Computational data mining

Fatemeh Mansoori

- For real symmetric matrices S : real eigenvalues and orthogonal eigenvectors
- *A* is not square then $Ax = \lambda x$ is impossible eigenvectors fail (left side in R^m, right side in R^n)
- Singular Value Decomposition fills this gap
- Suppose *A* is often a matrix of data.
  - rows could tell us the age and height of 1000 children
  - *A* is 2 by 1000: definitely rectangular
  - rank is $r = 2$ and that matrix *A* has two positive singular values σ1 and σ2 .

- need two sets of singular vectors the *u's* and the *v's*
- real *m* by *n* matrix :
  - the *n* right singular vectors v1 , ... , vn are orthogonal in R^n
  - The *m* left singular vectors u1 , ... , *um* are perpendicular to each other in R^m
  - For singular vectors, each *Av* equals σu

$$\boxed{Av_1 = \sigma_1 u_1 \quad \cdot\cdot \quad Av_r = \sigma_r u_r} \qquad \boxed{Av_{r+1} = 0 \quad \cdot\cdot \quad Av_n = 0}$$

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

  - last *n- r v's* are in the nullspace of *A*
  - last m- *r* u's are in the nullspace of A^T

- All of the right singular vectors v1 to *vn* go in the columns of *V*
- left singular vectors u1 to *um* go in the columns of *U*
- The columns of V and U are orthogonal unit vectors

$$V^{\mathrm{T}} = V^{-1} \text{ and } U^{\mathrm{T}} = U^{-1}$$

$$\boldsymbol{AV} = \boldsymbol{U\Sigma} \qquad A\begin{bmatrix} \\ \boldsymbol{v}_1 \,..\, \boldsymbol{v}_r \,..\, \boldsymbol{v}_n \\ \\ \end{bmatrix} = \begin{bmatrix} \\ \boldsymbol{u}_1 \,..\, \boldsymbol{u}_r \,..\, \boldsymbol{u}_m \\ \\ \end{bmatrix} \left[\begin{array}{ccc|c} \boldsymbol{\sigma_1} & & & \\ & \ddots & & \mathbf{0} \\ & & \boldsymbol{\sigma_r} & \\ \hline & \mathbf{0} & & \mathbf{0} \end{array}\right]$$

- basis of *v's* for the row space of *A* and then *u's* for the column space.

$$AX = X\Lambda. \text{ But } AV = U\Sigma$$

$$\boxed{\textbf{The Singular Value Decomposition of } \boldsymbol{A} \textbf{ is} \quad \boldsymbol{A = U\Sigma V^{\mathrm{T}}}.}$$

Pieces of the SVD

$$\boxed{\boldsymbol{A = U\Sigma V^{\mathrm{T}} = \sigma_1 u_1 v_1^{\mathrm{T}} + \cdots + \sigma_r u_r v_r^{\mathrm{T}}}.}$$

Example 1
$AV = U\Sigma$

$$\begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} = \frac{1}{\sqrt{10}} \begin{bmatrix} 1 & -3 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 3\sqrt{5} & \\ & \sqrt{5} \end{bmatrix}$$

- first piece is more important than the second piece because the first singular values is grater
- To recover *A,* add the pieces $\sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathrm{T}} + \sigma_2 \boldsymbol{u}_2 \boldsymbol{v}_2^{\mathrm{T}}$
-

$$\frac{3\sqrt{5}}{\sqrt{10}\sqrt{2}} \begin{bmatrix} 1 \\ 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \end{bmatrix} + \frac{\sqrt{5}}{\sqrt{10}\sqrt{2}} \begin{bmatrix} -3 \\ 1 \end{bmatrix} \begin{bmatrix} -1 & 1 \end{bmatrix} = \frac{3}{2} \begin{bmatrix} 1 & 1 \\ 3 & 3 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} 3 & -3 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix}$$

