SVD Sparsity

Michael Wilensky

December 2023

1 Introduction

The point of this document is to expose why singular value decomposition (SVD) tends to produce such sparse respresentations of matrices (other than that, obviously, you have reduced NM numbers to $\min(N, M)$ numbers. The key turns out that singular vectors of a matrix $\mathbf{M}$ solve a least-squares objective function.

2 Approximating a Matrix

Suppose we have a unit-length vector, $\mathbf{v}$, i.e $\mathbf{v}^\dagger\mathbf{v} = 1$, and we would like to know which unit-length vector $\mathbf{u}$ best approximates a matrix, $\mathbf{M}$, like so:

$$\mathbf{M} \approx \sigma\mathbf{u}\mathbf{v}^\dagger \tag{1}$$

for some value, σ. Best is of course a user-defined concept, so let's just use least-squares for convenience. Since $\mathbf{M}$ is a matrix, it looks a little funny, but here it is

$$L = \mathrm{tr}\left((\mathbf{M} - \sigma\mathbf{u}\mathbf{v}^\dagger)(\mathbf{M}^\dagger - \sigma^*\mathbf{v}\mathbf{u}^\dagger)\right) \tag{2}$$

It turns out this is satisfied when

$$\sigma\mathbf{u} = \mathbf{M}\mathbf{v} \tag{3}$$

$$|\sigma|^2 = \mathbf{v}^\dagger\mathbf{M}^\dagger\mathbf{M}\mathbf{v} \tag{4}$$

This second equation just ensures $\mathbf{u}$ is unit-norm. If $\mathbf{v}$ transforms to a large vector under $\mathbf{M}$, then σ will be large in amplitude, and the approximation will be tight compared to a vector that does not transform to as large a vector as $\mathbf{v}$. Note that we could flip this around and apply a similar derivation to $\mathbf{v}$ given $\mathbf{u}$.

3 What does this have to do with SVD?

It turns out that if two vectors $\mathbf{u}$ and $\mathbf{v}$ satisfy $\sigma\mathbf{u} = \mathbf{M}\mathbf{v}$ and $\sigma\mathbf{v} = \mathbf{M}^\dagger\mathbf{u}$, then they are left- and right-singular vectors of $\mathbf{M}$, respectively. So given the above derivation, it is no surprise that SVD works so well to approximate matrices. It is a successive pairing of vectors that produce finer and finer approximations of $\mathbf{M}$ when their outer products are summed (assuming singular values have been sorted largest to smallest).

An important fact about SVD that is not apparent in the derivation of a single pair of vectors $\mathbf{u}$ and $\mathbf{v}$ above is that you can always find a set of orthogonal singular vectors (i.e. a unitary matrix $\mathbf{U}$ whose columns are the left-singular vectors and similar for the right-singular ones). This results in something exceptionally sparse since if you take any vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ and produce $\mathbf{u}_1 = \mathbf{M}\mathbf{v}_1$ and $\mathbf{u}_2 = \mathbf{M}\mathbf{v}_2$, you are not guaranteed that $\mathbf{u}_1$ and $\mathbf{u}_2$ are orthogonal to one another. If $\mathbf{u}_1$ and $\mathbf{u}_2$ are not orthogonal (and similar for the $\mathbf{v}$'s), then it is not clear that each successive pairing makes the approximation better and better. The fact that an SVD like this always exists is a great boon.