

The Singular Value Decomposition

Rob McCulloch

1. Singular Value Decomposition
2. Column space, Row space, and rank
3. Linear is just a bunch of linear
4. Reduced Form
5. SVD and Least Squares
6. SVD and Spectral
7. Condition Number of a Matrix
8. Moore Penrose Generalized Inverse
9. Matrix Approximation

1. Singular Value Decomposition

This is a key decomposition that applies to *any* matrix A , $m \times n$.

2. Column space, Row space, and rank

We will see that the first r columns of U are an orthonormal basis for the column space of A .

We will see that the first r columns of V are an orthonormal basis for the row space of A .

Hence, the column rank = the row rank, which is then the rank.

So, r is the rank of the matrix.

Note:

A is $m \times n$, $A = [a_1, a_2, \dots, a_n]$, $a_i \in R^m$.

The column space is the span of the a_i which is the set $\{Ab, b \in R^n\}$.

Suppose B is $n \times n$ invertible.

Then

$$\{Ab, b \in R^n\} = \{ABb, b \in R^n\}$$

so that the column space of A is the same as the column space of AB .

Similar result for premultiplying by an invertible matrix for the row space.

Similarly, the first r columns of V are an orthonormal basis for the row space of A .

So, the row rank = the column rank = the rank, all which are equal to r in our notation.

The $i = r + 1, \dots, n$ columns of V form a basis for the subspace of R^n orthogonal to the row space.

3. Linear is just a bunch of linear

$$A_{m \times n} = U_{m \times m} \Sigma_{m \times n} V_{n \times n}^T$$

$$U = [u_1, u_2, \dots, u_m] \quad V = [v_1, v_2, \dots, v_n]$$

$$A v_j = U \Sigma V^T v_j = U \Sigma \begin{bmatrix} \langle v_1, v_j \rangle \\ \langle v_2, v_j \rangle \\ \vdots \\ \langle v_n, v_j \rangle \end{bmatrix}$$
$$= U \Sigma e_j$$

$$j > r \Rightarrow \Sigma e_j = 0$$
$$\Rightarrow A v_j = 0$$

$$1 \leq j \leq r \quad \Sigma e_j =$$

$$A v_j = U \Sigma e_j = \sigma_{jj} u_j$$

$$\begin{bmatrix} 0 & \dots & 0 \\ \vdots & \vdots & \vdots \\ 0 & \dots & 0 \end{bmatrix} \text{ - } j\text{th}$$

A truly remarkable result !!!

$$Av_j = \sigma_{jj} u_j, \quad 1 \leq j \leq r,$$

$$Av_j = 0, \quad (r + 1) \leq j \leq n.$$

$$A : R^n \rightarrow R^m.$$

- ▶ $N(A) = \{x \in R^n, \text{ s.t. } Ax = 0\}$, a subspace of dim $n - r$ with orthonormal basis $\{v_{r+1}, \dots, v_n\}$.
- ▶ $R(A) = \{Ax, x \in R^n\}$, a subspace of dim r with orthonormal basis $\{u_1, u_2, \dots, u_r\}$.

$$A: \mathbb{R}^n \rightarrow \mathbb{R}^m \quad A \text{ is } m \times n$$

$$x \in \mathbb{R}^n \quad x = \sum_{j=1}^n \tilde{x}_j v_j \quad (\tilde{x}_j = \langle v_j, x \rangle)$$

$$A(x) = \sum_{j=1}^n \tilde{x}_j A v_j = \sum_{j=1}^n \tilde{x}_j v_{j,j} u_j$$

$$y = Ax = \sum_{j=1}^m \tilde{y}_j u_j \quad (\tilde{y}_j = \langle u_j, y \rangle)$$

In terms of the orthonormal bases
 $\{u_j\}_{j=1}^m$ $\{v_j\}_{j=1}^n$

$$\tilde{x} = \begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \\ \vdots \\ \tilde{x}_n \end{bmatrix} \xrightarrow{A} \begin{bmatrix} v_{1,1} \tilde{x}_1 \\ v_{2,2} \tilde{x}_2 \\ \vdots \\ v_{n,n} \tilde{x}_n \\ 0 \dots 0 \end{bmatrix} = \begin{bmatrix} v_{1,1} \tilde{x}_1 \\ \vdots \\ v_{n,n} \tilde{x}_n \\ \tilde{y}_m \end{bmatrix}$$