Reduced Form of the SVD

$$\boldsymbol{A}\boldsymbol{V}_r = \boldsymbol{U}_r\boldsymbol{\Sigma}_r \qquad A\left[\begin{array}{ccc} \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_r \\ & \text{row space} & \end{array}\right] = \left[\begin{array}{ccc} \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_r \\ & \text{column space} & \end{array}\right]\left[\begin{array}{ccc} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{array}\right]$$

First Proof of the SVD

- want to identify the two sets of singular vectors, the *u* 's and the *v's*

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = (V\Sigma^{\mathrm{T}}U^{\mathrm{T}})\ (U\Sigma V^{\mathrm{T}}) = \boldsymbol{V\Sigma^{\mathrm{T}}\Sigma V^{\mathrm{T}}}$$

$$\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} = (U\Sigma V^{\mathrm{T}})\ (V\Sigma^{\mathrm{T}}U^{\mathrm{T}}) = \boldsymbol{U\Sigma\Sigma^{\mathrm{T}}U^{\mathrm{T}}}$$

Both right hand sides have the special form $Q\Lambda Q^{\mathrm{T}}$.

Eigenvalues are in $\Lambda = \Sigma^{\mathrm{T}}\Sigma$

V contains orthonormal eigenvectors of $A^{\mathrm{T}}A$

U contains orthonormal eigenvectors of AA^{T}

σ_1^2 to σ_r^2 are the nonzero eigenvalues of both $A^{\mathrm{T}}A$ and AA^{T}

v's then u's $\quad A^{\mathrm{T}}Av_k = \sigma_k^2 v_k$ and then $u_k = \dfrac{Av_k}{\sigma_k}$ for $k = 1, \ldots, r$

$$\boldsymbol{v}\textbf{'s then } \boldsymbol{u}\textbf{'s} \qquad \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{v}_k = \boldsymbol{\sigma}_k^2\boldsymbol{v}_k \quad \text{and then} \quad \boldsymbol{u}_k = \frac{\boldsymbol{A}\boldsymbol{v}_k}{\boldsymbol{\sigma}_k} \quad \text{for} \quad k = 1, \ldots, r$$

- Prove that u's are the eigenvectors of AA^T

- Prove that u's ate orthogonal

Question

Find the matrices U, Σ, V for $A = \begin{bmatrix} \mathbf{3} & \mathbf{0} \\ \mathbf{4} & \mathbf{5} \end{bmatrix}$.

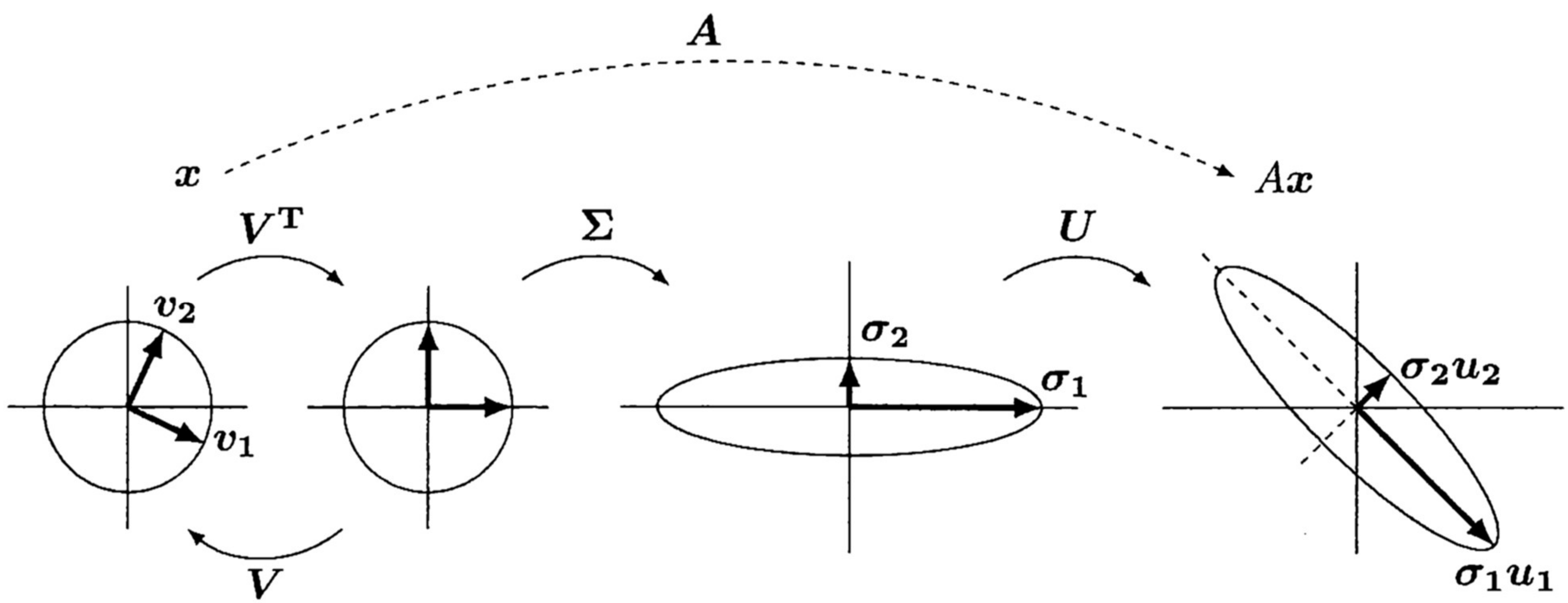
$$A^{\mathrm{T}}A = \begin{bmatrix} 25 & 20 \\ 20 & 25 \end{bmatrix} \qquad \sigma_1^2 = 45 \text{ and } \sigma_2^2 = 5.$$

