Lecture Note:
Singular Value Decomposition

Marc Toussaint

Learning & Intelligent Systems Lab, TU Berlin

August 23, 2024

One can view a matrix $A \in \mathbb{R}^{3\times 4}$ as a collection of rows, $A = \begin{pmatrix} v_1^\top \\ v_2^\top \\ v_3^\top \end{pmatrix}$, $v_i \in \mathbb{R}^4$. Applying A on x is then scalar-producting all rows with x and outputs $Ax = \begin{pmatrix} v_1^\top x \\ v_2^\top x \\ v_3^\top x \end{pmatrix}$. The rows span an input space $I = \text{span}\, v_{1:3}$. All x that are orthogonal to I will be mapped to zero. The rows only pick up components of x that lie within I.

Or one can view a matrix A as a collection of columns, $A = \begin{pmatrix} u_1 & u_2 & u_3 & u_4 \end{pmatrix}$, $u_i \in \mathbb{R}^3$. Applying A on x then gives the linear combination $Ax = x_1u_1 + .. + x_4u_4$ of these columns, with x_i being the linear coefficients. All outputs $y = Ax$ will lie in the output space $O = \text{span}\, u_{1:4}$.

This view of matrices as input space spanning rows, or output space spanning columns, is useful and clarifies that matricies transport from some input space to some output space. But given a matrix A in raw form we don't really have an explicit understanding of that transport: Regarding the input space, some rows might be linearly dependent, so that the input dimension could be less than n. And the rows may not be orthonormal, so we do not have an explicit orthonormal basis describing the input space. The same two points hold for the output space (columns being linearly dependent and not orthonormal).

The SVD rewrites a matrix in a form where we really have an orthonormal basis for the input and output spaces, and a clear understanding which input directions are mapped to which output directions. Here the theorem:

For any matrix $A \in \mathbb{R}^{m\times n}$ there exists a $k \leq m, n$, orthonormal vectors $v_1, .., v_k \in \mathbb{R}^n$, orthonormal vectors $u_1, .., u_k \in \mathbb{R}^m$, and scalar numbers $\sigma_k > 0$, such that

$$A = \sum_{i=1}^{k} u_i \sigma_i v_i^\top = USV^\top\,, \quad \text{where} \quad S = \text{diag}(\sigma_{1:k}),\ U = u_{1:k} \in \mathbb{R}^{m\times k},\ V = v_{1:k} \in \mathbb{R}^{n\times k}\ . \tag{1}$$

In this form, we see that $V^\top$ spans the input space with orthonormal rows $v_i^\top$, and U spans the output space with orthonormal columns u_i. Further, we understand what's happening "in between": Each component $u_i\sigma_i v_i^\top$ first projects x on the ith input direction v_i, then scales this with the factor σ_i, then out-projects it to the output direction u_i. This is done "independently" for all $i = 1, .., k$, as all v_i and u_i are orthogonal. In short, what the matrix does it: it transports each input direction v_i to the output direction v_i and scales by σ_i in between. The number k tells us how many dimensions are actually transported (could be less than m and n).

k is called the rank of the matrix (note that we required $\sigma_i > 0$) and σ_i are called the singular values.

The matrices U and V are orthonormal and in some explanations characterized as rotations, and the

equation $A = USV^\top$ described as rotation-scaling-rotation. That's ok, but this story does not work well if $m \neq n$ (we have different input and output spaces), or $k < m, n$ (we don't have full rank). I think the above story is better.

Matrices of the form $xy^\top$ (which is also called outer product of x and y) are of rank 1 (the singular value would be $\sigma_1 = |x||y|$, and $u = x/|x|, v = y/|y|$). One can think of rank 1 matrices as minimalistic matrices: they pick up a single input direction, scale, and out-project to a single output direction. The sum notation $A = \sum_{i=1}^{k} \sigma_i u_i v_i^\top$ describes A as a sum of rank 1 matrices, i.e., every matrix A can be thought of as a composition of rank 1 matrices. This clarifies in what sense rank 1 matrices are minimalistic building blocks of higher rank matrices.