18.06

Zack Remscrim

Matrix Factorization

Given a matrix A can we write it as a product

$A = QR$

or $A = LU$

or $A = U\Sigma V^\top$

or…

Where the matrices on the right hand side satisfy certain particular properties (rotation matrix, lower triangular matrix, diagonal matrix, …)

Before discussing if we can do this, why would we want to do this?

Polynomial Factorization

Given a polynomial $f(x) = x^5 - 3x^4 - 5x^3 + 5x^2 - 6x + 8$

Want to find roots of f (values x such that $f(x) = 0$)

f is complicated, so this is not so easy

Idea: Factor f

$f(x) = (x-1)(x+2)(x-4)(x^2+1)$

Easy to see roots are $1, -2, 4, i, -i$

Polynomial Factorization

Why does this work?

Every polynomial (over $\mathbb{R}$) can be factored into a product of linear and quadratic polynomials (Fundamental theorem of algebra)

It is easy to find the roots of linear and quadratic polynomials

It is easy to determine the roots of a polynomial from the roots of its factors

Matrix Factorization

Given some matrix, A we want to “understand” A as a function

A is complicated, so this is not so easy

Idea: Factor A

For example, SVD, $A = U\Sigma V^\top$

Easy to understand A, it is a “generalized rotation” by $V^\top$, then a scaling by Σ, then a “generalized rotation” by U

Matrix Factorization

Why does this work?

Every matrix (over $\mathbb{R}$) can be factored into a product $A = U\Sigma V^\top$
(SVD theorem)

It is easy to understand the matrices $U, \Sigma, V^\top$

It is easy to understand a matrix given an understanding of its factors

Singular Value Decomposition

"Full" SVD: Given $m \times n$ matrix A

Write $A = U\Sigma V^\top$

U is $m \times m$ orthogonal matrix ($U^\top U = I_m$)

V is $n \times n$ orthogonal matrix ($V^\top V = I_n$)

Σ is $m \times n$ diagonal matrix

Diagonal entries: $\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \cdots \geq \sigma_{\min(m,n)} \geq 0$

The σ_i all called singular values

Computing the SVD

Warmup:

Write $5 = apb$ with p prime and a, b prime or 1

Let $D = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, what is an SVD of D?

$D = IDI$

Let Q be square orthogonal, what is an SVD of Q?

$Q = QII$

or $Q = IIQ$

SVD not unique, but close

Computing the SVD

How do we compute a the SVD in general?

Hard to compute by hand

Easy for a computer

Procedures good for hand calculation are different than those a computer should use

We will not discuss this

However, can sometimes compute SVD via "advanced guessing"

Using properties of U, V, Σ

Singular Value Decomposition

Block form: Say A has r positive singular values

Write $A = [U_1 \quad U_2] \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^\top \\ V_2^\top \end{bmatrix}$

Σ_r $r \times r$ diagonal matrix, with all positive singular values on diagonal

U_1 first r columns of U

V_1 first r columns of V

U_1, U_2, V_1, V_2 are all orthogonal matrices

Singular Value Decomposition

$$
\begin{aligned}
A &= \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^\top \\ V_2^\top \end{bmatrix} \\
&= \begin{bmatrix} U_1 & U_2 \end{bmatrix} \left(\begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^\top \\ V_2^\top \end{bmatrix} \right) \\
&= \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma_r V_1^\top \\ 0 \end{bmatrix} \\
&= U_1 \Sigma_r V_1^\top \ \text{(rank-}r\text{ SVD)}
\end{aligned}
$$

Why is this useful?

Column Space

For an $m \times n$ matrix A

Let $A_1, A_2, \dots, A_n$ denote columns of A

Column space col(A)= $\{c_1A_1 + c_2A_2 + \cdots + c_nA_n | c_i \in \mathbb{R}\}$

(all linear combinations of columns)

$= \{Ax | x \in \mathbb{R}^n\}$

(all $y \in \mathbb{R}^m$, such that $y = Ax$, for some x)

Equation $Ax = b$ has a solution exactly when $b \in$ col(A)