$$\begin{bmatrix} 25 & 20 \\ 20 & 25 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \mathbf{45}\begin{bmatrix} \mathbf{1} \\ \mathbf{1} \end{bmatrix} \qquad \begin{bmatrix} 25 & 20 \\ 20 & 25 \end{bmatrix}\begin{bmatrix} -1 \\ 1 \end{bmatrix} = \mathbf{5}\begin{bmatrix} \mathbf{-1} \\ \mathbf{1} \end{bmatrix}$$

$$v_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad v_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} -1 \\ 1 \end{bmatrix} \qquad \textbf{Left singular vectors } \boldsymbol{u}_i = \frac{A\boldsymbol{v}_i}{\sigma_i}$$

$$\boxed{\boldsymbol{U} = \frac{1}{\sqrt{10}}\begin{bmatrix} 1 & -3 \\ 3 & 1 \end{bmatrix} \qquad \boldsymbol{\Sigma} = \begin{bmatrix} \sqrt{45} & \\ & \sqrt{5} \end{bmatrix} \qquad \boldsymbol{V} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}}$$

The Geometry of the SVD



Question

Question : If $S = Q\Lambda Q^{\mathrm{T}}$ is symmetric positive definite, what is its SVD ?

Question : If $A = Q$ is an orthogonal matrix, why does every singular value equal 1 ?

What is the SVD decomposition of Q

Question : Why are all eigenvalues of a square matrix A less than or equal to σ_1 ?

Question : If $A = \boldsymbol{x}\boldsymbol{y}^{\mathrm{T}}$ has rank 1, what are $\boldsymbol{u}_1$ and $\boldsymbol{v}_1$ and σ_1 ? **Check that $|\lambda_1| \leq \sigma_1$.**

First Singular Vector v1

Maximize the ratio $\frac{||Ax||}{||x||}$. **The maximum is** σ_1 **at the vector** $x = v_1$.

