CISC 271   Class 14

SVD – Singular Value Decomposition

Text Correspondence: §7.1

Main Concepts:

- *Eigenvectors of AA^T: left singular vectors*
- *Eigenvectors of A^TA: right singular vectors*
- *Non-zero eigenvalues are the same for AA^T and A^TA*
- *SVD describes "eigenvectors" of a non-square matrix*

Sample Problem, Machine Inference: For a non-square data matrix, how can we find "eigenvalues"?

So far in this course, we have found many uses for eigenvalues and eigenvectors of square matrices. In practice, we often encounter non-square matrices of data, such as empirical findings that are summarized in a table of numbers. How can we perform something like an eigenvalue analysis of a data matrix? We can reason from what we know already and try to find an approach.

We know that, if A is a square invertible matrix, then the columns of A are linearly independent. We can also infer that, for such a matrix A, that A^TA is a square and symmetric positive-definite matrix. What can we infer if A is not square?

14.1 Eigenvectors of the matrix $[A^TA]$

Suppose that A is a "tall thin" matrix of full rank. By this we mean that $A \in \mathbb{R}^{m\times n}$ has more rows than columns, so $m > n$, and that the rank of A is n. Using a result from the previous class, we can infer that $[A^TA]$ is a square symmetric positive definite matrix. Because $[A^TA]$ is symmetric, it can be diagonalized as the matrix decomposition

$$[A^TA] = V\Lambda V^T \tag{14.1}$$

where

$$V = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_n \end{bmatrix}$$

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ 0 & 0 & \ddots & \vdots \\ \vdots & \vdots & & \lambda_n \end{bmatrix} \tag{14.2}$$

We have permuted the eigenvalues in Equation 14.2 so that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n > 0$. The rank of the matrix $[A^T A]$ is n and we also have to permute the rows of V when we order the eigenvalues by their magnitudes.

What if the matrix A is not full rank? We know that the diagonal matrix Λ of Equation 14.1 will have some zero values on the diagonal entries. We will have to modify Equation 14.2 to account for the zero eigenvalues, which would give us

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & & \cdots & & & 0 \\ 0 & \lambda_2 & & \cdots & & & 0 \\ 0 & 0 & \ddots & & & & \vdots \\ \vdots & \vdots & & \lambda_r & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{bmatrix} \tag{14.3}$$

From Equation 14.3, we can see that the rank of the matrix $[A^T A]$ is r.

Suppose, instead, that the matrix A is a "short wide" matrix. Using the same kind of reasoning as we used for a "tall thin" matrix, we can draw the same inference: the columns of the matrix V in Equation 14.1 are eigenvectors of $[A^T A]$.

14.2 Eigenvectors of the matrix $[AA^T]$

What happens when we consider the other product of a matrix with its transpose, which is AA^T? We begin with the clearest case and proceed to more complicated cases.

Suppose that A is a "short wide" full-rank matrix. The same kinds of reasoning allow us to draw the same conclusions about the matrix $[AA^T]$. This will be a square symmetric matrix of size $m \times m$.

Suppose that A is either square and rank-deficient, or is a "tall thin" matrix. For such a matrix, $[AA^T]$ will also be positive semi-definite and can be diagonalized. The decomposition will be like that in Equation 14.1 but the orthogonal matrix will have a different size. The decomposition can be written as

$$[AA^T] = U\Lambda U^T$$

The $m \times m$ orthogonal matrix U, and the diagonal matrix Λ, will be

$$
\begin{aligned}
U &= \begin{bmatrix} \vec{u}_1 & \vec{u}_2 & \cdots & \vec{u}_m \end{bmatrix} \\
\Lambda &= \begin{bmatrix}
\lambda_1 & 0 & & \cdots & & & 0 \\
0 & \lambda_2 & & \cdots & & & 0 \\
0 & 0 & \ddots & & & & \vdots \\
\vdots & \vdots & & \lambda_r & & & \\
 & & & & 0 & & \\
 & & & & & \ddots & \\
0 & 0 & & \cdots & & & 0
\end{bmatrix}
\end{aligned}
\tag{14.4}
$$

An important result, which we will prove at the end of these notes, is that the r non-zero eigenvalues of AA^T are the same as those of A^TA. We can see this easily if A is square and diagonalizable. What if A is square and not diagonalizable, or if A is rectangular?

14.3 The Singular Value Decomposition, or SVD

One of the most useful theorems of 20th–century linear algebra is the singular-value decomposition, commonly called the SVD. It is a general result for rectangular complex matrices, but for real matrices the unitary factors are simply orthogonal matrices.

For any real-valued matrix $A \in \mathbb{R}^{m \times n}$ there is an orthogonal matrix $U \in \mathbb{R}^{m \times m}$, an orthogonal matrix $V \in \mathbb{R}^{n \times n}$ and a "diagonal" matrix $\Sigma \in \mathbb{R}^{m \times n}$ such that

$$
A = U \Sigma V^T \tag{14.5}
$$

where

$$
\Sigma = \begin{bmatrix}
\sigma_1 & 0 & & \cdots & & & 0 \\
0 & \sigma_2 & & \cdots & & & 0 \\
0 & 0 & \ddots & & & & \vdots \\
\vdots & \vdots & & \sigma_r & & & \\
 & & & & 0 & & \\
 & & & & & \ddots & \\
 & & & & & & 0 \\
0 & & & \cdots & & & 0
\end{bmatrix}
$$

The matrix Σ has four crucial properties:

(a) $\sigma_j \in R$
(b) $\sigma_j > 0$ for $j \leq r$
(c) $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ *and*
(d) r is the rank of A

The existence of the SVD was known in 1936 but the earliest reliable computation of the SVD dates to approximately 1965.

14.4 Using the SVD

The SVD has a vast number of applications. It is widely used in linear algebra and is also found in many applications that are formulated in linear algebra. It can be used to prove matrix properties, find the "nearest" orthogonal matrix to a given square matrix, and determine the "best" orthonormal basis vectors for a vector space. More recently, it was used in recommendation computations, such as the Netflix algorithm for predicting user ratings of videos.

An example use is that, knowing the SVD, we can very easily prove the decompositions that we deduced for the products of a matrix and its transpose. Using the rule of matrix transposition, and the fact that the transpose of an orthogonal matrix is its inverse, we can see that

$$\begin{aligned}
[AA^T] &= [U\Sigma V^T][U\Sigma V^T]^T \\
&= [U\Sigma V^T][V\Sigma^T U^T] \\
&= U\Sigma[V^T V]\Sigma^T U^T \\
&= U\Sigma\Sigma^T U^T \\
&= U\Lambda U^T \\
[A^T A] &= [U\Sigma V^T]^T[U\Sigma V^T] \\
&= [V\Sigma^T U^T][U\Sigma V^T] \\
&= V\Sigma^T[U^T U]\Sigma V^T \\
&= V\Sigma^T\Sigma V^T \\
&= V\Lambda V^T
\end{aligned}$$

This confirms that our reasoning was correct. It also demonstrates that the singular values of the matrix A, in the "diagonal" matrix Σ, are the square roots of the eigenvalues of $[AA^T]$ and $[A^T A]$; these eigenvalues are the same because of the SVD of the matrix A.