Linear Algebra

Sartaj Ul Hasan

Department of Mathematics
Indian Institute of Technology Jammu
Jammu, India - 181221

Email: sartaj.hasan@iitjammu.ac.in

Lecture 41

(Nov 15, 2019)

Singular Value Decomposition (SVD)

Singular Value Decomposition

Observation: Let A be an $m \times n$ matrix – i.e A represents a linear transformation $X \mapsto AX$ from $\mathbb{R}^n$ to $\mathbb{R}^m$. Then A^TA, being a symmetric $n \times n$ matrix, can be orthogonally diagonalized. Let $\{v_1, v_2, \ldots, v_n\}$ be an orthonormal basis for $\mathbb{R}^n$ consisting of eigenvectors of A^TA with corresponding eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. Then:

$$\begin{aligned}
||Av_i||^2 &= < Av_i, Av_i > \\
&= (Av_i)^T(Av_i) = v_i^T(A^TA)v_i \\
&= v_i^T\lambda_i v_i \text{ (since } v_i \text{ is an eigenvector of } A^TA \text{ for } \lambda_i) \\
&= \lambda_i v_i^T v_i \\
&= \lambda_i < v_i, v_i >= \lambda_i ||v_i||^2 = \lambda_i
\end{aligned}$$

In other words, all the eigenvalues of the matrix A^TA are non-negative. Hence we can if necessary re-arrange the eigenvalues so that: $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$.

Singular Value Decomposition (Cont'd)

- **Definition:** Let A be an $m \times n$ matrix. The **singular values** of A are the square roots of the eigenvalues of $A^T A$, denoted by $\sigma_1, \sigma_2, \ldots, \sigma_n$ arranged in descending order, i.e., $\sigma_i = \sqrt{\lambda_i}$ for $i = 1, \ldots, n$.
- Note that the singular values are the lengths of the vectors $Av_1, Av_2, \ldots, Av_n$.
- **Proposition 58:** Suppose that $\{v_1, v_2, \ldots, v_n\}$ is an orthonormal basis for $\mathbb{R}^n$ consisting of eigenvectors of $A^T A$ with corresponding eigenvalues arranged so that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$. Suppose that A has r nonzero singular values. Then $Av_1, Av_2, \ldots, Av_r$ is an orthogonal basis for Col A, and rank $A = r$.

Proof of Proposition 58

Proof of Proposition 58: Clearly the vectors $Av_1, Av_2, \dots, Av_n$ belong to Col $A \subseteq \mathbb{R}^m$. Also, for $j > r$, we have $||Av_j|| = \sqrt{\lambda_j} = \sigma_j = 0$, so $Av_j = 0$. For $i, j \leq r$, we have:

$$\begin{aligned} < Av_i, Av_j > &= Av_i . Av_j = (Av_i)^T (Av_j) = v_i^T (A^T A) v_j \\ &= v_i^T \lambda_j v_j, \text{ since } v_j \text{ is an eigenvector of } A^T A \text{ for } \lambda_j \\ &= \lambda_j (v_i . v_j) = 0, \text{since } \{v_1, \dots, v_n\} \text{ is orthonormal basis for } \mathbb{R}^n. \end{aligned}$$

Thus, the vectors $Av_1, Av_2, \dots, Av_r$ form an orthogonal set of non-zero vectors and are therefore LI. Finally, let $Y \in$ Col A be arbitrary element. Then $Y = AX$ for some vector $X \in \mathbb{R}^n$. Note that X can be expressed in terms of the basis $\{v_1, v_2, \dots, v_n\}, X = c_1 v_1 + c_2 v_2 + \cdots + c_n v_n$. Then $Y = AX = A(c_1 v_1 + c_2 v_2 + \cdots + c_n v_n) = c_1 Av_1 + c_2 Av_2 + \cdots + c_r Av_r$, since remaining terms are 0, as noted at the start of the proof. Thus, the vectors $Av_1, Av_2, \dots, Av_r$ span Col A. Moreover, rank $A = r$, which is equal to the no. of non-zero singular values of A (= no. of non-zero eigenvalues of A).

Singular Value Decomposition (Cont'd)

- **Theorem 8 (Singular Value Decomposition (SVD) of a Matrix):** Let A be an $m \times n$ matrix with rank r. Then A can be factored as a product as follows:

$$A = U\Sigma V^T, \text{ where}$$

  - Σ is an $m \times n$ matrix containing an $r \times r$ diagonal matrix D with the r non-zero singular values of A, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$, along the main diagonal. D is placed in upper left hand corner of Σ. Remaining entries of Σ are zero.
  - U is an $m \times m$ orthogonal matrix. In order to obtain U, we take the r vectors Av_i corresponding to the nonzero singular values, extend them to an orthogonal basis of $\mathbb{R}^m$ using the Gram-Schmidt Process (this step is needed only in case $r < m$), and finally normalize the vectors to obtain an orthonormal basis $\{u_1, u_2, \ldots, u_m\}$. U has the vectors u_i as its columns.
  - V is an $n \times n$ orthogonal matrix. The matrix V has as its columns the orthonormal basis $\{v_1, v_2, \ldots, v_n\}$ of eigenvectors of $A^T A$.

