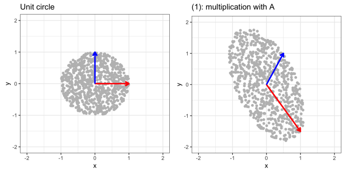
Algorithms and Data Structures

Matrix Approximation
Singular Value Decomposition & Principal Component Analysis



Learning goals

- Singular value decomposition
- Principal component analysis

REMINDER: SINGULAR VALUE DECOMPOSITION

For a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ of rank r, there exists a decomposition

$$\mathbf{A} = \mathbf{U} \boldsymbol{\Sigma} \boldsymbol{V}^\top$$

with $\boldsymbol{U} \in \mathbb{R}^{m \times m}$ and $\boldsymbol{V} \in \mathbb{R}^{n \times n}$ are orthogonal matrices, and $\boldsymbol{\Sigma} \in \mathbb{R}^{m \times n}$ is a diagonal matrix with non-negative diagonal entries sorted in descending order, i.e. $\sigma_1 \geq \sigma_2 \geq \ldots$

$$\left(\begin{array}{ccc|ccc}
\sigma_1 & & & & \vdots & \\
 & \ddots & & \cdots & 0 & \cdots \\
 & & \sigma_r & & \vdots & \\
\hline
 & \vdots & & & \vdots & \\
\cdots & 0 & \cdots & \cdots & 0 & \cdots \\
 & \vdots & & & \vdots &
\end{array}\right)$$

REMINDER: SINGULAR VALUE DECOMPOSITION / 2

Definition:

- The diagonal elements of the matrix $\mathbf{\Sigma}$ are known as **singular values** of the matrix $\mathbf{A}$
- The column vectors of $\mathbf{U}$ are called **left singular vectors**
- The row vectors of $\mathbf{V}$ are called **right singular vectors**

A non-negative real number σ is a singular value if both left and right singular vectors $\mathbf{u}$ and $\mathbf{v}$ exist, such that

$$\begin{aligned} \mathbf{A}\mathbf{v} &= \sigma\mathbf{u} \\ \mathbf{A}^\top\mathbf{u} &= \sigma\mathbf{v} \end{aligned}$$

REMINDER: SINGULAR VALUE DECOMPOSITION / 3

A **truncated** singular value decomposition of rank $k \leq r$ is given by

$$\mathbf{U}_k \boldsymbol{\Sigma}_k \mathbf{V}_k^\top$$

where $\boldsymbol{\Sigma}_k \in \mathbb{R}^{k \times k}$ only contains the k largest singular values and $\mathbf{U}_k$ and $\mathbf{V}_k$ the corresponding left/right singular vectors.

REMINDER: SINGULAR VALUE DECOMPOSITION / 4

"**Intuition**":

- Each matrix defines a matrix transformation $\mathbf{x} \mapsto \mathbf{A}\mathbf{x}$. The singular value decomposition splits this transformation into a rotation / mirror ($\mathbf{x} \mapsto \mathbf{V}^\top\mathbf{x}$), a scaling ($\mathbf{x} \mapsto \Sigma\mathbf{x}$) and another rotation / mirror ($\mathbf{x} \mapsto \mathbf{U}\mathbf{x}$).
- In 2D, the singular values can be interpreted as the magnitude of the semiaxis of the ellipse defined by $\mathbf{A}$.
- The columns of $\mathbf{U}$ form an orthonormal basis for the column space of $\mathbf{A}$, the columns of $\mathbf{V}$ span the row space of $\mathbf{A}$.

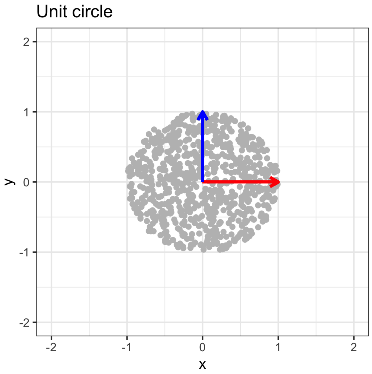
REMINDER: SINGULAR VALUE DECOMPOSITION

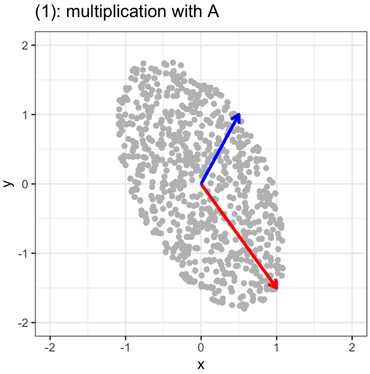
/ 5

Example:

Consider $\mathbf{A} = \begin{pmatrix} 1 & \frac{1}{2} \\ -\frac{3}{2} & 1 \end{pmatrix}$.

