CISC 271 Class 15

Orthonormal Basis Vectors and the SVD

Text Correspondence: §7.1

Main Concepts:

- *Left singular vectors: orthonormal basis of* ***data*** *vectors*
- *Right singular vectors: orthonormal basis of* ***weight*** *vectors*
- *Singular values: Positive real numbers, generalized "eigenvalues"*

Sample Problem, Machine Inference: For a set of data vectors, what are the "best" vectors that approximate the vector space of the data?

There are many lessons that we can draw from the SVD of a matrix. In this course we will use the SVD primarily to find a set of basis vectors for a vector space, so we will explore the decomposition for square and non-square matrices.

15.1 SVD of a Square Matrix

If a matrix $A \in \mathbb{R}^{m \times m}$ has m rows and m columns, then the columns are vectors in a data space $\mathbb{R}^m$. The SVD of the matrix A will be

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

where all of the factors on the right side are square $m \times m$ matrices. They have basic properties:

- ***U*** is an orthogonal matrix
  the columns of U $\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_m$ are a basis for the data space $\mathbb{R}^m$
- **Σ** is a diagonal matrix of non-negative real numbers
  - the first diagonal entry σ_1 is the largest number in Σ
  - the smallest non-zero entry in Σ is σ_r
  - the rank of the matrix A is r
- ***V*** is an orthogonal matrix
  its columns are a basis for the weight space $\mathbb{R}^m$

From these properties, we can infer that the first r columns of U are a basis for the column space of A.

Example: Square asymmetric matrix of full rank. Consider

$$A_1 = \begin{bmatrix} 3 & 4 \\ 0 & 3 \end{bmatrix}$$

By inspection, the columns of A_1 are linearly independent so the column span is $\mathbb{R}^2$. The eigenvalues – which are the diagonal entries because A_1 is upper triangular – are $\lambda_1 = 3$ and $\lambda_2 = 3$. The eigenvectors can be computed as

$$\lambda_1(A_1) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad \lambda_2(A_1) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

The asymmetric matrix A_1 is not diagonalizable. This is because the eigenvectors are not linearly independent.

The SVD of A_1 can be found by hand, using the textbook algorithm, or can be estimated by computation. Doing the latter, using two digits of numerical precision, we get the approximate values

$$U_1 = \begin{bmatrix} 0.88 & -0.47 \\ 0.47 & 0.88 \end{bmatrix} \qquad \Sigma_1 = \begin{bmatrix} 5.60 & 0 \\ 0 & 1.61 \end{bmatrix} \qquad V_1 = \begin{bmatrix} 0.47 & -0.88 \\ 0.88 & 0.47 \end{bmatrix}$$

We can see that the columns of U_1 are an orthonormal basis for the data space, and the columns of V_1 are an orthonormal basis for the weight space.

A remarkable property of the SVD of A_1 is that it produces a decomposition having a diagonal matrix, even though A_1 is not diagonalizable!

Example: Square symmetric rank-deficient matrix. Consider

$$A_2 = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

This matrix is, by inspection, symmetric and rank-deficient because the second column is -1 times the first column. Being symmetric it is diagonalizable, so we expect the SVD to have $U_2 = V_2$. Computing the SVD of A_2, we find that

$$U_2 = \begin{bmatrix} -0.71 & 0.71 \\ 0.71 & 0.71 \end{bmatrix} \qquad \Sigma_2 = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \qquad V_2 = \begin{bmatrix} -0.71 & 0.71 \\ 0.71 & 0.71 \end{bmatrix}$$

Because only one singular value of A_2 is non-zero, the rank of A_2 is 1. The first singular value, which is 2, indicates that the first column of U_2 is a basis vector for the column space of A_2. The second column of U_2 is orthogonal to the first column, so it is a basis vector for the complement of the column space of A_2.

15.2 SVD of a Non-Square Matrix

If a matrix $A \in \mathbb{R}^{m \times n}$ has m rows and n columns, with $m \neq n$, then the columns are vectors in a data space $\mathbb{R}^m$ and they act on a weight vector in a weight space $\mathbb{R}^n$. The SVD of the matrix A will always be

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

but we must be careful when we interpret the singular vectors.

Example: Non-square matrix of full rank. Consider the "tall thin" matrix

$$A_3 = \begin{bmatrix} 1 & -1 \\ -1 & 1 \\ 2 & 0 \end{bmatrix}$$

Computing the SVD of A_3, we find that

$$U_3 = \begin{bmatrix} -0.50 & 0.50 & -0.71 \\ 0.50 & -0.50 & -0.71 \\ -0.71 & -0.71 & 0.00 \end{bmatrix} \qquad \Sigma_3 = \begin{bmatrix} 2.61 & 0 \\ 0 & 1.08 \\ 0 & 0 \end{bmatrix} \qquad V_3 = \begin{bmatrix} -0.92 & -0.38 \\ 0.38 & -0.92 \end{bmatrix}$$

The SVD of A_3 tells us that the first two columns of U_3 are an orthonormal basis for the column space of A_3. This may seem unusual because the columns of A_3 are also a basis. The distinction is that U_3 is, in a numerical and mathematical sense, the "best" basis for the vector space *in the absence of other information*. Later in the course, we will look at how to find an orthonormal basis by using the matrix A directly.

The columns of V_3 are an orthonormal basis for the weight space, which is $\mathbb{R}^2$ because A_3 is full rank. Here, too, the SVD has selected a basis that a human might not have selected.

