Math 415 - Lecture 37

Singular Value Decomposition

Friday December 4th 2015

Textbook reading: Chapter 6.3

Suggested practice exercises: Chapter 6.3, # 1, 2, 3, 5, 8, 9, 15

Strang lecture: Lecture 29: Singular Value Decomposition

1 Review

- Spectral theorem: If A is an $n \times n$ symmetric matrix, then it has an orthonormal basis of eigenvectors $\mathbf{v}_1 \dots \mathbf{v}_n$, and all eigenvalues are real.
- We can write

$$A = \underbrace{\begin{bmatrix} | & & | \\ \mathbf{v}_1 & \dots & \mathbf{v}_n \\ | & & | \end{bmatrix}}_{\text{matrix of eigenvectors}} \underbrace{\begin{bmatrix} \lambda_1 & 0 & & 0 \\ 0 & \lambda_2 & & 0 \\ & & \ddots & \\ 0 & 0 & & \lambda_n \end{bmatrix}}_{\text{matrix of eigenvalues}} \begin{bmatrix} - & \mathbf{v}_1^T & - \\ & \vdots & \\ - & \mathbf{v}_n^T & - \end{bmatrix}$$

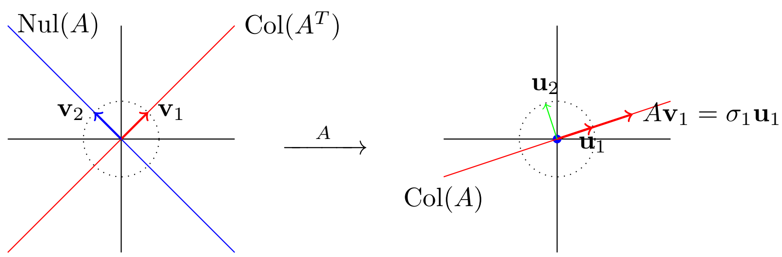
- Today: There is a similar decomposition for any $m \times n$ matrix A.
  - Doesn't even have to be square!
  - The price we pay: different bases on the left and right sides.

2 Singular Value Decomposition

2.1 Goals

Starting with an $m \times n$ matrix A we want to

- Describe the *geometry* of the corresponding map $\mathbb{R}^n \to \mathbb{R}^m$,
- Find a way to *approximate* A by simpler matrices, that are easier/cheaper to calculate with.



How?Remember: for each A we get 4 subspaces

- Input space $\mathbb{R}^n$ contains row space Col(A^T) and Null space Nul(A). Dimensions are r and $n - r$.
- Output space $\mathbb{R}^m$ contains columns space Col(A) and left null space Nul(A^T). Dimensions are r and $m - r$.

2.2 Idea

Idea:
Find an orthonormal basis in each of the input space subspaces, and watch what happens to these basis vectors if we multiply by A.

Choose a basis $\mathbf{v_1}, \ldots, \mathbf{v_r}$ of the row space $Col(A^T)$, and a basis $\mathbf{v_{r+1}}, \ldots, \mathbf{v_n}$ of the null space Nul(A). Then think of what happens when we apply A to each of the basis vectors:

- $A\mathbf{v_{r+1}} = 0 = \cdots = A\mathbf{v_n}$, since each vector belongs to Nul(A).
- The other vectors $A\mathbf{v_i}$, $i = 1, 2, \ldots, r$, will all be nonzero, in fact will be give a basis of Col(A)!
- Rescale these basis vectors to get unit vectors $\mathbf{u_1}$,$\mathbf{u_2}, \ldots, \mathbf{u_r}$. By a miracle they turn out to be orthogonal, if we choose the $\mathbf{v}_1, \mathbf{v}_2, \ldots$ in the right way.
- We get $A\mathbf{v_i} = \sigma_i \mathbf{u_i}$ for $i = 1, 2, \ldots, r$. The stretch factors $\sigma_i > 0$, $i = 1, 2, \ldots, r$ are called the *Singular Values* of A
- Extend the $\mathbf{u_i}$ basis of Col(A) to a basis $\mathbf{u_{r+1}}, \ldots, \mathbf{u_m}$ of the output space.

2.3 What is SVD?

Motto
In Linear Algebra everything is a matrix factorization.

