**CSE 840: Computational Foundations of Artificial Intelligence** October 2, 2023

Lecture 9

*Instructor: Vishnu Boddeti* *Scribe: Ayaan Shaik, Javen Zamojcin, Max Brombach*

# 1 Singular Value Decomposition

**Proposition 1** *Consider $A \in \mathbb{R}^{mxn}$ of rank $r$. Then we can write $A$ in the form*

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

*where $U \in \mathbb{R}^{mxm}$, $V \in \mathbb{R}^{nxn}$ are orthogonal matrices and $\Sigma \in \mathbb{R}^{mxn}$ is diagonal and exactly $r$ of the diagonal values $\sigma_1, \sigma_2, ...$ are non-zero.*

**Definition 2** *Singular Value Decomposition (SVD) is a matrix factorization technique that decomposes a matrix into three simpler matrices, revealing the underlying structure and relationship with the orthogonal matrix.*

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

*Where $U$ is a left orthogonal (m x m) matrix and m is the number of rows in A, $V^T$ is the transpose of a right orthogonal (n x n) matrix and n is the number of columns in A, and $\Sigma$ is a (m x n) diagonal matrix with non-negative real numbers on its diagonal, called singular values.*

**Proof of Proposition 1:** Consider a matrix $A \in \mathbb{R}^{mxn}$ with rank $r$. Now consider a matrix $B := A^TA \in \mathbb{R}^{mxn}$, which is therefore symmetric and positive semi-definite (PSD).

$$Symmetric : (A^TA)^T = A^T(A^T)^T = A^TA$$

$$PSD : x^TBx = \ < x, Bx > \ = \ < x, A^TAx > \ = \ < Ax, Ax > \ = ||Ax||^2 \geq 0$$

Therefore, there exists an orthonormal basis of eigenvectors $x_1, x_2, ..., x_n$ with eigenvalues $\lambda_1, \lambda_2, ..., \lambda_n \geq 0$.
Let $\Sigma \in \mathbb{R}^{mxn}$ be the diagonal matrix of singular values, $\sigma_i$, where $\sigma_i = \sqrt{\lambda_i}$. We know we can take the square root of the eigenvalues because PSD matrices are equivalently characterized as matrices with non-negative eigenvalues.
We can now construct a unit eigenvector $r_i$ of matrix $B$ as $r_i := \frac{(Ax_i)}{\sigma_i}$. Now define matrices $U$ and $V$ with columns of $r_i$, and $x_i$, respectively. Columns of $U\Sigma$ are given as $\sigma_i r_i = \sigma_i \frac{Ax_i}{\sigma_i} = Ax_i$.
Multiply the columns of $U\Sigma$ by $V^T$. Consider that the rows of $V^T$ are the $x_i$ vectors; and if $i \neq j$, then $x_i \perp x_j$ and $||x_i|| = 1$. The terms consisting of $i, j$ with $x_i \perp x_j$ cancel, and the terms with $i = j$ will equal 1.
What remains afterward will be matrix $A$, therefore proving that $A = U\Sigma V^T$. $\square$

# 2 Key Differences Between SVD and Eigendecomposition

**Remark 3** *SVD always exists, no matter how matrix A looks like.*

**Remark 4** *Matrices $U, V$ are orthogonal, which is not true of eigenvectors in general.*

**Remark 5** *Singular values are always real and non-negative.*

**Remark 6** *If $A \in \mathbb{R}^{nxn}$ is symmetric, then the SVD is "nearly" the same as the eigenvalue decomposition. If $(\lambda_i, v_i)$ are the eigenvalue/eigenvector pairs of $A$, then $|\lambda_i|, v_i$ are the singular value / singular vector pairs of $A$. In particular, left-and-right singular vectors are the same.*

**Remark 7** *Left-singular vectors of $A$ are the eigenvectors of $AA^T$.*

**Remark 8** *Right-singular vectors of $A$ are the eigenvectors of $A^TA$.*

**Remark 9** *Right-singular vectors of $A$ are the eigenvectors of $A^TA$.*

**Remark 10** $\lambda_i \neq 0$ *is an eigenvalue of* $AA^T \vee A^TA \iff \sqrt{\lambda_i} \neq 0$ *is a singular value of* $A$.