So, for a linear $R \Rightarrow R$ we have the simple form:

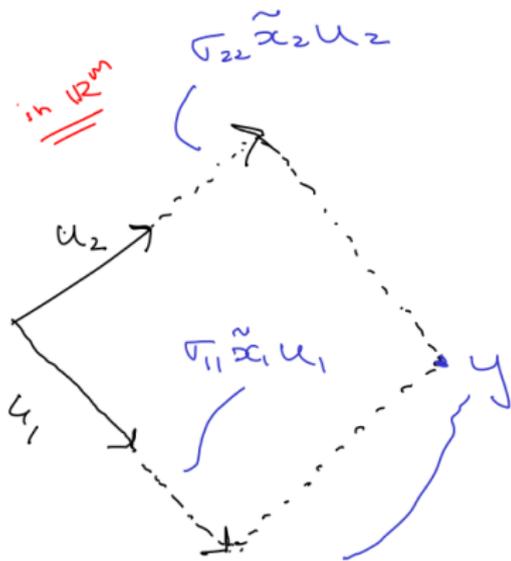
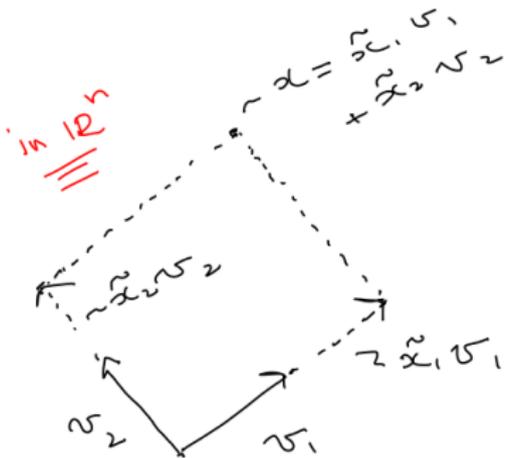
$$y = ax$$

where $A = [a]$, 1×1 .

In general, after you rotate to certain orthogonal bases, a rank r linear transformation $R^n \Rightarrow R^m$ is just the simple one r times.

$$\tilde{y}_i = \sigma_{ii} \tilde{x}_i, \quad i = 1, 2, \dots, r.$$

$$r = 2.$$



$$A: \mathbb{R}^n \rightarrow \mathbb{R}^m$$

$$\dim(\text{R}(A)) = r$$

$$\dim(\text{N}(A)) = n - r$$

$$y = \sigma_{11} \tilde{x}_1 u_1 + \sigma_{22} \tilde{x}_2 u_2$$

$$= \tilde{y}_1 u_1 + \tilde{y}_2 u_2$$

Note:

U orthogonal.

$$1 = |I| = |U'U| = |U'| |U| = |U|^2 \Rightarrow |U| = \pm 1.$$

Note:

A , $m \times m$ square of full rank so $r = m$.

Obviously, $\tilde{x}_i \rightarrow \sigma_{ii} \tilde{x}_i$ changes the volume by $\prod_{i=1}^m \sigma_{ii}$.

$$|A| = |U| |\Sigma| |V'| = (\pm 1) |\Sigma| = (\pm 1) \prod_{i=1}^m \sigma_{ii}.$$

Note:

Inverse of $\tilde{x}_i \rightarrow \tilde{y}_i = \sigma_{ii} \tilde{x}_i$ is

$$\tilde{y}_i \rightarrow \tilde{x}_i = \frac{1}{\sigma_{ii}} \tilde{y}_i$$

which is exactly $V \Sigma^{-1} U'$.

4. Reduced Form

You can simplify the construction to the “reduced form” by getting rid of the some zeros in Σ and corresponding columns in U and/or V .

