MATH 312 LECTURE 21 : Singular Value Decomposition.

I.) Recall that any (square real symmetric) positive semidefinite matrix A can be factorised as follows:

$A = Q D Q^T$, where Q is orthogonal
D is diagonal with nonnegative entries.

This can be rewritten as

$A = \lambda_1 \vec{q}_1 \vec{q}_1^T + \dots + \lambda_r \vec{q}_r \vec{q}_r^T$, where

- $\lambda_1, \dots, \lambda_r \equiv$ nonzero eigenvalues of A (may be repeated)
- $\vec{q}_1, \dots, \vec{q}_r \equiv$ eigenvectors of A corresponding to $\lambda_1, \dots, \lambda_r$.
  (columns of Q: so orthonormal vectors).
- $r = \text{rank}(A)$ (The spectral theorem ensures that we can find r orthonormal eigenvectors).

Remark: In Q there are n−r additional orthonormal columns, corresponding to $\lambda = 0$ (nullspace of A).

Remark: The order of the λ's in D has to be the same as the order of $\vec{q}$'s in Q.

→ Notice that we have decomposed A as a sum of r matrices with rank 1.

We could order these using the weights $\underline{\underline{\lambda}}$'s.

→ Each piece is a projection matrix onto the line given by the corresponding eigenvector.

→ Since A is symmetric, $\{\vec{q}_1, \dots, \vec{q}_r\}$ is both an orthonormal basis of the column space and row space.

$(C(A) = C(A^T)$ since $A = A^T$!$)$.

For same reason, $\{\vec{q}_{r+1}, \dots, \vec{q}_n\}$ is a basis for nullspace and left nullspace.

- Now, we want to generalise this decomposition for any real matrix A $m \times n$.

  ↳ Singular Value Decomposition.

<u>Remark</u>: We will be able to check that for (symmetric) positive semidefinite matrix the SVD is the same as the above one.

So now, we want to show that any real $m \times n$ matrix A of rank r can be decomposed as follows:

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

1) Σ is an $m \times n$ (pseudo) diagonal matrix, with nonnegative entries.

2) V is an $n \times n$ orthogonal matrix. (*)

3) U is an $m \times m$ orthogonal matrix

Let's first see one way to find such Σ, U, V that always works. Then we will prove why.

(*) Reminder: Orthogonal matrix means orthonormal cols. and rows.

- Recipe for $A = U \Sigma V^T$.

1) $\underline{\underline{\Sigma}}$

1.1) Find eigenvalues of A^TA or AA^T

<u>Remark</u>: The nonzero ones are the same for A^TA and AA^T.

[So use the smaller matrix!]

1.2) Order the nonzero eigenvalues:

$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r > 0$

1.3) Let $\sigma_i = \sqrt{\lambda_i}$ for $i = 1, \dots, r$. $\left\| \sigma\text{'s} \equiv \text{Singular Values of } A. \right\|$

1.4) Finally, construct Σ of size $\underline{m \times n}$:

$$\Sigma = \begin{array}{c} \left[\begin{array}{ccc|ccc} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & \\ \hline & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{array}\right] \begin{array}{l} \left.\vphantom{\begin{array}{c}1\\1\\1\end{array}}\right\} r \\ \left.\vphantom{\begin{array}{c}1\\1\\1\end{array}}\right\} m-r \end{array} \\ \begin{array}{cc} \underbrace{\hphantom{\sigma_1 \ \ddots \ \sigma_r}}_{r} & \underbrace{\hphantom{0 \ \ddots \ 0}}_{n-r} \end{array} \end{array}$$

Example:

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad (5 \times 4)$$

(different in class)

Singular values of A? Σ?

$A^T A \rightarrow 4 \times 4$

$A A^T \rightarrow 5 \times 5$

$$\hookrightarrow A^T A = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 9 \end{bmatrix} \rightarrow \lambda_1 = 9,\ \lambda_2 = 4,\ \lambda_3 = 1,\ \lambda_4 = 0.$$

Thus, $\sigma_1 = 3$, $\sigma_2 = 2$, $\sigma_1 = 1$ and

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

- **Remark:** Recall that rank(A) is equal to the number of singular values (thee in this case).

