Singular Value Decomposition and Applications

Uri Shaham

October 11, 2022

1 Introduction

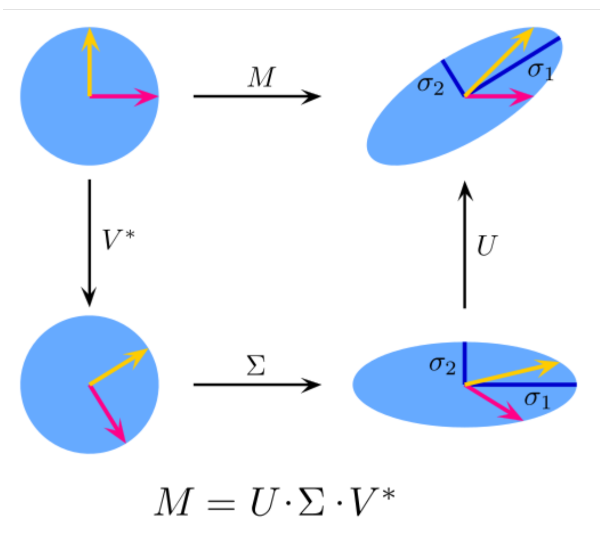
Definition 1.1 (SVD). *Let $A \in \mathbb{R}^{n\times m}$ be a real-valued matrix. The singular value decomposition (SVD) of A is a matrix factorization*

$$A = U\Sigma V^T, \tag{1}$$

where U is $n\times n$ orthogonal matrix (i.e., $UU^T = I_{n\times n}$), V is $m\times m$ orthogonal matrix and Σ is $n\times m$ diagonal matrix (i.e., $\Sigma_{ij} = 0$ for $i \neq j$.

Observation 1.2. $U\Sigma V^T = \sum_{i=1}^{r} \sigma_i u_i v_i^T$ *where* $r = \min\{n, m\}$.

Observation 1.3. *When $n = m$, A can be viewed as an operator from $\mathbb{R}^n$ to $\mathbb{R}^n$, acting on any vector x by rotation (possibly with reflection), axis rescaling and another rotation.*



1.1 Existence and uniqueness of SVD

Theorem 1.4. *Existence of SVD Any matrix $A \in \mathbb{R}^{n\times m}$ has a SVD.*

Proof. The matrix A^TA is symmetric and positive semi-definite (to see this, assume that $\lambda < 0$ is an eigenvalue and let x be the corresponding eigenvector. Then $X^TA^TAx < 0$). Then A^TA has an eigendecomposition $A^TA = V\Lambda V^T$ with real eigenvectors and non-negative eigenvalues. Let $r = \text{rank}(A^TA)$. Wlog, assume that $\lambda_1 \geq \lambda_2, \ldots \geq \lambda_r > 0$ and $\lambda_{r+1} = \ldots = \lambda_m = 0$. Set $\sigma_i = \sqrt{\lambda_i}$, for $i = 1, \ldots, m$. Define $u_i = \frac{Av_i}{\sigma_i}$ for $i = 1, \ldots, m$. Then $u_1, \ldots, u_m$ are orthonormal:

$$u_i^T u_j = \frac{v_i T A^T A v_j}{\sigma_i \sigma_j} = \frac{(Av_i)^T A v_j}{\sigma_i \sigma_j} = \delta_{ij},$$

and

$$U\Sigma V^T = AV\Sigma^{-1}\Sigma V^T = A$$

. □

Theorem 1.5. *Let $A = U\Sigma V^T$ and $\Sigma_{ii} \geq \Sigma_{jj}$ for $i < j$. Then Σ is uniquely determined.*

Proof. Let $i \in \{1, \ldots, \min\{n, m\}\}$, and let e_i be the $i'th$ standard basis vector of $\mathbb{R}^m$. Then Since U, V are orthonormal, they don't affect $\|Ax\|$ for any $x \in \mathbb{R}^m$, hence $\|Aei\| = \sigma_i$, so σ_i is uniquely determined. □

1.2 Power iteration

Observation 1.6. *Let $A = U\Sigma V^T$ and let $B = A^TA$. Then $B = V\Sigma^2 V^T$ and more generally, $B^k = V\Sigma^{2k}V^T$. In addition, Σ^{2k} is diagonal with entries σ_i^{2k}.*

Assume that $\sigma_1 > \sigma_2 > \ldots \sigma_n$. Then for k large enough $\sigma_1^{2k} \gg \sigma_2^{2k}$, hence $B^k =\approx \sigma_1^{2k} v_1 v_1^T$. Therefore, $B^k x$ is approximately in the direction of v_1. This gives an approach to find v_1: Starting from any vector x not orthogonal to v_1, $\frac{B^k x}{\|B^k x\|} \to v_1$ as $k \to \infty$.