Consider the case where $m > n$ and the rank is n so that the columns of A , $m \times n$ are linearly independent.

$$r = n.$$

Full SVD

$$\begin{aligned}
 \begin{bmatrix} A \\ m \times n \end{bmatrix} &= \begin{bmatrix} u_1 \\ \vdots \\ u_2 \\ \vdots \\ m \times (m-n) \end{bmatrix} \begin{bmatrix} \tilde{\Sigma} \\ \text{O} \\ n \times n \end{bmatrix} \begin{bmatrix} V^T \\ n \times n \end{bmatrix} \\
 &= \begin{bmatrix} u_1 \\ \vdots \\ m \times n \end{bmatrix} \begin{bmatrix} \tilde{\Sigma} \\ n \times n \end{bmatrix} \begin{bmatrix} V^T \\ n \times n \end{bmatrix}
 \end{aligned}$$

Reduced SVD

$$\begin{aligned}
 AV &= u_1 \tilde{\Sigma} \\
 AV_j &= \sigma_{jj} u_j \\
 j &= 1, 2, \dots, n
 \end{aligned}$$

U_2 is just an orthonormal basis for $R(A)^\perp$, you don't need it.

In general we have:

$$A_{m \times n} = \underbrace{[u_1, u_2]}_{\substack{\{ \\ m \times r \quad m \times (m-r)}}} \begin{bmatrix} \tilde{\Sigma} & 0 \\ 0 & 0 \end{bmatrix} \underbrace{\begin{bmatrix} v_1^T \\ v_2^T \end{bmatrix}}_{r \times n}$$
$$= u_1 \tilde{\Sigma} v_1^T$$

Columns of U_1 are an orthonormal basis for the column space of A .

Columns of V_1 are an orthonormal basis for the row space of A .

5. SVD and Least Squares

Let's see how the SVD decomposition can be used to compute the least squares solution.

Let's assume that X , $n \times p$ is of full rank p , where of course,

$$y = X\beta + \epsilon$$

is our model.

We simplify the SVD by using the reduced form.

$$X = U_1 \underbrace{\tilde{\Sigma}}_{2^{p \times p}} \underbrace{V^T}_{p \times p}$$

$\begin{matrix} n \times p & & \end{matrix}$

$$\tilde{\Sigma} = \text{diag}(\sigma_{ii}) \quad i=1, 2, \dots, p$$

$$V \text{ orthogonal, } U_1^T U_1 = I_p$$

$$(X^T X) = V \tilde{\Sigma} U_1^T U_1 \tilde{\Sigma} V^T = V \tilde{\Sigma}^2 V^T$$

$$(X^T X)^{-1} = V \tilde{\Sigma}^{-2} V^T$$

$$X^T y = V \tilde{\Sigma} U_1^T y$$

$$\begin{aligned} (X^T X)^{-1} X^T y &= [V \tilde{\Sigma}^{-2} V^T] [V \tilde{\Sigma} U_1^T y] \\ &= V \tilde{\Sigma}^{-1} U_1^T y = \sum_{i=1}^p v_i \frac{\langle u_i, y \rangle}{\sigma_{ii}} \end{aligned}$$

This just says:

Want to solve $y \approx Xb$

First replace y with $\hat{y} = \sum_{i=1}^p \langle y, u_i \rangle u_i$
 $= \sum_{i=1}^p \tilde{y}_i u_i$

$$b = \sum_{i=1}^p \tilde{x}_i v_i$$

So we have to solve: $\tilde{y}_i = \tilde{x}_i \sigma_{ii}$
for \tilde{x}_i given $\tilde{y}_i = \langle y, u_i \rangle$

$$\Rightarrow \tilde{x}_i = \frac{\tilde{y}_i}{\sigma_{ii}} \Rightarrow x = \sum_{i=1}^p \frac{\langle y, u_i \rangle}{\sigma_{ii}} v_i$$

6. SVD and Spectral

$$A, m \times n. A = U\Sigma V'.$$

$$A'A = [V\Sigma'U'] [U\Sigma V'] = V\Sigma'\Sigma V'$$

$$\begin{aligned} \Sigma^T \Sigma &= \begin{bmatrix} \Sigma^2 & 0 \\ 0 & 0 \end{bmatrix}_{n \times n} \begin{bmatrix} \Sigma^2 & 0 \\ 0 & 0 \end{bmatrix}_{m \times n} \\ &= \begin{bmatrix} \Sigma^2 & 0 \\ 0 & 0 \end{bmatrix}_{n \times n} \equiv \Sigma_n^2 \end{aligned}$$

$$\text{So, } A'A = V\Sigma_n^2 V'.$$

$$\text{Similarly, } AA' = U\Sigma_m^2 U'.$$

:

In our our svd, the rank of A , $m \times n$ is the number of non-zero diagonals of Σ which is r in our basic notation.