2) $\underline{\underline{V}}$

2.1) Find r orthonormal eigenvectors of A^TA:

$$\begin{array}{cccccccc} \lambda_1 & \geq & \lambda_2 & \geq & \dots & \geq & \lambda_r & > 0 \\ | & & | & & \dots & & | & \\ \vec{v}_1 & & \vec{v}_2 & & & & \vec{v}_r & \end{array}$$

2.2) Find n−r orthonormal eigenvectors for $\lambda = 0$ in A^TA..

<u>Remark</u>: • In both cases, Gram-Schmidt is needed if an eigenvalue is repeated.

• Vectors have to be unit length.

$$V = \left[\underbrace{\vec{v}_1 \ \dots \ \vec{v}_r}_{r} \ \Big| \ \underbrace{\vec{v}_{r+1} \dots \vec{v}_n}_{n-r} \right] \quad \begin{array}{c}\text{orthogonal} \\ n \times n\end{array}$$

→ The first r cols. have to be in order ‖.

<u>Example</u>:

$$\lambda_1 = 9 \rightarrow A^TA - 9I = \begin{bmatrix} -9 & 0 & 0 & 0 \\ 0 & -8 & 0 & 0 \\ 0 & 0 & -5 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \rightarrow \vec{v}_1 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \quad \dots \lambda_2, \lambda_3, \lambda_4$$

$$\hookrightarrow \quad V = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$
$$\qquad\qquad \begin{array}{cccc} \uparrow & \uparrow & \uparrow & \uparrow \\ \vec{v}_1 & \vec{v}_2 & \vec{v}_3 & \vec{v}_4 \end{array}$$

3) $\underline{\underline{\mathcal{U}}}$

3.1) For $i = 1, \dots, r$, let $\vec{u}_i = \dfrac{A\vec{v}_i}{\sigma_i}$.

3.2) Find $m-r$ orthonormal eigenvectors of AA^T corresponding to $\lambda = 0$:

$\vec{u}_{r+1}, \dots, \vec{u}_m$ (i.e. find an orthonormal basis for $N(AA^T)$).

$$\mathcal{U} = \left[\begin{array}{ccc|ccc} \frac{A\vec{v}_1}{\sigma_1} & \cdots & \frac{A\vec{v}_r}{\sigma_r} & \vec{u}_{r+1} & \cdots & \vec{u}_m \end{array}\right] \quad \begin{array}{c}\text{orthogonal}\\ m \times m\end{array}$$

(order matters!

→ these will be orthonormal (if not, previous mistake!).

Ex:

$$A\begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \vec{v}_3 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 2 & 0 \\ 3 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \left(\vec{u}_i = \frac{A\vec{v}_i}{\sigma_i}\right) \Rightarrow$$

$$\Rightarrow \vec{u}_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \ \vec{u}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \ \vec{u}_3 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

We need $\vec{u}_4, \vec{u}_5 \in N(AA^T)$:

$$AA^T = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & 0 \\ 0 & 0 & 9 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \rightarrow \vec{u}_4 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \ \vec{u}_5 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \rightarrow \mathcal{U} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

[→ Check: $A = \mathcal{U}\Sigma V^T$?]

- We want to see now that such U, Σ, V always exist and that $A = U \Sigma V^T$.

1) $\underline{\underline{\Sigma}}$

1.1) $\text{rank}(A) = \text{rank}(AA^T) = \text{rank}(A^TA)$ $\longrightarrow$

nonzero eigenvalues of $\left| \begin{matrix} A^TA \\ AA^T \end{matrix} \right.$ are the same (thus, with same multiplicity)

1.2) A^TA, AA^T positive semidefinite $\Rightarrow \lambda \geq 0$.

Therefore, there are r positive eigenvalues $\lambda_1, \dots, \lambda_r$ ~~and they~~ ~~[illegible]~~ (maybe repeated),

so we can define the singular values $\sigma_i = \sqrt{\lambda_i}$ $(i = 1, \dots, r)$.

2) $\underline{\underline{V}}$

A^TA is symmetric $\Rightarrow$ There is an orthonormal basis of eigenvectors.