Polar decomposition

- Show that every matrix A could be Factor into a orthogonal matrix and a symmetric matrix

Polar decomposition	$A = U\Sigma V^{\mathrm{T}} = (UV^{\mathrm{T}})(V\Sigma V^{\mathrm{T}}) = (Q)(S).$

- Show that the product of orthogonal matrix is orthogonal

Check for the matrix $A = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix}$

Principle component and best low rank matrix

- principal components of A are its singular vectors, the columns uj and vj of the orthogonal matrices U and V.
- Principal Component Analysis (PCA) uses the largest *singular values* connected to the first *u's* and *v's* to understand the information in a matrix of data

We are given a matrix A, and we extract its most important part A_k (**largest σ's**):

$$\boldsymbol{A_k = \sigma_1 u_1 v_1^{\mathrm{T}} + \cdots + \sigma_k u_k v_k^{\mathrm{T}}} \quad \textbf{with rank } \boldsymbol{(A_k) = k}.$$

- The closest rank k matrix to A is A_k

Eckart-Young **If** B **has rank** k **then** $\|\|A - B\|\| \geq \|\|A - A_k\|\|$.

Matrix norm

- What is the meaning of the symbol || A ||
- "norm" of the matrix *A :*
  - a measure of its size (like the absolute value of a number)

Spectral norm $||\boldsymbol{A}||_{\mathbf{2}} = \max \frac{||A\boldsymbol{x}||}{||\boldsymbol{x}||} = \boldsymbol{\sigma_1}$ (often called the $\boldsymbol{\ell^2}$ **norm**)

Frobenius norm $||\boldsymbol{A}||_{\boldsymbol{F}} = \sqrt{\sigma_1^2 + \cdots + \sigma_r^2}$ (12) and (13) also define $||\boldsymbol{A}||_{\boldsymbol{F}}$

Nuclear norm $||\boldsymbol{A}||_{\boldsymbol{N}} = \boldsymbol{\sigma_1} + \boldsymbol{\sigma_2} + \cdots + \boldsymbol{\sigma_r}$ (the trace norm)

- Find the norms of I and Q

$$||I||_2 = 1 \qquad ||I||_F = \sqrt{n} \qquad ||I||_N = n$$
$$||Q||_2 = 1 \qquad ||Q||_F = \sqrt{n} \qquad ||Q||_N = n.$$

- the norm of A did not change when A is multiplied on eighter side by an orthogonal matrix

Question

- Show that the singular values do not change when U and V are multiplied by Q. (Qu or Qv)

- show that the norm of A did not change when A is multiplied on eighter side by an orthogonal matrix

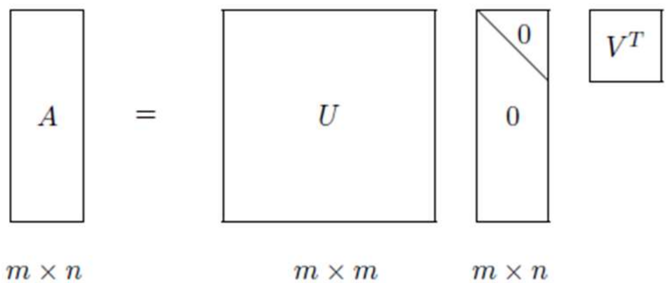
$$||\boldsymbol{A}||_{\boldsymbol{F}}^{\boldsymbol{2}} = |a_{11}|^2 + |a_{12}|^2 + \cdots + |a_{mn}|^2 \quad \text{(every } a_{ij}^2\text{)}$$

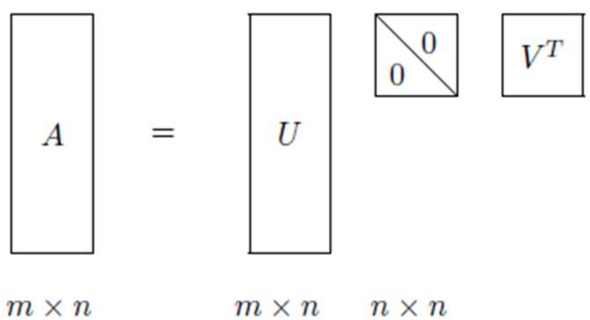
$$||\boldsymbol{A}||_{\boldsymbol{F}}^{\boldsymbol{2}} = \text{trace of } A^{\mathrm{T}}A = (A^{\mathrm{T}}A)_{11} + \cdots + (A^{\mathrm{T}}A)_{nn}$$

$$||\boldsymbol{A}||_{\boldsymbol{F}}^{\boldsymbol{2}} = \boldsymbol{\sigma}_1^2 + \boldsymbol{\sigma}_2^2 + \cdots + \boldsymbol{\sigma}_r^2$$