For a symmetric matrix S , the rank is the number of non-zero eigen values which is the number of non-zero elements if the diagonal matrix D in $S = PDP'$.

So, from the previous slide we have that the rank of A is the same as the rank of $A'A$ and AA' .

Of course, the rank of $X'X$ is relevant.

7. Condition Number of a Matrix

If the columns (or rows) of a matrix A are linearly dependent, then it can cause a problem, depending on what you want to do.

In linear regression, if X is the design matrix, then if the columns are linearly dependent, you cannot invert $X'X$.

More generally, if the columns are *close* to being linearly dependent then computation will become numerically unstable. That is, if some of the σ_{jj} are close to 0 for $j \in 1, 2, \dots, r$ this can cause trouble.

We saw that computing the coefficients for the projection on the the column space involved $1/\sigma_{jj}$ so you can see if these are very small, we have trouble.

Suppose $X, n \times p$ is of full column rank so that $p = r$.

Then,

$$X = U_1 \tilde{\Sigma} V'$$

as we discussed above when we looked at the reduced form.

Here, U_1 is $n \times p$, $\tilde{\Sigma}$ is $p \times p$, and V is $p \times p$.

The diagonals of $\tilde{\Sigma}$ are $\sigma_{jj}, j = 1, 2, \dots, p, \sigma_j > \sigma_{j+1, j+1} > 0$.

The degree to which ill-conditioning prevents a matrix from being inverted accurately depends on the ratio of its largest to smallest singular value, a quantity known as the condition number: which is

$$\text{Condition number} = \frac{\sigma_{11}}{\sigma_{pp}}$$

8. Moore Penrose Generalized Inverse

In solving the least squares problem, we have generally assumed that the design matrix X , $n \times p$ is of full rank p .

If X is not of full rank then there are many solutions to

$$\min_b \|y - Xb\|^2$$

The Moore Penrose inverse chooses a solution for us.

Suppose we want to solve

$$y = Xb \quad \text{for } b$$

given y and X .

If X is $n \times n$ full rank

$$\hat{b}^* = X^{-1}y \quad \text{is an exact solution.}$$

If X is $n \times p$ full rank (p) then

$$\hat{b} = (X^T X)^{-1} X^T y$$

is the closest we can get in that

$$X\hat{b} \approx y.$$

If X is not of full rank then
 $\exists \alpha_i, i=1, 2, \dots, p$, not all 0
such that $\sum_{i=1}^p \alpha_i \alpha_i = 0$.

$$\text{or } X\alpha = 0$$

$$\begin{aligned} \text{thus } X(b + c\alpha) &= Xb + cX\alpha \\ &= Xb \end{aligned}$$

So if b^* is a solution to

$$\min_b \|y - Xb\|^2 \text{ so is } b^* + c\alpha$$

— there is not a unique solution.

Note: Same for any $\alpha \in N(X)$.

Let $X = U_1 \tilde{\Sigma} V_1^T$
- the reduced form
SVD of X

$$\text{Let } X^+ \equiv V_1 \tilde{\Sigma}^{-1} U_1^T$$

the Moore-Penrose generalized
inverse of X .

Claim $b^0 = X^+ y$ is a solution to
$$\min_b \|y - Xb\|^2$$

Note: U_1 is $n \times r$, $\tilde{\Sigma}$ is $r \times r$, V_1 is $p \times r$.

The columns of U_1 are an orthonormal basis for the column space of X we like to project y onto.

The columns of V_1 are an orthonormal basis for the space of coefficient vectors that do NOT map to 0.

