Elements of linear algebra

SVD (singular value decomposition)

Any real $m \times n$ matrix $A_{m\times n}$ can be decomposed as $A_{m\times n} = \mathcal{U}_{m\times m} S_{m\times n} V^T_{n\times n}$,

where $\mathcal{U}^T_{m\times m}\,\mathcal{U}_{m\times m} = \mathbb{I}_m = \mathcal{U}_{m\times m}\,\mathcal{U}^T_{m\times m}$
(orthonormal columns)

Each column of $\mathcal{U}$ is a left singular vector;

$$V^T_{n\times n} V_{n\times n} = V_{n\times n} V^T_{n\times n} = \mathbb{I}_n$$

(orthonormal rows & columns)

Each column of V is a right singular vector

$S_{m\times n}$ contains $r = \min(m,n)$ singular values $\sigma_i \geq 0$ on the main diagonal, $\emptyset$ everywhere else:

$$S_{m\times n} = \left.\overbrace{\begin{pmatrix} \sigma_1 & & 0 \\ & \sigma_2 \ddots & \\ 0 & & \sigma_r \\ \hline & 0 & \end{pmatrix}}^{n \text{ columns}}\right\} m \text{ rows}$$

$m > n$,
$r = n$ (# columns)

'tall, skinny' matrices

$$S_{m\times n} = \left(\begin{array}{cccc|c} \sigma_1 & & & 0 & \\ & \sigma_2 & & & 0 \\ 0 & & \ddots & & \\ & & & \sigma_r & \end{array}\right) \Bigg\} \begin{array}{l} m \\ \text{rows} \end{array}$$

$m < n$,
$r = m$ (# rows)

n columns

'short, fat' matrices

Cost of computing SVD decomposition:
$\theta(\min(mn^2, m^2n))$.

Now, consider $A^TA = (VS^TU^T)(USV^T) =$

$= V(S^TS)V^T$.

Then $(A^TA)V = V(S^TS)$

$\underset{n\times m}{S^T} S_{m\times n} \equiv D_n$ $n\times n$ diag. matrix containing σ_i^2

Likewise, $AA^T = (USV^T)(VS^TU^T) =$
$= U(SS^T)U^T$, s.t.

$(AA^T)U = U(SS^T)$

$\underset{m\times n}{S} S^T_{n\times m} \equiv D_m$ $m\times m$ diag. matrix containing σ_i^2

So, $\begin{cases} (A^TA)\,V = D_n V, \\ (AA^T)\,\mathcal{U} = D_m \mathcal{U}. \end{cases}$

V contains eigenvectors of A^TA as columns, while the diagonal of D_n contains the corresponding eigenvalues. Likewise, $\mathcal{U}$ & D_m play the same roles for AA^T.

Now, if A is square and symmetric and real $\rightarrow$ all eigenvalues real, all eigenvectors orthonormal

$$A = A^T, \quad n = m$$

Then $AA^T = A^TA = A^2 \Rightarrow V = \mathcal{U}$. $\quad \mathcal{U}^T = \mathcal{U}^{-1}$

$$\underset{\substack{\uparrow \\ \text{column of } \mathcal{U}}}{A^2 \vec{u}_i} = A(A\vec{u}_i) = \lambda_i^2 \vec{u}_i \Rightarrow \lambda_i^2 = \sigma_i^2$$

$$\Downarrow$$

$$\boxed{\lambda_i = \pm\sigma_i}$$

$$\text{So,} \quad A = \mathcal{U} \underset{\substack{\uparrow \\ n\times n \text{ diag. matrix with} \\ \lambda_i,\ i=1,\dots,n \\ \text{on the diagonal}}}{S} V^T = \mathcal{U} S \mathcal{U}^T = \underset{\substack{\uparrow \\ \text{eigenvalue decomposition} \\ \text{of } A}}{\underline{\underline{\mathcal{U} S \mathcal{U}^{-1}}}}.$$

Rank-nullity theorem

Recall that

$$A = \mathcal{U} S V^T = \sigma_1 \begin{pmatrix} \vec{u}_1 \end{pmatrix} \underline{\quad \vec{v}_1 \quad} + \dots + \sigma_r \begin{pmatrix} \vec{u}_r \end{pmatrix} \underline{\quad \vec{v}_r \quad}$$

Then

$$A\vec{x} = \sum_{j=1}^{r} \sigma_j (\vec{v}_j^T \cdot \vec{x}) \vec{u}_j$$

Thus, any $A\vec{x}$ can be written as a linear combination of r left singular vectors $\vec{u}_j$ $(j=1,\dots,r)$ $\Rightarrow$ $\text{rank}(A) = r$, the $\vec{u}_j$'s span the range of A

What is the basis for the null space?

Consider $\vec{y} = \sum_{j=r+1}^{n} c_j \vec{v}_j$ ← right singular vectors, columns of V

$$A\vec{y} = \sum_{j=1}^{r} \sigma_j \underbrace{(\vec{v}_j^T \cdot \vec{y})}_{=0 \text{ for } \forall j=1,\dots,r\text{, by construction}} \vec{u}_j = \vec{0}$$

Hence $\text{nullspace}(A) = \text{span}\{\vec{v}_j \text{ s.t. } \sigma_j = 0\}$ ← i.e., $j = r+1, \dots, n$

So, $\dim(\text{nullspace}(A)) = n - r$.