SVD

- How SVD give us the orthogonal basis for 4 subspaces of a matrix





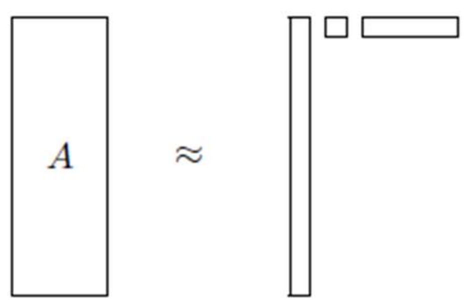
$$y = Ax = \sum_{i=1}^{r} \sigma_i u_i v_i^T x = \sum_{i=1}^{r} (\sigma_i v_i^T x) u_i = \sum_{i=1}^{r} \alpha_i u_i.$$

$$Az = \left(\sum_{i=1}^{r} \sigma_i u_i v_i^T \right) \left(\sum_{i=r+1}^{n} \beta_i v_i \right) = 0.$$

Matrix Approcimation

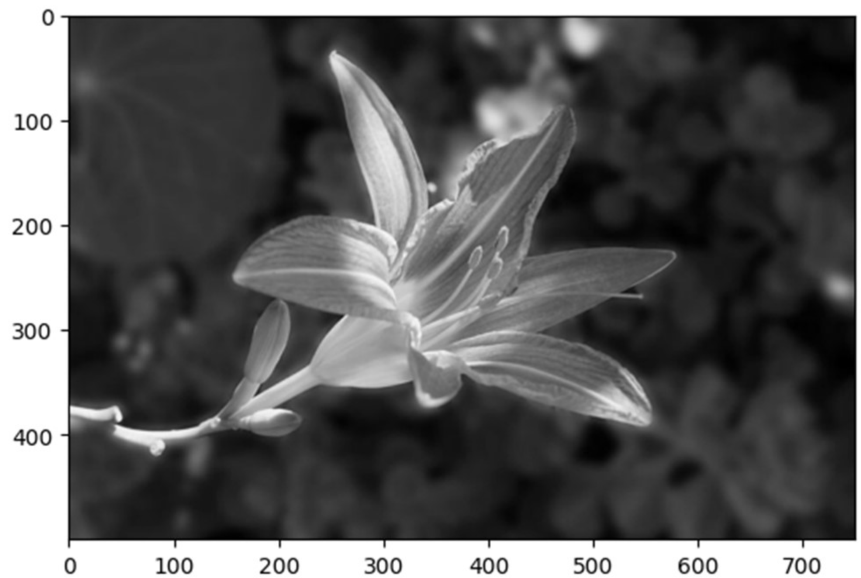
$$A = \sum_{i=1}^{n} \sigma_i u_i v_i^T \approx \sum_{i=1}^{k} \sigma_i u_i v_i^T =: A_k.$$

