CS 3220: The Singular Value Decomposition

Instructor: Anil Damle

Disclaimer

These notes are intended to highlight key points from some of the lectures. However, they are not written nor intended to be substitutes for the lectures. Furthermore, the resources section of the website contains extensive written material covering the topics in class, material that was developed with the explicit intention of being a written presentation of this material. Many of those books are several editions into their existence and, therefore, have been refined in a way these notes are not. In addition, the textbooks given are invariably far more comprehensive and thorough in their treatment of the topics. Lastly, I take full responsibility for any typos included herein; nevertheless, these notes are a work in progress and if you find anything amiss please let me know.

The Singular value decomposition

For the purposes of this course there are several aspects of the so-called Singular Value Decomposition (SVD) of a matrix that we will be interested in—these notes primarily pertain to algebraic properties of the SVD. We start with a definition:

Definition 1 (The full SVD)**.** *Given $A \in \mathbb{R}^{m\times n}$ there exist orthogonal matrices $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$ with columns $\{u_i\}_{i=1}^m$ and $\{v_i\}_{i=1}^m$ along with a "diagonal matrix" $\Sigma \in \mathbb{R}^{m\times n}$ (in the sense that $\Sigma_{i,j} = 0$ if $i \neq j$) with diagonal entries $\sigma_i \equiv \Sigma_{i,i}$ satisfying $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\min(m,n)} \geq 0$ such that*

$$A = U\Sigma V^T.$$

We call the columns of U the left singular vectors of A and the columns of V the right singular vectors of A. The σ_i's are the singular values of A. Collectively, the matrices U, Σ, and V are referred to as the SVD of A.

While not quite unique, the factorization is essentially unique and all of the properties we will discuss are valid regardless of the specific SVD given/found; therefore, we omit any further consideration of this point.

Connections to fundamental subspaces

In class, we defined the following four subspaces pertaining to any matrix $A \in \mathbb{R}^{m\times n}$:

Range: The range of a matrix (synonymously, the column space) is defined as

$$\{z \in \mathbb{R}^m | z = Ax \text{ for some } x \in \mathbb{R}^n\}.$$

Colloquially, this is the set of vectors that one can produce by taking arbitrary linear combinations of the columns of A.

Co-Range: The co-range of a matrix (synonymously, the row space) is defined as

$$\{z \in \mathbb{R}^n | z = A^T x \text{ for some } x \in \mathbb{R}^m\}.$$

Colloquially, this is the set of vectors that one can produce by taking arbitrary linear combinations of the rows of A. This is equivalent to the range of A^T.

Kernel: The kernel (synonymously, the null space) of A is defined as

$$\{x \in \mathbb{R}^n | Ax = 0\}.$$

The kernel of a matrix is the set of vectors mapped to zero. Note that 0 is always in the kernel and we call the kernel of a matrix A non-trivial if there exist non-zero vectors in it (*i.e.,* the dimension of the null space is greater than zero).

Co-Kernel: Quite simply, the kernel of A^T. More formally this is defined as

$$\{x \in \mathbb{R}^m | A^T x = 0\}.$$

Here, our first consideration related to these subspaces is that the SVD reveals an orthonormal basis for each of these spaces. In particular, let $r \leq \min(m, n)$ be the number of non-zero singular values. Then, we have that:

- $\text{range}(A) = \text{span}\{u_1, \ldots, u_r\}$
- $\text{null}(A) = \text{span}\{v_{r+1}, \ldots, v_n\}$ (If $r = n$ the null space of A is just the zero vector.)
- $\text{range}(A^T) = \text{span}\{v_1, \ldots, v_r\}$
- $\text{null}(A^T) = \text{span}\{u_{r+1}, \ldots, u_n\}$ (If $r = m$ the null space of A^T is just the zero vector.)

Several standard observations stem from these facts. First, we immediately see that $\text{range}(A) \perp \text{null}(A^T)$ and $\text{range}(A^T) \perp \text{null}(A)$ because U and V are orthogonal matrices. Second, we have that $\text{Dim}(\text{range}(A)) = r$, $\text{Dim}(\text{null}(A)) = n - r$, $\text{Dim}(\text{range}(A^T)) = r$, and $\text{Dim}(\text{null}(A^T)) = m - r$. This means that the range of A and null space of A^T decompose $\mathbb{R}^m$ into two orthogonal subspaces that jointly span all of $\mathbb{R}^m$. Similarly, the null space of A and range of A^T decompose $\mathbb{R}^n$ into two orthogonal subspaces that jointly span all of $\mathbb{R}^n$. All of the preceding may be inferred from examination of the SVD.

The rank of a matrix

Keeping with the theme of framing everything in terms of the SVD, we define the rank of a matrix in the following manner.

Definition 2 (Rank of a matrix)**.** *The rank of a matrix $A \in \mathbb{R}^{m \times n}$ is the number of non-zero singular values of A.*

Observe that by definition, we have that $\text{rank}(A) \leq \min(m, n)$. In addition, we also have that the rank of a matrix is the dimension of the column space (and row space). (This also motivates the use of a subscript r in the previous section.) When $m = n$ we say a matrix is full rank if $r = n$. For $m > n$ we say a matrix is full column rank if $r = n$ and if $m < n$ we say a matrix is full row rank if $r = m$. (Though, sometimes for rectangular matrices we may omit "row" or "column" and simply refer to them as full rank if $r = \min(m, n)$.)

The SVD as a sum of rank one matrices

An alternative view of the SVD is that it yields a representation of A as a sum of rank-one matrices. In particular, the SVD yields

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

Best approximation via the SVD

Building off of the preceding interpretation of the SVD, it is interesting to think about what happens if we truncate the sum. In particular, let

$$A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^T$$

for $k \leq r$. Clearly this matrix is related to A in some way. It turns out that it solves a very specific approximation problem, that of approximating A by a lower rank matrix. While we do not dwell on why low-rank approximations are useful here, one example may be found in the SVD demo code available on the course website.

Here, we discuss the so-called Eckart-Young-Mirsky theorem. This Theorem tells us that A_k is the best approximation of A by a rank k matrix, in fact it is so in two different norms.

Theorem 1 (Eckart-Young-Mirsky, informally)**.** *Let* $A \in \mathbb{R}^{m\times n}$ *have SVD* $A = U\Sigma V^T$ *and define* $A_k = A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^T$. A_k *solves the optimization problem*

$$\min_{B \text{ of rank } k} \|A - B\|_*$$

for $* = \{2, F\}$. *Furthermore*

$$\|A - A_k\|_2 = \sigma_{k+1} \quad \textit{and} \quad \|A - A_k\|_F = \left(\sum_{i=k+1}^{\min{(m,n)}} \sigma_i^2 \right)^{1/2}.$$