In summary, $\underbrace{\dim(\text{range}(A))}_{\text{i.e., rank}(A)} + \underbrace{\dim(\text{nullspace}(A))}_{\text{i.e., nullity}(A)} =$

$= \text{rank} + \text{nullity} = r + (n-r) = n$.

So, $\underbrace{\text{rank} + \text{nullity} = n}$

$\text{rank}(A) = \#\text{ non-zero } \sigma_i\text{'s}$

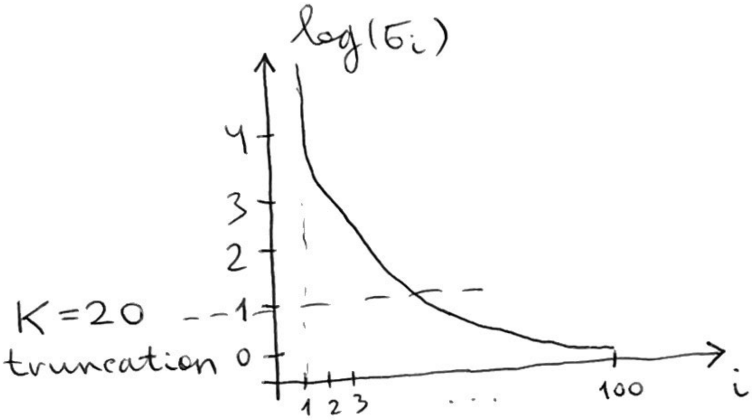
Truncated SVD

Instead of $A = U S V^T$, use $\hat{A}_K = U_K S_K V_K^T$,

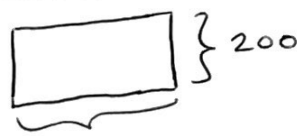
where $K \le r = \text{rank}(A)$.

U_K, V_K contain the first K columns of U, V, respectively.

S_K contains first K σ_i's (sorted by magnitude)



Original image: 200×320



$r = 200$

With $K=20$, the reconstructed 200×320 image has $r=20$, but still looks pretty good.

Moore-Penrose pseudo-inverse

MP pseudo-inverse of A : A^+

satisfies

$$\begin{cases} AA^+A = A, \\ A^+AA^+ = A^+ \end{cases} \qquad \begin{cases} (AA^+)^T = AA^+ \\ (A^+A)^T = A^+A \end{cases}$$

If A is square & non-singular (i.e., A^{-1} exists) : $A^+ = A^{-1}$

If $\underline{m > n}$, ('tall, skinny' A matrix) and
$r = \text{rank}(A) = n$ (full rank):

$$A^+_{n \times m} = (A^T A)^{-1} A^T .$$

In this case, $\underbrace{A^+ A}_{n \times n} = (A^T A)^{-1} A^T A = \mathbb{I}_n$, so that
A^+ is a left inverse of A.

However, $\underbrace{A A^+}_{m \times m} = A (A^T A)^{-1} A^T \neq \mathbb{I}_m$ in general

If $\underline{m < n}$ ('short, fat' A matrix) and
$r = \text{rank}(A) = m$ (i.e., $A^T_{n \times m}$ is full rank, its columns are linearly independent):

$$A^+_{n \times m} = A^T (A A^T)^{-1} .$$

In this case, $A A^+ = A A^T (A A^T)^{-1} = \mathbb{I}_m$.
A^+ is a right inverse but not a left inverse.

It turns out that if A is SVD - decomposed : $A = U S V^T$, then

$$A^+_{n \times m} = V_{n \times n} S^{-1}_{n \times m} U^T_{m \times m} ,$$ where

$$S^{-1}_{n \times m} = \underbrace{\begin{pmatrix} \ddots & & 0 \\ & \ddots & \\ 0 & & \end{pmatrix}}_{m \text{ columns}} \Big\} \; n \text{ rows}$$

$m < n$ as an example

$r = m$ σ_i^{-1}'s on the diagonal

What about $\begin{cases} (AA^+)^T = AA^+ \\ (A^+A)^T = A^+A \end{cases}$?

If $m>n$, $A^+A = \mathbb{I}_n$, so that $(A^+A)^T = A^+A$ trivially.

For $(AA^+)^T$, recall

$$\begin{cases} (AB)^T = B^T A^T, \\ (AB)^{-1} = B^{-1}A^{-1}, \\ (A^T)^{-1} = (A^{-1})^T. \end{cases}$$

Then $(A\overbrace{(A^TA)^{-1}A^T}^{A^+})^T = A(\underbrace{(A^TA)^{-1}}_{\substack{\shortparallel \\ A^{-1}(A^{-1})^T}})^T A^T =$

$= A\underbrace{(A^TA)^{-1}A^T}_{A^+} = AA^+$, as claimed

$A^{-1}(A^{-1})^T$, yielding $(A^{-1}(A^{-1})^T)^T = A^{-1}(A^T)^{-1} = (A^TA)^{-1}$

The $m<n$ case can be treated similarly.