The complicated story with orthonormal basis and singular values for A gives a factorization, called Singular Value Decomposition:

$$A = \underbrace{\begin{bmatrix} | & & | \\ \mathbf{u}_1 & \dots & \mathbf{u}_m \\ | & & | \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} \sigma_1 & 0 & \\ 0 & \sigma_2 & \\ & & \ddots \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} - & \mathbf{v}_1^T & - \\ & \vdots & \\ - & \mathbf{v}_n^T & - \end{bmatrix}}_{V^T}$$

- $A = U\Sigma V^T$. This is just $Av_i = \sigma_i u_i$ rearranged in matrix form.
- U, V are orthogonal. We need to choose the input basis v_i carefully in order for the output basis u_i to be orthonormal.
- Columns of U are an orthonormal basis for $\mathbb{R}^m$. U is $m \times m$.
- Rows of V are an orthonormal basis for $\mathbb{R}^n$. V is $n \times n$.
- Σ is rectangular $m \times n$ and diagonal, the r non zero diagonal entries are called **singular values**, they are positive.

2.4 How to Compute SVD

Here is a recipe for computing SVD:

Compute $A^T A$. This is a symmetric matrix!! (Why?)

Make V:
- Find orthonormal eigenvectors $\mathbf{v_1}, \dots, \mathbf{v_n}$ of $A^T A$. (Why can we do this?)
- Magic: The eigenvalues are always positive or zero! $\lambda_1 \geq \cdots \geq \lambda_r > 0$, $\lambda_{r+1} = 0 = \cdots = \lambda_n$.
- Order $\mathbf{v_1}, \dots, \mathbf{v_n}$ according to the size of their eigenvalues.
- Put $\mathbf{v}_1, \dots, \mathbf{v}_n$ into matrix V,

Make Σ: Set $\sigma_i = \sqrt{\lambda_i}$ for $i = 1 \dots r$ and $\sigma_{r+1} = \cdots = \sigma_n = 0$. Put these into diagonal of **rectangular** $m \times n$ matrix Σ.

Make U:
- Set $\mathbf{u_1} = \frac{1}{\sigma_1} A\mathbf{v_1}$, ..., $\mathbf{u_r} = \frac{1}{\sigma_r} A\mathbf{v_r}$.
- Magic: The $\mathbf{u}_1, \dots, \mathbf{u}_r$ are orthogonal
- Extend $\mathbf{u_1}, \dots, \mathbf{u_r}$ to an orthonormal basis $\mathbf{u_1}, \dots, \mathbf{u_m}$ for $\mathbb{R}^m$.
- Put $\mathbf{u_1}, \dots, \mathbf{u_m}$ into matrix U.

Now you have $A = U\Sigma V^T$!

Example 1. Compute the SVD of

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

Solution. Compute $A^T A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

Make V: Basis of eigenvectors for A^TA: $\mathbf{v_1} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\mathbf{v_2} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

$$V = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Make Σ: Eigenvalues are 1 and 1. So, $\sigma_1 = \sigma_2 = \sqrt{1} = 1$.

$\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

Make U: $\mathbf{u}_1 = \frac{1}{\sigma_1} A\mathbf{v_1} = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$ and $\mathbf{u}_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$.

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

We get

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

Note: $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ has no real eigenvalues. It's not diagonalizable with real matrices! But, it has an SVD! Wolfram Alpha

Example 2. Compute the SVD of

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

Solution. Compute $A^TA = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}$.

Make V: Basis of eigenvectors for A^TA:

$$\mathbf{v}_1 = \begin{bmatrix} \frac{1}{\sqrt{6}} \\ -\frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix}$$

with eigenvalues $\lambda_1 = 3, \lambda_2 = 1, \lambda_3 = 0$.

Make Σ: Eigenvalues were $\lambda_1 = 3, \lambda_2 = 1, \lambda_3 = 0$, so $\sigma_1 = \sqrt{3}, \sigma_2 = 1$.