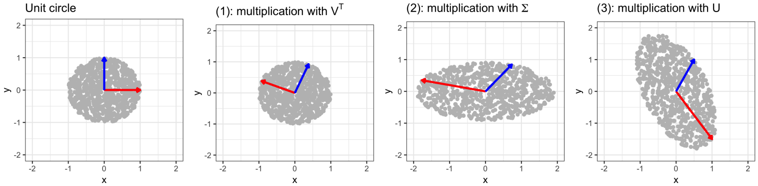
The **A** matrix defines a linear transformation.





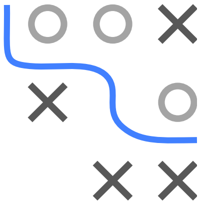
REMINDER: SINGULAR VALUE DECOMPOSITION
/ 6

It can be decomposed using the singular value decomposition:



Note: The red / blue vectors are the canonical unit vectors $(1, 0)^\top$ and $(0, 1)^\top$ and their transformations after the respective matrix multiplications.

EXCURSUS: PRINCIPAL COMPONENT ANALYSIS



Given: n data points with p features $^{(*)}$ each

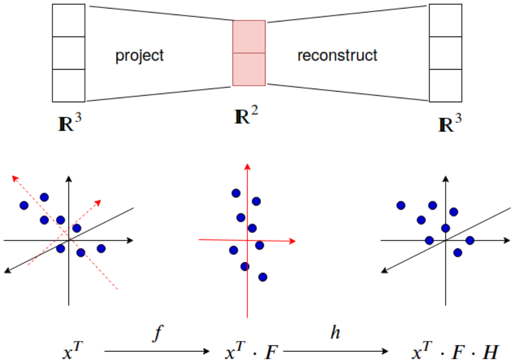
Goal: Projection of the n data points into a k-dimensional space ($k < p$) with as little information loss as possible

Idea:

- Find a linear tranformation $f : \mathbb{R}^p \to \mathbb{R}^k$, which maps each observation $\boldsymbol{x} \in \mathbb{R}^p$ to a k-dimensional point $\boldsymbol{z}$.
- Lose as little information as possible through this dimensionality reduction.
- As little information as possible is lost if we can reconstruct the point $\boldsymbol{z}$ as good as possible, i.e. we can use a linear function $h : \mathbb{R}^k \to \mathbb{R}^p$, such that $\mathbf{x} \approx h(\boldsymbol{z})$.

$^{(*)}$ We assume the data points are centered around 0.

EXCURSUS: PRINCIPAL COMPONENT ANALYSIS / 2



The linear transformations f, h are described by matrix multiplication:
$f : \mathbf{x}^\top \mapsto \mathbf{x}^\top \boldsymbol{F} =: \boldsymbol{z}$ and $h : \boldsymbol{z}^\top \mapsto \boldsymbol{z}^\top \boldsymbol{H}$

Note: Here, we are writing $\mathbf{x}$ as a **row vector** $\mathbf{x}^\top$, to be in line with the matrix notation in the following slides (the observations are the rows of the design matrix $\boldsymbol{X}$).