Is a vector space

Column Space

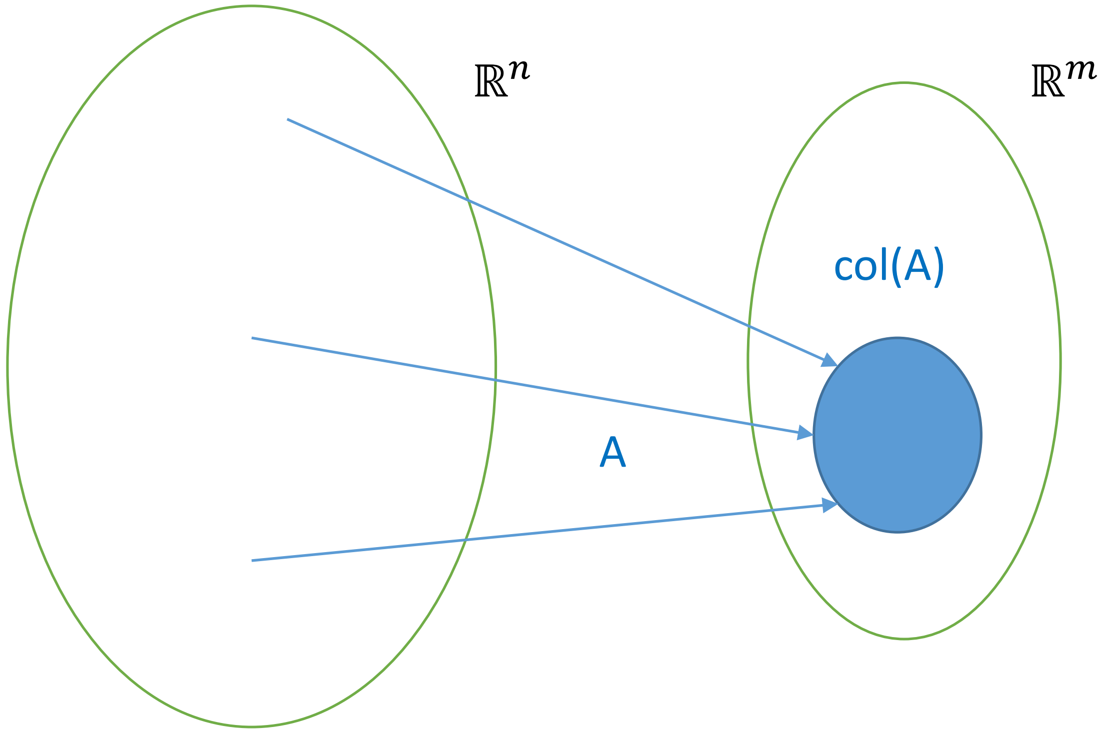
For an $m \times n$ matrix A

A is a function

input: $x \in \mathbb{R}^n$

output: $y \in \mathbb{R}^m$

$y = Ax$



Column Space

Claim: If $A = U_1 \Sigma_r V_1^\top$ (rank-r SVD),

Then $\text{col}(A) = \text{col}(U_1)$

Why?

If $b \in \text{col}(A)$

Then $b = Ax$ for some x

$$\begin{aligned} b &= Ax_1 \\ &= U_1\ \Sigma_r V_1^\top x \\ &= U_1\ (\Sigma_r V_1^\top x) \end{aligned}$$

Then $b \in \text{col}(U_1)$

So $\text{col}(A) \subseteq \text{col}(U_1)$

Column Space

Claim: If $A = U_1 \Sigma_r V_1^\top$ (rank-r SVD),

Then $\text{col}(A) = \text{col}(U_1)$

Why?

If $b \in \text{col}(U_1)$

Then $b = U_1 x$ for some x

$$
\begin{aligned}
b &= U_1 x_1 \\
&= A V_1 \Sigma_r^{-1} x \\
&= A(V_1 \Sigma_r^{-1} x)
\end{aligned}
$$

Then $b \in \text{col}(A)$

So $\text{col}(U_1) \subseteq \text{col}(A)$

Column Space

Claim: If $A = U_1 \Sigma_r V_1^\top$ (rank-r SVD),

Then $\text{col}(A) = \text{col}(U_1)$

Why is this useful?

U_1 is orthogonal matrix, much easier to write down its column space