2 Applications

2.1 Low rank approximation

Definition 2.1 (spectral norm)**.** *Let $A = U\Sigma V^T = \sum_{i=1}^r u_i v_i^T$ be $n \times m$ matrix, and assume that $\sigma_1 \geq \ldots \geq \sigma_r$. The spectral norm of A is defined as $\|A\| = \sigma_1$.*

Theorem 2.2 (spetral norm is matrix 2-norm)**.** $\|A\| = \sup_{\|x\|_2 = 1} \|Ax\|_2$

Theorem 2.3 (Eckart-Young 1936)**.** *The best rank k approximation of A in spectral norm is $A_k = \sum_{i=1}^k u_i v_i^T$.*

Proof. First, note that $\|A - A_k\| = \|\sum_{i=k+1}^r u_i v_i^T\| = \sigma_{k+1}$. Let B_k be any $n \times m$ rank k matrix, i.e., $B_k = XY_T$, where X and Y have k columns each. Since Y has k columns, is a non-trivial linear combination of the first $k+1$ columns that $w := \sum_{i=1}^{k+1} \gamma_i v_i$ gives $Yw = 0$. Then $B_k w = 0$. Wlog $\|w\| = 1$, i.e., $\sum_{i=1}^{k+1} \gamma^2 = 1$ (by Pythagoras). Hence we have

$$\|A - B_k\|^2 \geq \|(A - B_k)w\|^2 = \|Aw\|^2 = \sum_{i=1}^{k+1} \sigma_i^2 \gamma_i^2 \geq \sigma_{k+1} = \|A - A_k\|.$$

□

2.2 Pseudo inverse

Definition 2.4. *The Pseudo inverse of matrix $A = U\Sigma V^T$ is $A^\dagger = U\Sigma^\dagger V^T$, where $\Sigma^\dagger$ is obtained from Σ by replacing all nonzero singular values by their reciprocals.*

Pseudo inverse can be used to solve least squares problems as follows. Let $z = A^\dagger b$. Then $\|Az - b\| \leq \|Ax - b\|$ for all $x \in \mathbb{R}^m$.

2.3 Matrix square root

Let A be a symmetric $n \times n$ PSD matrix with SVD $A = V\Sigma V^T$ (see p3@homework). Let $\Sigma^{\frac{1}{2}}$ be $\text{diag}(\sqrt{\sigma_1}, \ldots, \sqrt{\sigma_n})$. Then for $B = U\Sigma^{\frac{1}{2}}V^T$ we have $BB = A$.

2.4 Sampling from multivariate normal distribution

To sample from a $\mathcal{N}(\mu, K)$ normal distribution:

1. sample a $\mathcal{N}(0, I)$ vector x (easy - each coordinate separately from a $\mathcal{N}(0, 1)$ Gaussian distribution).
2. set $y = \mu + K^{\frac{1}{2}}x$ (note that as K is covariance, it is PSD).

2.5 PCA

Let A be a $n \times d$ matrix with mean-centered columns, representing n data points in d dimensions. Then the sample covariance matrix is $\frac{1}{n-1}A^TA$. In PCA, the principal directions are the eigenvectors V of the covariance matrix, and the new representation is $U = AV$. Observe that these are exactly the matrices in the SVD of A, $A = U\Sigma V^T$, as the covariance is $V\Sigma^2V^T$, and we already now that U can be obtained from V via $U = AV\Sigma^{-1}$

Homework

1. Let $A = U\Sigma V^T = \sum_{i=1}^{r} u_i v_i^T$. Prove that $\{u_1, \ldots, u_r\}$ are an orthonormal basis of $\text{col}(A)$ and that $\{v_1, \ldots, v_r\}$ are an orthonormal basis of $\text{row}(A)$
2. How power iteration can be used to find v_2?
3. Let A be a symmetric matrix with SVD $A = U\Sigma V^T$.
   - Prove that $U = V$, up to column sign flips.
   - If A is also positive semi-definite, prove that $U = V$.
4. WRT to sampling from multivariate Gaussian, prove that: (i) K is PSD (ii) $y \sim \mathcal{N}(\mu, K)$.
5. Programming: implement power iteration, computing SVD of a random 100×2 matrix, and compare to the numpy result.