So V would be orthogonal by construction.

3) $\underline{\underline{U}}$

$$U = \left[\frac{A\vec{v}_1}{\sigma_1} \cdots \frac{A\vec{v}_r}{\sigma_r} \;\vdots\; \underbrace{\vec{u}_{r+1} \cdots \vec{u}_m}_{\substack{m-r \text{ orthonormal} \\ \text{vectors in } N(AA^T).}} \right] \equiv \begin{array}{l} m-r \text{ orthonormal} \\ \text{eigenvectors of } AA^T \text{ for } \lambda = 0. \end{array}$$

(possible since $\dim N(AA^T) = m - r$).

Need to check:

3.1) Are the first r columns of U orthonormal?

3.2) Are the first r orthogonal to the other m-r?

3.1) Let's check if: ($i, j \in \{1, \dots, r\}$)

$$\underbrace{\frac{A\vec{v}_i}{\sigma_i}}_{\vec{u}_i} \cdot \underbrace{\frac{A\vec{v}_j}{\sigma_j}}_{\vec{u}_j} = \frac{(A\vec{v}_i)^T}{\sigma_i} \frac{A\vec{v}_j}{\sigma_j} = \frac{\vec{v}_i^T A^T A \vec{v}_j}{\sigma_i \sigma_j} \underset{\substack{\uparrow \\ \vec{v}_j \text{ is eigenvector of } A^TA \\ \text{by our construction of } V}}{=} \frac{\vec{v}_i^T \lambda_j \vec{v}_j}{\sigma_i \sigma_j} =$$

$$= \frac{\lambda_j}{\sigma_i \sigma_j} \vec{v}_i^T \vec{v}_j = \left\{ \begin{array}{ll} 0 & \text{if } i \neq j \\ \underset{\substack{\uparrow \\ \text{by our definition of } \sigma_i.}}{\frac{\lambda_i}{\sigma_i^2} = 1} & \text{if } i = j \end{array} \right| \equiv \begin{array}{l} \{\vec{u}_1, \dots, \vec{u}_r\} \\ \text{are orthonormal} \\ \text{vectors.} \end{array}$$

3.2) Notice that

$$\begin{cases} \vec{u}_{r+1}, \dots, \vec{u}_m \in N(AA^T) = N(A^T) & \left(\text{proved in previous lectures}\right) \\ \underbrace{\frac{A\vec{v}_1}{\sigma_1}}_{\vec{u}_1}, \dots, \underbrace{\frac{A\vec{v}_r}{\sigma_r}}_{\vec{u}_r} \in C(A) & \end{cases}$$

[*) Recall that $A\vec{x}$ is $x_1\vec{a}_1 + x_2\vec{a}_2 + \dots + x_n\vec{a}_n$, i.e., lin. comb. of the cols. of A.]

So, $\left.\begin{array}{l} \vec{u}_{r+1}, \dots, \vec{u}_m \in N(A^T) \\ \vec{u}_1, \dots, \vec{u}_r \in C(A) \end{array}\right\}$ and $C(A) = N(A^T)^{\perp}$,

so all the $\vec{u}_1, \dots, \vec{u}_r$ are perpendicular to all $\vec{u}_{r+1}, \dots, \vec{u}_m$.

- Finally, we have to check that $A = U \Sigma V^T$ holds:

V orthogonal $\Rightarrow V^T = V^{-1} \Rightarrow \{ A = U\Sigma V^T \iff AV = U\Sigma \}$.

So let's check $AV = U\Sigma$:

$$A\left[\vec{v}_1 \dots \vec{v}_r \,\middle|\, \underbrace{\vec{v}_{r+1} \dots \vec{v}_n}_{\substack{\in N(A^TA) \\ = N(A)}}\right] = \left[\vec{u}_1 \dots \vec{u}_r \,\middle|\, \vec{u}_{r+1} \dots \vec{u}_m\right] \left[\begin{array}{ccc|c} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ \hline & & & \underbrace{\quad}_{n-r} \end{array}\right] \}\, m-r$$

