**Singular Value Decomposition (SVD) tutorial**

BE.400 / 7.548

Singular value decomposition takes a rectangular matrix of gene expression data (defined as A, where A is a *n*x *p*matrix) in which the *n*rows represents the genes, and the *p*columns represents the experimental conditions. The SVD theorem states:

**A*nxp*= U*nxn* S*nxp* VT*pxp***

Where

**U**T**U** = **I**nxn

**V**T**V** = **I**pxp(i.e. U and V are orthogonal)

Where the columns of U are the left singular vectors (*gene coefficient vectors*); S (the same dimensions as *A*) has singular values and is diagonal (*mode amplitudes*); and VT has rows that are the right singular vectors (*expression level vectors*). The SVD represents an expansion of the original data in a coordinate system where the covariance matrix is diagonal.

Calculating the SVD consists of finding the eigenvalues and eigenvectors of *AAT*and *ATA*. The eigenvectors of *ATA* make up the columns of *V*, the eigenvectors of *AAT*make up the columns of *U*. Also, the singular values in **S** are square roots of eigenvalues from *AAT* or *ATA*.  The singular values are the diagonal entries of the *S*matrix and are arranged in descending order. The singular values are always real numbers. If the matrix *A*is a real matrix, then *U*and *V*are also real.

To understand how to solve for SVD, let’s take the example of the matrix that was provided in Kuruvilla *et al*:

**A close up of a logo

Description automatically generated**

In this example the matrix is a 4x2 matrix. We know that for an n x n matrix W, then a nonzero vector **x** is the eigenvector of W if:

W **x** =  **x**

For some scalar .  Then the scalar  is called an eigenvalue of A, and **x** is said to be an eigenvector of A corresponding to .

So to find the eigenvalues of the above entity we compute matrices *AAT*and *ATA*.  As previously stated , the eigenvectors of *AAT*make up the columns of *U*so we can do the following analysis to find U.

**A close up of a logo

Description automatically generated**

Now that we have a n x n matrix we can determine the eigenvalues of the matrix W.

Since W **x** =  **x** then   (W- I) **x** = 0

**A close up of a logo

Description automatically generated**

For a unique set of eigenvalues to determinant of the matrix (W-I) must be equal to zero.  Thus from the solution of the characteristic equation, |W-I|=0 we obtain:

=0, =0;  = 15+Ö221.5 ~ 29.883;   = 15-Ö221.5 ~ 0.117 (four eigenvalues since it is a fourth  degree polynomial).  This value can be used to determine the eigenvector that can be placed in the columns of U.  Thus we obtain the following equations:

19.883 x1 + 14 x2 = 0

14 x1 + 9.883 x2 = 0

x3  = 0

x4 = 0

Upon simplifying the first two equations we obtain a ratio which relates the value of x1 to x2.   The values of x1 and x2 are chosen such that the elements of the S are the square roots of the eigenvalues.   Thus a solution that satisfies the above equation x1 = -0.58 and x2 = 0.82 and x3 = x4 = 0 (this is the second column of the U matrix).

Substituting the other eigenvalue we obtain:

            -9.883 x1 + 14 x2 = 0

14 x1 - 19.883 x2 = 0

x3  = 0

x4 = 0

Thus a solution that satisfies this set of equations is x1 = 0.82 and x2 = -0.58 and x3 = x4 = 0 (this is the first column of the U matrix). Combining these we obtain:

A close up of a logo

Description automatically generated

Similarly *ATA* makes up the columns of *V* so we can do a similar analysis to find the value of V.

**A close up of a logo

Description automatically generated**

and similarly we obtain the expression:



Finally as mentioned previously the S is the square root of the eigenvalues from *AAT* or *ATA.* and can be obtained directly giving us:

A close up of a logo

Description automatically generated

Note that:  1 > 2 > 3 > … which is what the paper was indicating by the figure 4 of the Kuruvilla paper.  In that paper the values were computed and normalized such that the highest singular value was equal to 1.

Proof:

**A**=**USV**T and **A**T=**VSU**T

**A**T**A** = **VSU**T**USV**T

**A**T**A** = **VS**2**V**T

**A**T**AV** = **VS**2

**References**

* Alter O, Brown PO, Botstein D. (2000) Singular value decomposition for genome-wide expression data processing and modeling. *Proc Natl Acad Sci U S A*, **97**, 10101-6.
* Golub, G.H., and Van Loan, C.F. (1989) Matrix Computations, 2nd ed. (Baltimore: Johns Hopkins University Press).
* Greenberg, M.  (2001) Differential equations & Linear algebra (Upper Saddle River, N.J. : Prentice Hall).
* Strang, G.  (1998) Introduction to linear algebra (Wellesley, MA : Wellesley-Cambridge Press).