of U (the left singular vectors of W). Find them!

Eigenvalues λ_i of $W^T W$ and WW^T

The non-zero eigenvalues are the same as for WW^T: $\lambda_1 = \lambda_2 = 0.5 > \lambda_3 = \lambda_4 = \lambda_5 = 0$.

Check: divide the eigenvalues of $4WW^T$ by 4

$\det(4WW^T - \lambda I) = -\lambda^3(\lambda^2 - 4\lambda + 4) = 0$.

Instead of directly finding eigenvectors of WW^T, use Claim 24.2: $\vec{u}_i = \dfrac{1}{\sigma_i} W\vec{v}_i$ for $i = 1, 2$ and get $\vec{u}_1 = \frac{1}{\sqrt{2}}(1, 0, 0, -1, 0)^T$, $\vec{u}_2 = \frac{1}{\sqrt{2}}(0, 1, 0, 0, -1)^T$.

If $\lambda_3 = \lambda_4 = \lambda_5 = 0$, then $\vec{u}_3 = (0, 0, 1, 0, 0)^T$,

$\vec{u}_4 = \frac{1}{\sqrt{2}}(1, 0, 0, 1, 0)^T$, $\vec{u}_5 = \frac{1}{\sqrt{2}}(0, 1, 0, 0, 1)^T$.

SVD formula: $W = U\Sigma V^T$

$$W = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ -1 & 0 \\ 0 & -1 \end{pmatrix},\ U = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & \frac{1}{\sqrt{2}} & 0 & 0 & \frac{1}{\sqrt{2}} \\ 0 & 0 & 1 & 0 & 0 \\ -\frac{1}{\sqrt{2}} & 0 & 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & -\frac{1}{\sqrt{2}} & 0 & 0 & \frac{1}{\sqrt{2}} \end{pmatrix},$$

$$\Sigma = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 \\ 0 & \frac{1}{\sqrt{2}} \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix},\ V = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$ Check now that $W = U\Sigma V^T$.

- Singular value decomposition $W = U\Sigma V^T$, where U, V are orthogonal, Σ is diagonal with square roots of eigenvalues of $W^T W$.