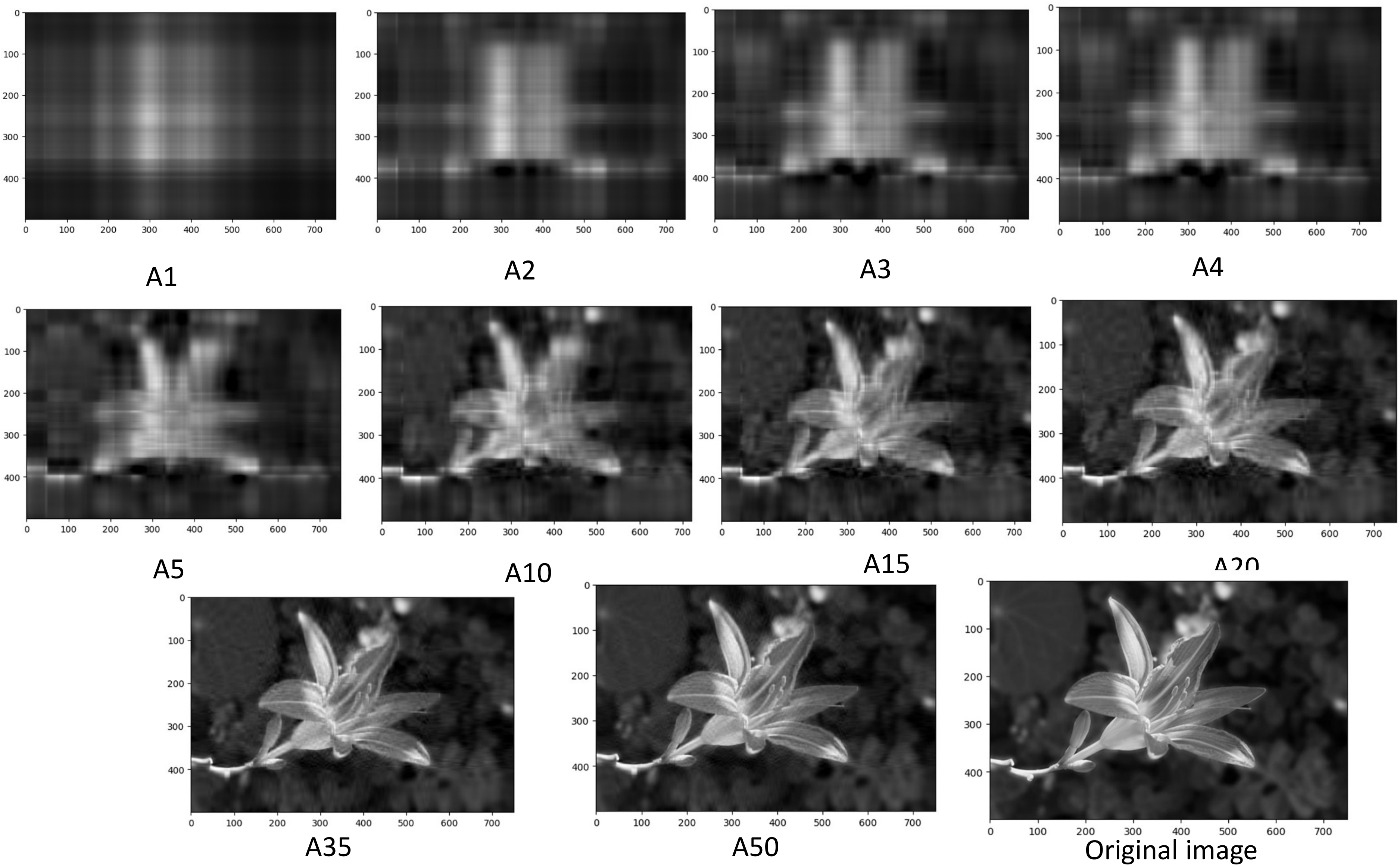
The low-rank approximation of a matrix is illustrated as



$$= \quad U_k \Sigma_k V_k^T.$$

A sample gray scale image

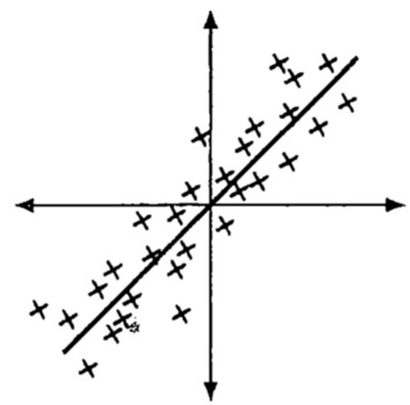



A1
A2
A3
A4
A5
A10
A15
A20
A35
A50
Original image

PCA

- We have n samples
- For each sample we measure m variables (like height and weight)
- data matrix A has n columns and m rows
- First step is to normalize the values
- Often those n points are clustered near a line or a plane or another low-dimensional subspace of Rm
- How will linear algebra find that closest line through (0, 0)
- It is in the direction of the first singular vector of A (direction of u1)