Make U: $\mathbf{u}_1 = \frac{1}{\sigma_1} A\mathbf{v}_1 = \frac{1}{\sqrt{3}} \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{6}} \\ -\frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{bmatrix} = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$ $\mathbf{u_2} = \frac{1}{\sigma_2} A\mathbf{v}_2 =$ $\frac{1}{1} \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$

Final result:

$$\begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} = \begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix}$$

Notice how A behaves in the basis $\{\mathbf{u}_1, \mathbf{u}_2\}$ and $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$:

$A\mathbf{v}_1 = \sigma_1 \mathbf{u}_1$:

$$\begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{6}} \\ -\frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{bmatrix} = \sqrt{3} \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

$A\mathbf{v}_2 = \sigma_2 \mathbf{u}_2$:

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

$A\mathbf{v}_3 = 0$.

A matrix might not be diagonalizable:

- If A is rectangular, it does not even have eigenvalues.

But A will always have an SVD! This comes at a cost:

- The SVD is not unique.
- The singular values σ_i are not eigenvalues.

Note the difference: for $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ the eigenvalues are $\lambda = i, -i$ but the singular values are $\sigma = 1, 1$.

2.5 Approximation

* To calculate matrix product AB we can use the **ROW** times **COLUMN** method: the ij component is the product $R_i B_j$, where R_i is row i of A and B_j is the jth column of B.

* What about **COLUMN** times **ROW**?

*

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} a+3b & 2a+4b \\ c+3d & 2c+4d \end{bmatrix} =$$

$$= \begin{bmatrix} a \\ c \end{bmatrix} \begin{bmatrix} 1 & 2 \end{bmatrix} + \begin{bmatrix} b \\ d \end{bmatrix} \begin{bmatrix} 3 & 4 \end{bmatrix}$$

* This works for any matrix multiplication: AB is a sum of **COLUMN** times **ROW** matrices.

It turns out we can write A as a sum:

$$A = \underbrace{\begin{bmatrix} | & & | \\ \mathbf{u}_1 & \dots & \mathbf{u}_m \\ | & & | \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} \sigma_1 & 0 & \\ 0 & \sigma_2 & \\ & & \ddots \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} - & \mathbf{v}_1^T & - \\ & \vdots & \\ - & \mathbf{v}_n^T & - \end{bmatrix}}_{V^T}$$

$$= \sigma_1 \mathbf{u_1} \mathbf{v_1}^T + \sigma_2 \mathbf{u_2} \mathbf{v_2}^T + \cdots + \sigma_r \mathbf{u_r} \mathbf{v_r}^T$$

(Sanity check: An $m \times 1$ column vector times a $1 \times n$ row vector is an $m \times n$ matrix.)

Idea. We can get a good approximation to A by taking the entries of the sum with the largest singular values! We'll see this when we talk about image compression later.

Example 3. If $\mathbf{u}, \mathbf{v}$ are non-zero, then the matrix $\mathbf{u}\mathbf{v}^T$ has rank 1. Why?

Solution. Let $\mathbf{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$. Then

$$\mathbf{u}\mathbf{v}^T = \mathbf{u} \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}^T = \mathbf{u} \begin{bmatrix} v_1 & \dots & v_n \end{bmatrix} = \begin{bmatrix} v_1 \mathbf{u} & \dots v_n \mathbf{u} \end{bmatrix}.$$

Example 4. Use

$$\begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} = \begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix}$$

to write $\begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}$ as a sum of rank 1 matrices.

Solution.

$$\begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} = \sqrt{3} \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{bmatrix} + 1 \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{bmatrix}.$$

2.6 SVD and the Four Fundamental Subspaces

The SVD of A gives orthonormal bases for all four fundamental subspaces of A.
Given $\{\mathbf{u_1}, \dots, \mathbf{u_m}\}$ and $\{\mathbf{v_1}, \dots, \mathbf{v_n}\}$,

- $Col(A^T) = \text{Span}\{\mathbf{v_1}, \dots, \mathbf{v_r}\}$
- $Nul(A) = \text{Span}\{\mathbf{v_{r+1}}, \dots, \mathbf{v_n}\}$
- $Col(A) = \text{Span}\{\mathbf{u_1}, \dots, \mathbf{u_r}\}$
- $Nul(A^T) = \text{Span}\{\mathbf{u_{r+1}}, \dots, \mathbf{u_m}\}$

2.7 Practice Questions

Example 5. Suppose A is an invertible square matrix. Find a singular value decomposition of A^{-1}.

Example 6. If A is a square matrix, then $|\det(A)|$ is the product of the singular values of A. Why?

Example 7. Find the singular value decomposition of $A = \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{bmatrix}$.