Tutorial On The Singular Value Decomposition

Javier R. Movellan

1 SVD Theorem

Let A be a real valued $m \times n$ matrix, where $m \geq n$. Then A can be decomposed as follows:

$$A = UWV^T \tag{1}$$

where U is a real valued $m \times m$ orthonormal matrix: $UU^T = I_m$, W is an $m \times n$ matrix whose diagonal elements are non-negative real values and whose off-diagonal elements are zero, and V is a $n \times n$ real valued orthonormal matrix; $VV^T = I_n$. The columns of U are called the left singular vectors. the columns of V are called the right singular vector, the diaonal elements of W are called called the *singular values*. The singular value decomposition exists always and is unique up to 1) Same permutations in columns of U, W and V. 2) Linear combinations of colummns of U and V with equal singular values.

Note

$$AA' = UWV^TVW^TU^T = UWW^TU^T \tag{2}$$

Thus the columsn of U are the eigenvectors of AA^T and the diagonal elements of WW^T are its eigenvalues. Moreover

$$A'A = VW^TU^TUWV^T = VW^TWV^T \tag{3}$$

Thus the columsn of V are the eigenvectors of A^TA and the diagonals elements of W^TW its eigenvalues.

If X is a matrix $X_{i,j}$ will represent the element in the i^{th} row, j^{th} column, and X_i will represent the i^{th} column vector. Note $A_{i,j} = \sum_{k=1}^{n} w_{k,k} U_{i,k} V_{i,k}$. Thus we can approximate A well by deleting columns of U and V with small singular values.

2 Properties for Square Matrices

2.1 Definitions

1. A be an $n \times n$ square matrix.
2. $N_A = \{x \in \mathbb{R}^n : Ax = 0\}$, the **null space** of A.
3. $\dim N_A$, the dimensionality of the null space of A, also known as the **nullity** of A.
4. $R_A = \{x \in \mathbb{R}^n : Ax \neq 0\}$, the **range** of A.
5. $\dim R_A$, the dimensionality of the range of A, also know as **the rank** of A. It is well known that rank plus nullity equals n.

2.2 Properties

1. $\dim R_A = \text{card}\{w_i \in \text{diag}W : w_i > 0\}$
2. $\dim N_A = \text{card}\{w_i \in \text{diag}W : w_i = 0\}$
3. Let U_i the i^{th} column of U. Then $\{U_i : w_i > 0\}$ is a basis of R_A and $\{U_i : w_i = 0\}$ is a basis of N_A
4. Let $\tilde{W}$ be a diagonal matrix with

$$\tilde{w}_{i,i} = \begin{cases} \frac{1}{w_{i,i}} & \text{if } w_{i,i} > 0 \\ 0 & \text{if } w_{i,i} = 0 \end{cases} \tag{4}$$

5. If A is square (i.e., $m = n$) and all $w_i > 0$. Then $A^{-1} = VW^{-1}U^T$.

   To see why simply note that $(VW^{-1}U^T)(UWV^T) = I_n$

6. The value $\hat{x} = V\tilde{W}U^Tb$ solves for the linear equation $Ax = b$ in the following sense
   (a) If A is non-singular $\hat{x}$ is the unique solution to the equation.
   (b) If A is singular and $b \in R_A$ then $\hat{x}$ is the solution with smallest norm.
   (c) If A is singular and $b \in N_A$ then $\hat{x} = \text{argmin}_x |Ax - b|$

2.3 Rectangular matrices

1. If $m < n$ the system $Ax = b$ has less equations than unknowns. Patch A adn b with zeroes to form an $n \times n$ system. The solution found via SVD is minimum norm.
2. If $m > n$ there are more equations than unknowns. The solution found via SVD is least squares.

2.4 Neural net interpretation

Let rows of A represent exemplars of dimensionality n. We know dropping $n - p$ small singular values and their associated columns in U and V allows us to get an approximation to A. Let $\hat{U}$, $\hat{W}$ and $\hat{V}$ represent "chopped" versions of the corresponding matrices, and $\hat{A} = \hat{U}\hat{W}\hat{V}^T$. We can retrieve the approximation to the i^{th} exemplar by premultiplying $\hat{A}$ times a row vector with all components set to 0 except the i^{th} component which is set to 1.

$$\hat{A}_i^T = [0, \cdots, 0, 1, 0 \cdots, 0]\hat{A} = [0, \cdots, 0, 1, 0 \cdots, 0]\hat{U}\hat{W}\hat{V}^T \tag{5}$$

In a neural network interpretation, the input vector $[0, \cdots, 0, 1, \cdots, 0]$ is a local representation for the i^{th} exemplar in the database. This exemplar is transformed into a hidden representation by multiplication by the matrix of weights $\hat{U}\hat{W}$. This results into a hidden representation of dimensionality p. This hidden representation is then multiplied times the orthonormal matrix $\hat{V}^T$ to obtain a representation of dimensionality n. This representation should approximate the original dimensions of the i^{th} exemplar in the database. We can actually use the hidden layer representation as a concise representation of the exemplar.

3 SVD and eigen decompositions

Let X be an $m \times n$ matrix of data, where m is the number of observations and n the number dimensionality of each observation. In computer vision problems typically $m > n$. The covariance matrix is $C_x = \frac{X'X}{m}$. The matrix of eigenvectors P is such that $C_x == P\Lambda P^T$, where Λ is diagonal and P is orthonormal. Now define

$$A = \frac{X'}{\sqrt{m}}$$

therefore $C_x = AA'$.

A is an $n \times m$ matrix and can be decomposed using singular value decomposition (SVD) in $A = UWV^T$ We can then rewrite C_x as: $Cov = AA' = UWV^TVWU^T = UW^2U^T$. Thus U columns of U are the eigenvectors of C_x an the squared singular values are the eigenvalues of C_x.

3.0.1 Shortcuts

Consider $T = \frac{XX'}{m} = A'A$, an $m \times m$ matrix. In computer vision problems typically the number of images in the database is smaller than the number of pixels per image and thus it is preferable to work with T rather than C_n. If P is a matrix of eigenvectors of T then AP is an eigenvector of C_x. To see why note that e_i is an eigenvector of T iff $Te_i = A'Ae_i = \lambda_i e_i$. Thus $AA'Ae_i = C_x(Ae_i) = \lambda_i(Ae_i)$ and (Ae_i) is an eigenvector of C_s with eigenvalue $A\lambda_i$.

If we are interested on geting only p eigenvectors (e.g., if we have more dimensions (m) than samples (n) we certainly do not want more than m eigenvectors) we can use an "economy" version of the SVD. There are routines that compute only the first p columns of U and therefore the first m rows of W giving U of size $n \times p$, W of size $p \times p$ and V of size $p \times p$. before.

These shortcuts avoid the multiplication needed to obtain C_x and the computation of all the columns of U and W we are not interested in.