SVD Report

Donal Loitam

July 27, 2022

Contents

1 Introduction

- The Singular Value Decomposition (SVD) of a matrix is a factorization of that matrix into three matrices -**U, S, and V**
- **S** is the diagonal matrix of singular values. Think of singular values as the importance values of different features in the matrix
- The **rank** of a matrix is a measure of the unique information stored in a matrix. Higher the rank, more the information
- **Eigenvectors** of a matrix are directions of maximum spread or variance of data

Before we jump into the algorithm of SVD, let us discuss the concept of the **Rank of a Matrix**

RANK OF A MATRIX: The rank of a matrix is the maximum number of linearly independent row (or column) vectors in the matrix. A vector **r** is said to be linearly independent of vectors **r1** and **r2** if it cannot be expressed as a linear combination of **r1** and **r2**.

$$r \neq a_1 r_1 + a_2 r_2 \tag{1}$$

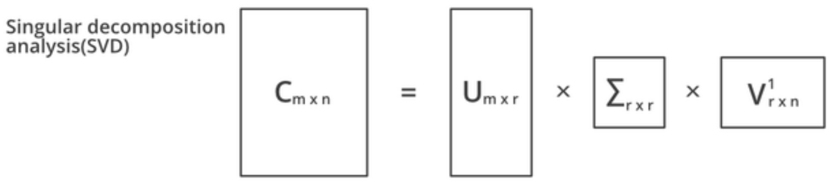
The rank of a matrix can be thought of as a representative of the amount of unique information represented by the matrix.

2 Key Points of the Algorithm

- Mathematics behind SVD :
  - The SVD of mxn matrix A is given by the formula :

$$A = UWV^T \tag{2}$$

where,
U: mxn matrix of the orthonormal eigenvectors of AA^T (**Left Singular Vectors**)
V^T: transpose of a nxn matrix containing the orthonormal eigenvectors of A^TA.
W: a nxn diagonal matrix of the **singular values (in decreasing order)** which are the square roots of the eigenvalues of A^TA . .



What are U and W and V (Given that U and V are orthogonal) ?

$$A = UWV^T \tag{3}$$
$$A^TA = (VW^TU^T)(UWV^T) \qquad [\because (AB)^T = B^TA^T] \tag{4}$$
$$= V(W^TW)V \qquad [\because U^TU = I] \tag{5}$$

'A' was a rectangular and completely general. But A^TA gives us a positive semidefinite matrix,and needless to say symmetric. It's **eigen vectors** should be orthogonal i.e V matrix and the **eigen-values** are positive and they're the **squares of the singular values** i.e

$$\lambda \text{ for } A^TA = \sigma^2 \text{ for A} \tag{6}$$

Similarly from AA^T , we can get U

- U is the eigenvectors for AA^T , and they have the same eigenvalues

- Why is SVD used in Dimensionality Reduction?
  - You might be wondering why we should go through with this seemingly painstaking decomposition. The reason can be understood by an alternate representation of the decomposition. See the fig below
  - The decomposition allows us to express our **original matrix as a linear combination of low-rank matrices.**
  - In a practical application, you will observe that only the first few, say k, singular values are large.The rest of the singular values approach zero. The second principal component is calculated in the same way, with the condition that it is uncorrelated with (i.e., perpendicular to) the first principal component and that it accounts for the next highest variance.

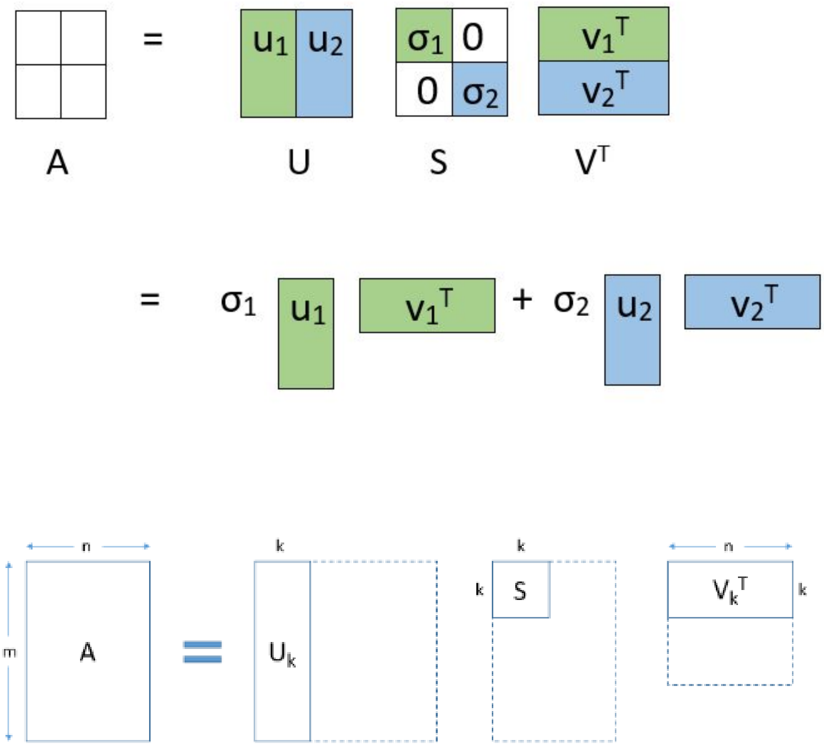


Figure 1: k-rank approximation of A

3 Some Questions

1. Explain the Curse of Dimensionality?
Ans.

- As the number of features increase, the number of samples increases, hence, the model becomes more complex
- The more the number of features, the more the chances of overfitting.

2. What are the pros and cons of Dimensionality Reduction ?
Ans. Some advantages for Dimensionality Reduction :-

- BETTER VISUALISATION : Dimensionality Reduction helps us visualize the data on 2D plots or 3D plots.
- REDUCED SPACE AND TIME COMPLEXITY : Fewer dimensions mean less computing. Less data means that algorithms train faster.

While some drawbacks are :-

- LESS INTERPRETABLE FEATURE : Transformed features are often hard to interpret
- Some information is lost, possibly degrading the performance of subsequent training algorithms.

3.
Ans :
4.
Ans.
5.
Ans.