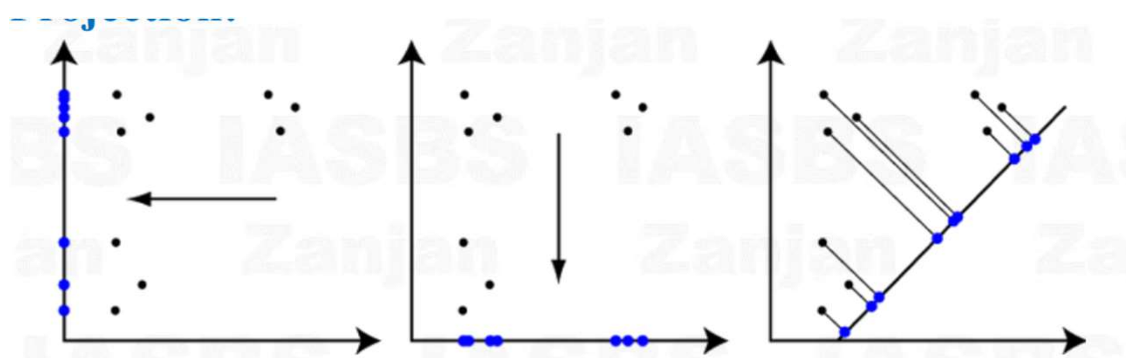
- In what sense will the line in the direction of *u1* be the *closest line* to the centered data

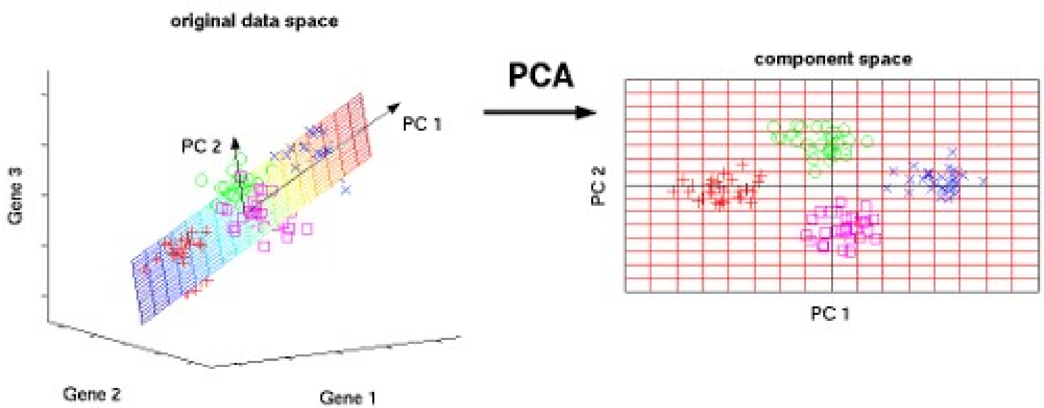


The variances are the diagonal entries of the matrix AA^{T}.

The covariances are the off-diagonal entries of the matrix AA^{T}.

Dimensionality reduction





Question

If A is a 2 by 2 matrix with $\sigma_1 \geq \sigma_2 > 0$, find $||A^{-1}||_2$ and $||A^{-1}||_F^2$.

Find a closest rank-1 approximation in the L^2 norm to $A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$

Find a closest rank-1 approximation to these matrices (L^2 or Frobenius norm)

$$A = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad A = \begin{bmatrix} 0 & 3 \\ 2 & 0 \end{bmatrix} \qquad A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

Question

Show that A^{T} has the same (nonzero) singular values as A. Then $||A|| = ||A^{\mathrm{T}}||$ for all matrices. But it's not true that $||A\boldsymbol{x}|| = ||A^{\mathrm{T}}\boldsymbol{x}||$ for all vectors. That needs $A^{\mathrm{T}}A = AA^{\mathrm{T}}$.