(proved in previous lecture, and was in the exam)

$$\Downarrow$$

$$\left.\begin{aligned}
&\rightarrow A \left[\begin{array}{ccc|ccc} \vec{v}_1 & \cdots & \vec{v}_r & \vec{v}_{r+1} & \cdots & \vec{v}_n \end{array}\right] = \left[\begin{array}{ccc|c} A\vec{v}_1 & \cdots & A\vec{v}_r & \underbrace{\vec{0} \cdots \vec{0}}_{n-r} \end{array}\right] \\
&\rightarrow \left[\begin{array}{ccc|ccc} \vec{u}_1 & \cdots & \vec{u}_r & \vec{u}_{r+1} & \cdots & \vec{u}_m \end{array}\right] \left[\begin{array}{ccc|c} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ \hline & & & \underbrace{\quad\quad}_{n-r} \end{array}\right] = \left[\begin{array}{ccc|c} \sigma_1\vec{u}_1 & \cdots & \sigma_r\vec{u}_r & \underbrace{\vec{0} \cdots \vec{0}}_{n-r} \end{array}\right]
\end{aligned}\right\}$$

Both things are equal if $\sigma_i \vec{u}_i = A\vec{v}_i$ for $i = 1, \dots, r$, but this is exactly our choice for the $\vec{u}$'s //.

• Some remarks

1) $\{\vec{v}_1, \dots, \vec{v}_r, \vec{v}_{r+1}, \dots, \vec{v}_n\}$ are eigenvectors of A^TA corresponding to eigenvalues $\sigma_1^2, \dots, \sigma_r^2$ and 0 ($n-r$ times).

Indeed:

$$A^TA = V\Sigma^T U^T U \Sigma V^T = V \underbrace{\Sigma^T \Sigma}_{\substack{\text{diagonal} \\ n \times n}} V^T$$

$$\text{diagonal}_{n\times n} \rightarrow \begin{bmatrix} \sigma_1^2 & & & & \\ & \ddots & & & \\ & & \sigma_r^2 & & \\ & & & 0 & \\ & & & & \ddots \; 0 \end{bmatrix} \Big\}{n-r}, \quad \underbrace{\quad}_{n-r}$$

This is nothing more than the spectral decomposition of A^TA

(V plays the role of the Q)

2) Similarly, $\{\vec{u}_1, \ldots, \vec{u}_r, \vec{u}_{r+1}, \ldots, \vec{u}_m\}$ is a basis of eigenvectors of AA^T corresponding to eigenvalues $\sigma_1^2, \ldots, \sigma_r^2$ and 0 ($m-r$ times)

$$\hookrightarrow AA^T = \mathcal{U} \Sigma V^T V \Sigma^T \mathcal{U}^T = \mathcal{U} \underbrace{\Sigma \Sigma^T}_{} \mathcal{U}^T$$

Spectral decomposition of AA^T

diagonal $m \times m$

$$\begin{bmatrix} \sigma_1^2 & & & & & \\ & \ddots & & & & \\ & & \sigma_r^2 & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix} \}\, m-r$$

$$\underbrace{\hphantom{000}}_{m-r}$$

3) We have a complete analogy with the spectral decomposition of positive semidefinite matrices:

$$A = \mathcal{U} \Sigma V^T = \underbrace{\sigma_1 \vec{u}_1 \vec{v}_1^T}_{} + \ldots + \underbrace{\sigma_r \vec{u}_r \vec{v}_r^T}_{}$$

rank 1 matrices

4)

$$V = \left[\begin{array}{c|c} \underbrace{\vec{v}_1 \ldots \vec{v}_r}_{\substack{\text{Basis for} \\ \text{row space} \\ C(A^T)}} & \underbrace{\vec{v}_{r+1} \ldots \vec{v}_n}_{\substack{\text{Basis for} \\ \text{nullspace} \\ N(A)}} \end{array}\right] \qquad \mathcal{U} = \left[\begin{array}{c|c} \underbrace{\vec{u}_1 \ldots \vec{u}_r}_{\substack{\text{Basis for} \\ \text{column space} \\ C(A)}} & \underbrace{\vec{u}_{r+1} \ldots \vec{u}_m}_{\substack{\text{Basis for} \\ \text{left-nullspace} \\ N(A^T)}} \end{array}\right]$$