Have to check

$$X^T (y - Xb^0) = 0 \quad \text{or} \quad X^T y = X^T X b^0$$

$$X = U, \tilde{\Sigma} V,^T \quad X^T = V, \tilde{\Sigma} U,^T$$

$$X^+ = V, \tilde{\Sigma}^{-1} U,^T \quad b^0 = X^+ y$$

$$X^T X b^0 = X^T X X^+ y$$

$$\begin{aligned} X^T X X^+ &= [V, \tilde{\Sigma} U,^T] [U, \tilde{\Sigma} V,^T] [V, \tilde{\Sigma}^{-1} U,^T] \\ &= V, \tilde{\Sigma} U,^T = X^T \end{aligned}$$

$$\text{so } X^T X b^0 = X^T X X^+ y = X^T y$$

Clearly, $XX^+ y = Xb^0$ projects y onto the column space of X .

$$X^+ = V_1 \tilde{\Sigma}^{-1} U_1^T$$

$$\begin{aligned} X^+ y &= V_1 \tilde{\Sigma}^{-1} \begin{bmatrix} \langle u_1, y \rangle \\ \langle u_2, y \rangle \\ \vdots \\ \langle u_r, y \rangle \end{bmatrix} \\ &= V_1 \begin{bmatrix} \langle u_1, y \rangle / \sigma_{11} \\ \langle u_2, y \rangle / \sigma_{22} \\ \vdots \\ \langle u_r, y \rangle / \sigma_{rr} \end{bmatrix} \\ &= \sum_{i=1}^r v_i \frac{\langle u_i, y \rangle}{\sigma_{ii}} \end{aligned}$$

- ▶ $U_1^T y$ = coefficients for $\hat{y} = P_V y$ in terms of U_1 orthonormal basis.
- ▶ $\tilde{\Sigma}^{-1} U_1^T y$ = coefficients of vector that goes to \hat{y} in V_1 orthonormal basis.
- ▶ $V_1 \tilde{\Sigma}^{-1} U_1^T y$ = vector in span of V_1 that goes to \hat{y} .

Generalized inverse just picks a vector in the row space that goes to \hat{y} .

XX^+ projects onto the column space of X .

$$X = U_1 \tilde{\Sigma} V_1', \quad X^+ = V_1 \tilde{\Sigma}^{-1} U_1'$$

$$XX^+ = [U_1 \tilde{\Sigma} V_1'] [V_1 \tilde{\Sigma}^{-1} U_1] = U_1 U_1'$$

X^+X projects onto the row space of X .

$$X^+X = [V_1 \tilde{\Sigma}^{-1} U_1'] [U_1 \tilde{\Sigma} V_1'] = V_1 V_1'$$

$X^+ X$ projects b onto the row space of X . gives us a characterization of the MP choice of solution.

$V = \{v_1, v_2\}$ ON basis for subspace
 orthogonal to the row space
 $\{v_1, v_2\}$ ON basis for row space of X

Any b in \mathbb{R}^p can be written

$$b = v_1 \alpha + v_2 \beta$$

$$\begin{aligned} X^+ X b &= [v_1, \tilde{\Sigma}^{-1} u_1^T] [u_1, \tilde{\Sigma} v_1^T] [v_1 \alpha + v_2 \beta] \\ &= v_1 \alpha \end{aligned}$$

$X^+ X$ is a projection onto the row space of X .

The column space and row space of X have the same dimension so we can define a 1-1 map between them.

Everything else gets projected away.

9. Matrix Approximation

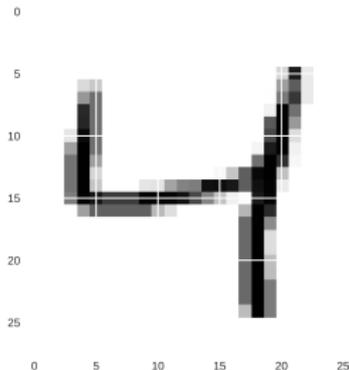
Suppose $\sigma_{11} \geq \sigma_{22} \geq \dots \geq \sigma_{rr}$ and after s they are small,
 $\sigma_{ij} \approx 0, i > s$.