# 3 Matrix Norms

Consider a matrix $A \in \mathbb{R}^{m \times n}$

**Definition 11**

$$||A||_{max} = ||A||_\infty = \max_{ij} |a_{ij}|$$

**Definition 12**

$$||A||_1 = \sum_{i,j} |a_{ij}|$$

**Definition 13** *Frobenius Norm*

$$\begin{aligned} ||A||_F &= \sqrt{\sum_{i,j} |a_{ij}^2|} = \sqrt{tr(A^TA)} \\ &= \sqrt{\sum \sigma_i^2} \quad \textit{where } \sigma_i \textit{ are the singular values of } A. \end{aligned}$$

**Definition 14** *Operator norm/Spectral Norm*

$$\begin{aligned} ||A||_2 &= \sigma_{max}(A) \quad where\ \sigma_{max} is the largest singular value. \\ &= \max_{x \neq 0} \frac{||Ax||}{||x||} \quad [\ \textit{where } ||Ax||, ||x|| \textit{ are Euclidean norm on vectors in } \mathbb{R}^m] \end{aligned}$$

# 4 Rank-k Approximations of Matrices

**Definition 15** *Consider a matrix $A = U\Sigma V^T withentries \sigma_1, \sigma_2, \ldots$ sorted in decreasing order. We define a new matrix $A_k$ as follows:*

$$A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^T \quad [\ u_i v_i \textit{is rank-1 matrix}\ ]$$

**Proposition 16** *Let $B$ be any rank-k matrix $\in \mathbb{R}^{m \times n}$. Then:*

$$||A - A_k||_F \leq ||A - B||_F$$

*$A_k$ is the best rank-k approximation (in Frobenius norm).*

**Proposition 17** *For any matrix $B$ of rank-k, $B \in \mathbb{R}^{m \times n}$,*

$$||A - A_k||_2 \leq ||A - B||_2$$

*where $|| \cdot ||_2$ denotes the operator norm. $A_k$ is the best rank-k approximation (in operator norm)*

# 5 Pseudo-Inverse of Matrix

**Definition 18** *For $A \in \mathbb{R}^{m \times n}$, a pseudo inverse of $A$ is defined as the matrix $A^\dagger \in \mathbb{R}^{m \times n}$ which satisfies the following properties:*

$$\left.\begin{array}{ll} (i) & AA^\dagger A = A \\ (ii) & A^\dagger A A^\dagger = A^\dagger \end{array}\right\} \text{"}nearly\ inverse\text{"}$$

$$\left.\begin{array}{ll} (iii) & (AA^\dagger)^T = AA^\dagger \\ (iv) & (A^\dagger A)^T = A^\dagger A \end{array}\right\} symmetry$$

**Intuition 19**

- *$A$ is a projection from $\mathbb{R}^3 \to \mathbb{R}^2$: $A \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x1 \\ x2 \end{pmatrix}$*
- *Cannot invert, obviously (inverting means reconstructing original)*
- *But I could "make up" a reconstruction: $R : \mathbb{R}^2 \to \mathbb{R}^3$, $R \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ 5 \end{pmatrix}$*
- *Now we have: $ARA = A \implies AA^\dagger A = A$*

**Proposition 20** *Let $A \in \mathbb{R}^{mxn}, A = U\Sigma V^T$ its SVD. Then: $A^\dagger = V\Sigma^\dagger U^T$ where $\Sigma^\dagger \in \mathbb{R}^{mxn}$.*

$$\Sigma_{ii}^\dagger = \begin{cases} \frac{1}{\Sigma_{ii}} & \textit{if } \Sigma_{ii} \neq 0 \\ 0 & \textit{otherwise} \end{cases}, \ \Sigma = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ 0 & & \sigma_n \end{pmatrix}, \ \Sigma = \begin{pmatrix} \frac{1}{\sigma_1} & & \\ & \ddots & \\ 0 & & \frac{1}{\sigma_n} \end{pmatrix}$$

**Intuition 21** *Assume $A \in \mathbb{R}^{nxn}$, invertible, has eigendecomposition $A = U\Lambda U^T$. Then:*

- *All entries of diag$(\Lambda)$ are $\neq 0$ (eigenvalues $\neq 0$)*
- $A^{-1} = U\Lambda^{-1}U^T$ *with* $\Lambda = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$, $\Lambda^{-1} = \begin{pmatrix} \frac{1}{\lambda_1} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\lambda_n} \end{pmatrix}$