EXCURSUS: PRINCIPAL COMPONENT ANALYSIS / 3

Goal: Minimize the reconstruction error between data $\boldsymbol{X} \in \mathbb{R}^{n \times p}$ and the projected and reconstructed data $\boldsymbol{XFH}$.

$$\min_{\boldsymbol{F} \in \mathbb{R}^{p \times k}, \boldsymbol{H} \in \mathbb{R}^{k \times p}} \|\boldsymbol{X} - \boldsymbol{XFH}\|_F^2$$

Defining $\boldsymbol{XF} =: \boldsymbol{W} \in \mathbb{R}^{n \times k}$, we write this as

$$\min_{\boldsymbol{W} \in \mathbb{R}^{n \times k}, \boldsymbol{H} \in \mathbb{R}^{k \times p}} \|\boldsymbol{X} - \boldsymbol{WH}\|_F^2.$$

This is the problem of matrix approximation. One solution is

$$\boldsymbol{XF} \quad = \quad \boldsymbol{W} = \mathbf{U}_k \Sigma_k; \quad \boldsymbol{H} = \mathbf{V}_k^\top,$$

with $\boldsymbol{U}_k \in \mathbb{R}^{n \times k}, \Sigma_k \in \mathbb{R}^{k \times k}, \boldsymbol{V}_k \in \mathbb{R}^{p \times k}$ chosen as truncated singular value decomposition of $\boldsymbol{X}$.

EXCURSUS: PRINCIPAL COMPONENT ANALYSIS / 4

$\boldsymbol{H} = \boldsymbol{V}_k^\top \in \mathbb{R}^{k \times p}$ is the reconstruction transformation matrix. The projection matrix $\boldsymbol{F} = \boldsymbol{V}_k \in \mathbb{R}^{p \times k}$ fulfills $\boldsymbol{XF} = \boldsymbol{U}_k \Sigma_k$:

$$\boldsymbol{XF} \;=\; \boldsymbol{XV}_k = \boldsymbol{U}\Sigma\boldsymbol{V}^\top\boldsymbol{V}_k = \boldsymbol{U}\Sigma\begin{pmatrix}\boldsymbol{I}_k \\ \boldsymbol{0}_{p-k}\end{pmatrix} = \boldsymbol{U}\begin{pmatrix}\Sigma_k \\ \boldsymbol{0}_{n-k}\end{pmatrix} = \boldsymbol{U}_k\Sigma_k,$$

- The rows of $\boldsymbol{XF} = \boldsymbol{U}_k\Sigma_k \in \mathbb{R}^{n \times k}$ are the projected observations.
- It can be shown (see next slide), that the rows of $\boldsymbol{H} = \boldsymbol{V}_k^\top \in \mathbb{R}^{k \times p}$ correspond to the k (pair-wise orthogonal) principal components.

EXCURSUS: PRINCIPAL COMPONENT ANALYSIS / 5

A more common motivation of PCA is the following: Linearly transform the data to a new coordinate system such that the greatest variance in the (transformed) data is along the first PC, the second greatest variance is along a second PC orthogonal to the first PC, etc.

- In this formulation, it can be shown that the k first principal components correspond to the k eigenvectors with the greatest eigenvalues of the covariance matrix $\boldsymbol{X}^\top \boldsymbol{X}$.
- The eigenvalue decomposition $\boldsymbol{X}^\top \boldsymbol{X}$ and the singular value decomposition of $\boldsymbol{X}$ are related. Given the singular value decomposition of $\boldsymbol{X}$, we can derive the eingevalue decomposition of $\boldsymbol{X}^\top \boldsymbol{X}$:

$$\boldsymbol{X}^\top \boldsymbol{X} \quad = \quad \boldsymbol{V} \Sigma \boldsymbol{U}^\top \boldsymbol{U} \Sigma \boldsymbol{V}^\top = \boldsymbol{V} \widehat{\Sigma}^2 \boldsymbol{V}^\top$$

with $\widehat{\boldsymbol{\Sigma}}^2 := \boldsymbol{\Sigma}^\top \boldsymbol{\Sigma} \in \mathbb{R}^{p \times p}$ having the squared singular values of $\boldsymbol{X}$ on the diagonal.

EXCURSUS: PRINCIPAL COMPONENT ANALYSIS / 6

- The right singular vectors $\boldsymbol{V}$ of $\boldsymbol{X}$ are equivalent to the eigenvectors of $\boldsymbol{X}^\top\boldsymbol{X}$, and the singular values of $\boldsymbol{X}$ are equal to the square-root of the eigenvalues of $\boldsymbol{X}^\top\boldsymbol{X}$. So we come up with the same solution for both approaches.