$$\begin{aligned} A_{m \times n} &= U_1 \tilde{\Sigma} V_1^T \\ &= \{u_1, u_2, \dots, u_r\} \begin{bmatrix} \sigma_{11} & & & \\ & \sigma_{22} & & \\ & & \ddots & \\ & & & \sigma_{rr} \end{bmatrix} \begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_r^T \end{bmatrix} \\ &= \{u_1, u_2, \dots, u_r\} \begin{bmatrix} \sigma_{11} v_1^T \\ \sigma_{22} v_2^T \\ \vdots \\ \sigma_{rr} v_r^T \end{bmatrix} = \sum_{i=1}^r \sigma_{ii} \underbrace{u_i}_{m \times 1} \underbrace{v_i^T}_{1 \times n} \\ &\approx \sum_{i=1}^s \sigma_{ii} u_i v_i^T = \{u_1, u_2, \dots, u_s\} \begin{bmatrix} \sigma_{11} v_1^T \\ \sigma_{22} v_2^T \\ \vdots \\ \sigma_{ss} v_s^T \end{bmatrix} \\ &= \hat{U} \hat{V}^T \end{aligned}$$

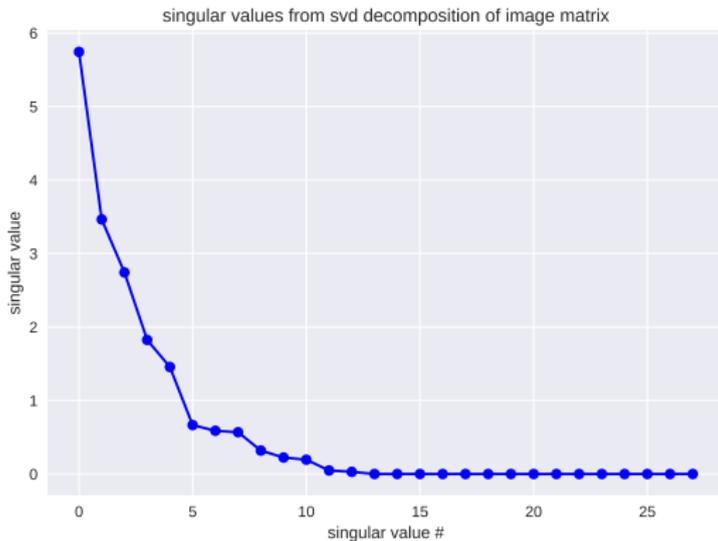
Example:

Here is a 28x28 grey scale image of a digit.

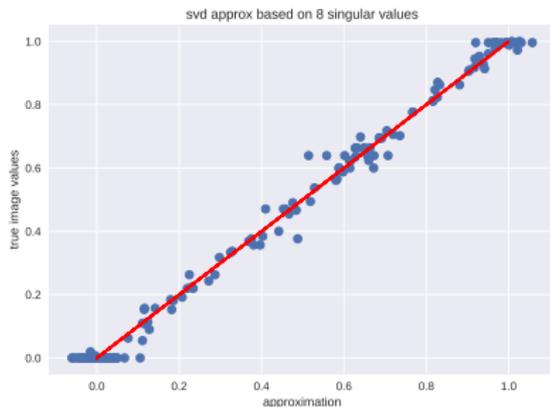
The (i,j) element of the matrix is $0:255$ indicating the grayscale.
I divided by 255.



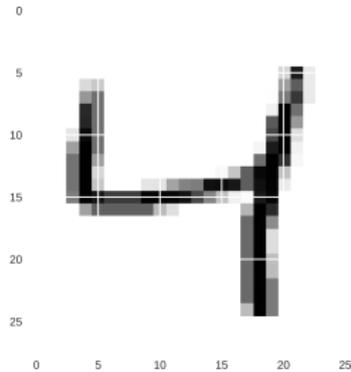
Here are the singular values from the 28x28 matrix.
This is called a scree plot.



Here is are the matrix values for the image plotted against the values from the approximation using 8 singular values.



Here is a 28x28 grey scale image of a digit.



and the approx:

