Singular Value Decomposition

Saturday, June 13, 2020 18:14

Compare with $S = Q \Lambda Q^T$

But now we do $A = U \Sigma V^T$

left singular vectors

right singular vectors

general matrix

$$\Sigma = \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \end{bmatrix}$$

diagonal matrix, $\sigma_i \geq 0$ called singular values.

Any matrix can be factored as $A = U \Sigma V^T$

The key to the math here: $A^T A$ is nice.

Let A be $m \times n$

So, $A^T A = [n \times m][m \times n] = [n \times n]$

symmetric matrix and PSD

every ≥ 0

$A^T A = V \Lambda V^T$

diagonal eigenvalues, ≥ 0

eigenvectors of $A^T A$ (orthogonal)

Consider AA^T.

This is $m \times m$, but has the same non-0 eigenvalues as $A^T A$.

$\sigma_1^2, \sigma_2^2, \ldots \sigma_r^2$

So, $AA^T = U \Lambda U^T$

The SVD will use the U and V

We're looking for

$$A v_1 = \sigma_1 u_1$$
$$A v_2 = \sigma_2 u_2$$
$$\vdots$$
$$A v_r = \sigma_r u_r$$

We could have more vectors but they would be in the null space.

$\because$ A may not be square, it may not have evectors x s.t. $Ax = \lambda x$ but it must have some v and u s.t. $Av = \sigma u$

So, we're looking for a bunch of orthogonal vectors v_i s.t.

Av_i gives us a bunch of orthogonal vectors u_i

So, we're looking for a bunch of orthogonal vectors v_i s.t. Av_i gives us a bunch of orthogonal vectors u_i

So, $A \begin{bmatrix} v_1 & v_2 & \cdots & v_r \end{bmatrix} = \begin{bmatrix} u_1 & u_2 & \cdots & u_r \end{bmatrix} \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \end{bmatrix}$

$$AV = U\Sigma$$

$$AVV^T = U\Sigma V^T \qquad (VV^T = I)$$

$$\underline{A = U\Sigma V^T}$$

↳ If this is what we hope, what will A^TA, AA^T look like?

$$A^TA = (U\Sigma V^T)^T (U\Sigma V^T) = (\Sigma V^T)^T \cancel{U^T U} \Sigma V^T$$

$$= V(\Sigma^T \Sigma)V^T$$

↳ eigenvectors of $A^TA = V$

and σ_i^2 are eigenvalues of A^TA and also of AA^T

$$AA^T = U\Sigma \cancel{V^T V} \Sigma^T U^T = U\Sigma\Sigma^T U^T$$

↳ eigenvectors of AA^T

→ Now let's see what the matrix $A = U\Sigma V^T$ does to a vector, as a transformation. Consider Ax

1. First, as V^T is orthogonal, length doesn't change for x, it may just be rotated or reflected.
2. Consider $\Sigma V^T x$ now: this stretches the vectors.
3. U is another orthogonal matrix, so another rotation/reflection

SVD essentially tells us that every linear transformation (matrix multiplication) factors into a rotation × a stretch × a possibly different rotation.

If A is PSD and symmetric, $A = Q\Lambda Q^T$

→ To actually compute the SVD, we don't use A^TA or AA^T bcoz it's very expensive.

To associate with geometric interpretation:

→ If A is 2×2 $= \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, we have 4 numbers a, b, c, d.

And we have $U \leftarrow$ 1 angle responsible for a 2D rotation +
$\Sigma \leftarrow$ 2 numbers for 2D vector stretch +
$V^T \leftarrow$ 1 angle for 2D rotation

$1+2+1 = 4$

→ If A is 3×3, we have 9 numbers:
3 for 3D rotation (roll, pitch, yaw) +
3 for stretching +
3 for another rotation.

Basically for $n \times n$ (n-dimensions), we need $\binom{n}{2}$ numbers for rotation — angles b/w all pairs of axes; and n numbers for stretching. So, total:

$$2\binom{n}{2} + n = \underline{n^2}.$$

↳ no. of matrix entries!

⇒ Just like the product of eigenvalues is the determinant, for a sq. matrix $A = U\Sigma V^T$, the product of singular values is the determinant ($\det U = \det V^T = 1$ as U, V orthogonal).

So,

$$A = \overset{m\times r}{\begin{bmatrix} & & \\ u_1 & \cdots & u_r \\ & & \end{bmatrix}} \overset{r\times r}{\begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix}} \overset{r\times n}{\begin{bmatrix} v_1^T \\ \vdots \\ v_r^T \end{bmatrix}}$$

$m\times m$ $m\times n$ $n\times n$

$$= \overset{m\times m}{\begin{bmatrix} & & \\ u_1 & \cdots & u_m \\ & & \end{bmatrix}} \overset{m\times n}{\begin{bmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ & & & & \ddots \; 0 \end{bmatrix}} \overset{n\times n}{\begin{bmatrix} v_1^T \\ \vdots \\ v_n^T \end{bmatrix}}$$

The extra u's are in the null-space of AA^T. Similar for the extra v's.

Accounting for null-space stuff.

Polar decomposition

Analogous to the polar form of a complex number:

$$re^{i\theta}$$

r → S: symmetric

$e^{i\theta}$ → Q: orthogonal

$$\Rightarrow \quad A = SQ \quad (\text{for every matrix } A)$$

Let's see this from the SVD.

Construct a symmetric matrix: $U\Sigma U^T$

Write

$$A = \underbrace{(U\Sigma U^T)}_{S}\,\underbrace{(UV^T)}_{Q}$$

$= I$

The polar decomp is used in engineering a lot.

Key fact:

If we have a big matrix of data, A and want to pull out the important parts (what data science has to do).

The biggest rank 1 part of a matrix $A = U \Sigma V^T$ is: $\boxed{u_1 \sigma_1 v_1^T}$

Here we see A as a sum of rank 1 matrices:

$$A = \sum_{i=1}^{r} u_i \sigma_i v_i^T$$

. Here r is the rank.