Proof of SVD Theorem (Theorem 8)

Let λ_i and v_i be as in Proposition 58. Thus, the v_i form an orthonormal basis for $\mathbb{R}^n$ consisting of eigenvectors of $A^T A$. Then $\{Av_1, \dots, Av_r\}$ is an orthogonal basis for Col A, which is a subspace of $\mathbb{R}^m$. Normalize each Av_i to obtain an orthonormal basis $\{u_1, u_2, \dots, u_r\}$ so that $u_i = \frac{Av_i}{||Av_i||} = \frac{1}{\sigma_i} Av_i$. Hence,

$$Av_i = \sigma_i u_i \tag{1}$$

Now, extend $\{u_1, \dots, u_m\}$ to an orthonormal basis of $\mathbb{R}^m$. Put $U = [u_1 \ u_2 \ \dots \ u_m]$ and $V = [v_1 \ v_2 \ \dots \ v_n]$. Note that U and V are orthogonal matrices by construction. Also,

$$\begin{aligned} AV &= A[v_1 \ v_2 \ \dots \ v_n] = [Av_1 \ Av_2 \ \dots \ Av_n] \\ &= [\sigma_1 u_1 \ \sigma_2 u_2 \ \dots \ \sigma_r u_r \ 0 \ \dots \ 0] \quad \text{(using (1))} \end{aligned} \tag{2}$$

Now, let D be $r \times r$ diagonal matrix with diagonal enteries $\sigma_1, \sigma_2, \dots, \sigma_r$ and we make D into an $m \times n$ matrix Σ (same size as A) by filling out with zeroes.

Proof of SVD Theorem (Theorem 8) (Cont'd)

Then:

$$
\begin{aligned}
U\Sigma &= [u_1 \ \ \ldots \ \ u_m] \left[\begin{array}{cccc|c}
\sigma_1 & 0 & \cdots & 0 & \\
0 & \sigma_2 & \cdots & 0 & \\
\vdots & \vdots & \ldots & \vdots & \mathbf{0} \\
0 & 0 & \cdots & \sigma_r & \\
\hline
 & & \mathbf{0} & & \mathbf{0}
\end{array}\right] \\
&= [\sigma_1 u_1 \ \ \ldots \ \ \sigma_r u_r \ 0 \ldots \ 0] \\
&= AV \quad \text{(from (2))}
\end{aligned}
$$

Since V is orthogonal, we have: $U\Sigma V^T = AVV^T = A$.

Remark: Any factorization $A = U\Sigma V^T$, with U and V orthogonal matrices, Σ as described above, is called a Singular Value Decomposition (SVD) of A. Note that U and V are not uniquely determined by A, but the diagonal entries of Σ are necessarily the singular values of A.

Summary

- Let A be any $m \times n$ matrix – then A represents a linear transformation $X \mapsto AX$ from $\mathbb{R}^n$ to $\mathbb{R}^m$.
- Recall the following:
  1. The matrix $A^T A$ is a symmetric $n \times n$ matrix. Therefore, we select an orthonormal basis for $\mathbb{R}^n$, $\{v_1, v_2, \ldots, v_n\}$ consisting of eigenvectors of $A^T A$ with corresponding eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$.
  2. $\lambda_i = ||Av_i||^2 \geq 0$; Hence we can arrange the eigenvalues so that: $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$.
  3. The **singular values** of A are the square roots of the eigenvalues of $A^T A$, i.e. $\sigma_i = \sqrt{\lambda_i}$ for $i = 1, \ldots, n$.
  4. **Proposition 58:** Suppose that A has r nonzero singular values. Then:
     - Rank $A = r$
     - $\{Av_1, Av_2, \ldots, Av_r\}$ is an orthogonal basis for Col A.

Solved Example for SVD

$$A = \begin{bmatrix} -3 & 1 \\ 6 & -2 \\ 6 & -2 \end{bmatrix}$$

$$\therefore A^T A = \begin{bmatrix} -3 & 6 & 6 \\ 1 & -2 & -2 \end{bmatrix} \begin{bmatrix} -3 & 1 \\ 6 & -2 \\ 6 & -2 \end{bmatrix} = \begin{bmatrix} 81 & -27 \\ -27 & 9 \end{bmatrix} = B, \text{say}$$

Char poly of B is $\det(B - \lambda I) = (81 - \lambda)(9 - \lambda) - 729 = \lambda(\lambda - 90)$.
Therefore the eigenvalues in decending order are $\lambda_1 = 90, \lambda_2 = 0$.