Example: Non-square matrix that is rank-deficient. Consider the "tall thin" matrix

$$A_4 = \begin{bmatrix} 1 & -1 \\ -1 & 1 \\ 2 & -2 \end{bmatrix}$$

Computing the SVD of A_4, we find that

$$U_4 = \begin{bmatrix} -0.42 & -0.91 & 0.00 \\ 0.42 & -0.18 & 0.89 \\ -0.82 & 0.37 & 0.45 \end{bmatrix} \qquad \Sigma_4 = \begin{bmatrix} 3.46 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \qquad V_4 = \begin{bmatrix} -0.71 & 0.71 \\ 0.71 & 0.71 \end{bmatrix}$$

The rank of A_4 is 1 because only the first singular value in Σ_4 is non-zero. The first column of U_4 is a basis vector for the column space of A_4; this is a unit-length version of either column of A_4 and might be what a human selected.

The first column of V_4 is a basis for the weight space of A_4. This informs us that, for a non-zero data vector $\vec{c}_4$, the only solution to $A_4\vec{w} = \vec{c}_4$ is a vector $\vec{w}$ for which the first entry is the negative of the second entry.

Even more telling is the second column of V_4. This is a basis for the null space of A_4 because any vector $\vec{w}$ for which the first entry equals the second entry is mapped to the zero vector $\vec{0}$. The reason is subtle and useful: in general, every zero diagonal entry of the matrix Σ selects a basis vector for the null space of the original matrix A.

15.3 The SVD as an Approximate Basis for a Vector Space

In this course, the SVD will be especially useful in performing numerical approximations.

For a matrix A we have seen that, if the first r entries of the matrix Σ are non-zero, then the rank of A is r. What if the r^{th} singular value is negligible?

To be negligible, we would mean that a singular value can be neglected. This will depend on the application but a good first way to address this problem is to consider all of the non-zero singular values as an ensemble. If σ_r is much smaller than σ_1, we might want to neglect it and just use $r-1$ basis vectors to approximate the vector space of the columns of the data in the matrix A. Two methods can be found to be in common current use:

- If σ_r/σ_1 is "small", neglect the effects of $\vec{u}_r$
- For the sum of preceding singular values

  $$l_r = \sum_{j=1}^{r}$$

  if σ_r/l is "small", neglect the effects of $\vec{u}_r$

To understand these methods in more depth, we can think of gathering the singular values into a vector $\vec{\sigma}$.

The first method uses the ratio of the largest singular value and the smallest singular value, which is an extension of the condition number to a non-square matrix. Because of how the singular values are ordered, the first entry of $\vec{\sigma}$ is the largest entry; this entry is the "L-infinity" or L_∞ norm, so we are basing the cut-off on $\sigma_r / \|\vec{\sigma}\|_\infty$.

The second method uses the sum of the singular values, which is L_1 norm of $\vec{\sigma}$ calculated up to and including σ_r. By taking into account all of the relevant singular values, we are basing the cut-off on $\sigma_r / \|\vec{\sigma}\|_1$.

Of course, these methods do not need to apply to only the smallest non-zero singular value σ_r. We might apply the methods to another singular value, perhaps this index k, which would select k columns of U as an approximate basis for the data in the matrix A. This is the concept that we will use when we perform principal-component analysis of large sets of data.

15.4 Some SVD Properties

Suppose that a matrix $A \in \mathbb{R}^{m \times n}$ is a "tall thin" matrix that has $m > n$ and $\text{rank}(A) = r$. The SVD of A is described in Equation 15.1, in which we can "read out" the four matrix spaces of the matrix.

$$A = \begin{bmatrix} U_{1\ldots r} & U_{(r+1)\ldots m} \end{bmatrix} \begin{bmatrix} \Sigma_{1\ldots r} & 0 \\ 0 & \Sigma_{(r+1)\ldots m} \end{bmatrix} \begin{bmatrix} V_{1\ldots r} & V_{(r+1)\ldots m} \end{bmatrix}^T \tag{15.1}$$

For the matrix A in Equation 15.1, we can see that:

- $U_{1\ldots r}$ is a basis for the column space of A
- $U_{(r+1)\ldots m}$ is a basis for the orthogonal complement of the column space of A
- $V_{1\ldots r}$ is a basis for the row space of A
- $V_{(r+1)\ldots n}$ is a basis for the null space of A

In summary, the SVD is a powerful matrix decomposition. Some of the properties that we may find useful include:

- $A \in \mathbb{R}^{m\times n} = U\Sigma V^T$ where U and V are orthogonal and Σ is "diagonal"
- Columns of $U \in \mathbb{R}^{m\times m}$ are an orthonormal basis for the data space $\mathbb{R}^m$
- Columns of $V \in \mathbb{R}^{n\times n}$ are an orthonormal basis for the weight space $\mathbb{R}^n$
- $\Sigma \in \mathbb{R}^{m\times n}$ has the same size as the matrix $A \in \mathbb{R}^{m\times n}$ that is factored
- Σ has zero in each off-diagonal entry
- The diagonal entries of Σ, written as σ_j, are non-negative real numbers that are ordered from largest to smallest
- If the smallest non-zero diagonal entry is σ_r then the rank of A is r
- The first r columns of U are a basis for the column space of A
- The first r columns of V are a basis for the weight space of A
- The last $(n-r)$ columns of V are a basis for the null space of A
- If A is diagonalizable then $U = V$

In this course we will neither prove these properties nor memorize them. Instead, we will use the properties to help us to find patterns in large sets of data.