Putting $\lambda_1 = 90 : B - \lambda_1 I = \begin{bmatrix} -9 & -27 \\ -27 & -81 \end{bmatrix} \sim \begin{bmatrix} 1 & 3 \\ 0 & 0 \end{bmatrix}$.

Solving, we get $X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_2 \begin{bmatrix} -3 \\ 1 \end{bmatrix}$ and by suitable choice and

normalizing, we get $v_1 = \begin{bmatrix} \frac{3}{\sqrt{10}} \\ \frac{-1}{\sqrt{10}} \end{bmatrix}$.

Solved Example for SVD (Cont'd)

Putting $\lambda_2 = 0$ and row reducing $B - \lambda_2 I = \begin{bmatrix} 81 & -27 \\ -27 & 9 \end{bmatrix}$, we similarly get $v_2 = \begin{bmatrix} \frac{1}{\sqrt{10}} \\ \frac{3}{\sqrt{10}} \end{bmatrix}$. Note that v_1 and v_2 are orthonormal as expected.

$$\therefore V = \begin{bmatrix} \frac{3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \\ \frac{-1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \end{bmatrix}$$

and singular values are $\sigma_1 = \sqrt{90} = 3\sqrt{10}$ and $\sigma_2 = 0$ and hence

$$\Sigma = \begin{bmatrix} \frac{3}{\sqrt{10}} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

We now need to compute U.

$$u_1 = \frac{1}{\sigma_1} A v_1 = \frac{1}{3\sqrt{10}} \begin{bmatrix} -3 & 1 \\ 6 & -2 \\ 6 & -2 \end{bmatrix} \begin{bmatrix} \frac{3}{\sqrt{10}} \\ \frac{-1}{\sqrt{10}} \end{bmatrix} = \begin{bmatrix} \frac{-1}{3} \\ \frac{2}{3} \\ \frac{2}{3} \end{bmatrix}$$

Solved Example for SVD (Cont'd)

However, $Av_2 = 0$. So we need to extend u_1 to an orthonormal basis of $\mathbb{R}^3$ by solving the system:

$$\begin{bmatrix} \frac{-1}{3} & \frac{2}{3} & \frac{2}{3} \end{bmatrix} X = 0 \text{ or } x_1 = 2x_2 + 2x_3 \text{ or } X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} x_2 + \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} x_3.$$

We need to get an orthonormal solution (if necessary we could use Gram-Schmidt process).

If we take $x_2 = 2/3, x_3 = -1/3$, giving: $u_2 = \begin{bmatrix} 2/3 \\ 2/3 \\ -1/3 \end{bmatrix}$

and $x_2 = -1/3, x_3 = 2/3$ giving: $u_3 = \begin{bmatrix} 2/3 \\ -1/3 \\ 2/3 \end{bmatrix}$.

[**Check:** $< u_2, u_3 >= 4/3 - 2/3 - 2/3 = 0$]

Solved Example for SVD (Cont'd)

So finally:

$$A = U\Sigma V^T = \begin{bmatrix} \frac{-1}{3} & \frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & \frac{-1}{3} \\ \frac{2}{3} & \frac{-1}{3} & \frac{2}{3} \end{bmatrix} \begin{bmatrix} \frac{3}{\sqrt{10}} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{3}{\sqrt{10}} & \frac{-1}{\sqrt{10}} \\ \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \end{bmatrix}$$

Check:

$$AV = \begin{bmatrix} -3 & 1 \\ 6 & -2 \\ 6 & -2 \end{bmatrix} \begin{bmatrix} \frac{3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \\ \frac{-1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \end{bmatrix} = \begin{bmatrix} -\sqrt{10} & 0 \\ 2\sqrt{10} & 0 \\ 2\sqrt{10} & 0 \end{bmatrix}$$

$$U\Sigma = \begin{bmatrix} \frac{-1}{3} & \frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & \frac{-1}{3} \\ \frac{2}{3}\,\frac{-1}{3}\,\frac{2}{3} & & \end{bmatrix} \begin{bmatrix} \frac{3}{\sqrt{10}} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} -\sqrt{10} & 0 \\ 2\sqrt{10} & 0 \\ 2\sqrt{10} & 0 \end{bmatrix}$$

A tip for doing this type of problem

Suppose we had instead been given a 2×3 matrix, say A_1. Then $A_1^T A_1$ would be a square 3×3 matrix. Finding its char poly, solving etc would be much harder.

What is the tip?

Instead work with $B_1 = A_1^T$, then $B_1^T B_1$ would be $(A_1^T)^T A_1^T = A_1 A_1^T$, i.e., $(2 \times 3) \times (3 \times 2)$ or 2×2 matrix, which is relatively easier to handle.

Suppose we solve to get

$$B_1 = U_1 \Sigma V_1^T \tag{3}$$

Then, $A_1 = B_1^T = (U_1 \Sigma V_1^T)^T = V_1 \Sigma^T U_1^T$, which we can easily get from (3).

Example 3 of Lay: Find SVD for $A = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix}$?

Answer: Using the tip given earlier, we find an SVD for

$$B = A^T = \begin{bmatrix} 4 & 8 \\ 11 & 7 \\ 14 & -2 \end{bmatrix}$$

and then take its transpose.

$$\text{Now, } B^T B = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix} \begin{bmatrix} 4 & 8 \\ 11 & 7 \\ 14 & -2 \end{bmatrix} = \begin{bmatrix} 333 & 81 \\ 81 & 117 \end{bmatrix} = C, \text{say.}$$

Therefore char poly of $C = (\lambda - 360)(\lambda - 90)$.
$\therefore \lambda_1 = 360, \sigma_1 = 6\sqrt{10}$ and $\lambda_2 = 90, \sigma_2 = 3\sqrt{10}$.
Hence,

$$\Sigma = \begin{bmatrix} 6\sqrt{10} & 0 \\ 0 & 3\sqrt{10} \\ 0 & 0 \end{bmatrix}.$$

We now need to find the eigenvectors of matrix C corresponding to λ_1 and λ_2.

Putting $\lambda_1 = 360 : C - \lambda_1 I = \begin{bmatrix} -27 & 81 \\ 81 & -243 \end{bmatrix} \sim \begin{bmatrix} 1 & -3 \\ 0 & 0 \end{bmatrix}$

$\therefore v_1 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$ is a suitable eigenvector.

Putting $\lambda_2 = 90 : C - \lambda_2 I = \begin{bmatrix} 243 & 81 \\ 81 & 27 \end{bmatrix} \sim \begin{bmatrix} 1 & 1/3 \\ 0 & 0 \end{bmatrix}$

$\therefore v_2 = \begin{bmatrix} 1 \\ -3 \end{bmatrix}$ is a suitable eigenvector.

Note that v_1 and v_3 are orthogonal, we will normalize them later.

We now need to find the u_i which are given by Bv_i (followed by normalizing at the last step)

$\therefore u_1 = \begin{bmatrix} 4 & 8 \\ 11 & 7 \\ 14 & -2 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 20 \\ 40 \\ 40 \end{bmatrix}$, so we cab take $\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$ as u_1

and $u_2 = \begin{bmatrix} 4 & 8 \\ 11 & 7 \\ 14 & -2 \end{bmatrix} \begin{bmatrix} 1 \\ -3 \end{bmatrix} = \begin{bmatrix} -20 \\ -10 \\ 20 \end{bmatrix}$, so we can take $\begin{bmatrix} -2 \\ -1 \\ 2 \end{bmatrix}$ as u_2.

Again u_1 and u_2 are orthogonal. To get an orthogonal basis for $\mathbb{R}^3$, we must find a vector orthogonal to both u_1 and u_2, i.e., we solve the system which has u_1^T and u_2^T as its rows.

$$\begin{bmatrix} 1 & 2 & 2 \\ -2 & -1 & 2 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 2 & 2 \\ 0 & 3 & 6 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 2 & 2 \\ 0 & 1 & 2 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & -2 \\ 0 & 1 & 2 \end{bmatrix},$$

so we get $u_3 = \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix}$ as a suitable vector.

To complete the process, we normalize the v's and u's giving:

$$v_1 = \frac{1}{\sqrt{10}} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{3}{\sqrt{10}} \\ \frac{1}{\sqrt{10}} \end{bmatrix}, v_2 = \frac{1}{\sqrt{10}} \begin{bmatrix} 1 \\ -3 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{10}} \\ \frac{-3}{\sqrt{10}} \end{bmatrix}$$

$$u_1 = \begin{bmatrix} 1/3 \\ 2/3 \\ 2/3 \end{bmatrix}, u_2 = \begin{bmatrix} -2/3 \\ -1/3 \\ 2/3 \end{bmatrix}, u_3 = \begin{bmatrix} 2/3 \\ -2/3 \\ 1/3 \end{bmatrix}$$

So we put $B = U\Sigma V^T$. Hence

$$A = B^T = V\Sigma^T U^T$$
$$= \begin{bmatrix} \frac{3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \\ \frac{1}{\sqrt{10}} & \frac{-3}{\sqrt{10}} \end{bmatrix} \begin{bmatrix} 6\sqrt{10} & 0 & 0 \\ 0 & 3\sqrt{10} & 0 \end{bmatrix} \begin{bmatrix} 1/3 & 2/3 & 2/3 \\ -2/3 & -1/3 & 2/3 \\ 2/3 & -2/3 & 1/3 \end{bmatrix}$$

The above is the required SVD of A.

Remark: One can work directly with A to get the answer.