Lecture 2:
The Singular Value Decomposition

Nando de Freitas
www.cs.ubc.ca/~nando/340-2009/
September 2009

Outline

The Singular Value Decomposition (SVD) is a matrix factorization that has many applications, including:

- Information retrieval,
- Least-squares problems,
- Image processing,
- Dimensionality reduction.

Eigenvalue decomposition

Let $\mathbf{A} \in \mathbb{R}^{m \times m}$. If we put the eigenvalues of $\mathbf{A}$ into a diagonal matrix $\mathbf{\Lambda}$ and gather the eigenvectors into a matrix $\mathbf{X}$, then the eigenvalue decomposition of $\mathbf{A}$ is given by

$$\mathbf{A} = \mathbf{X}\mathbf{\Lambda}\mathbf{X}^{-1}.$$

But what if $\mathbf{A}$ is not a square matrix? Then the SVD comes to the rescue.

EXAM Exercise, if $A \in \mathbb{R}^{m \times m}$, symmetric show that $A = X \Lambda X^T$

Formal definition of the SVD

Given $\mathbf{A} \in \mathbb{R}^{m \times n}$, the SVD of $\mathbf{A}$ is a factorization of the form

$$\boxed{\underset{m \times n}{\mathbf{A}} = \underset{m \times n}{\mathbf{U}}\,\underset{n \times n}{\mathbf{\Sigma}}\,\underset{n \times n}{\mathbf{V}^T}}$$

where $\mathbf{u}$ are the left **singular vectors**, σ are the **singular values** and $\mathbf{v}$ are the right singular vectors.

$\mathbf{\Sigma} \in \mathbb{R}^{n \times n}$ is diagonal with positive entries (singular values in the diagonal).

$\mathbf{U} \in \mathbb{R}^{m \times n}$ with orthonormal columns. $U^T U = I$

$\mathbf{V} \in \mathbb{R}^{n \times n}$ with orthonormal columns.

($\Rightarrow \mathbf{V}$ is orthogonal so $\mathbf{V}^{-1} = \mathbf{V}^T$)

$$\begin{cases} V^T V = I \\ V V^T = I \end{cases}$$

Formal definition of the SVD

The equations relating the right singular values $\{\mathbf{v}_j\}$ and the left singular vectors $\{\mathbf{u}_j\}$ are

$$\mathbf{A}\mathbf{v}_j = \sigma_j \mathbf{u}_j \qquad j = 1, 2, \ldots, n$$

$$\mathbf{AV} = \mathbf{U\Sigma}$$

$$\underset{v_i \in \mathbb{R}^n}{\mathbf{A} \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \ldots & \mathbf{v}_n \end{bmatrix}} = \underset{u_i \in \mathbb{R}^m}{\begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \ldots & \mathbf{u}_n \end{bmatrix}} \begin{bmatrix} \sigma_1 & & & 0 \\ & \sigma_2 & & \\ 0 & & \ddots & \\ & & & \sigma_n \end{bmatrix}$$

SVD properties

1. There is no assumption that $m \geq n$ or that $\mathbf{A}$ has full rank.

2. All diagonal elements of $\mathbf{\Sigma}$ are non-negative and in non-increasing order:

$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_p \geq 0$$

where $p = \min(m, n)$

SVD properties

Theorem 3 *Every matrix* $\mathbf{A} \in \mathbb{R}^{m \times n}$ *has singular value decomposition* $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$

Furthermore, the singular values $\{\sigma_j\}$ *are uniquely determined.*

If $\mathbf{A}$ *is square and* $\sigma_i \neq \sigma_j$ *for all* $i \neq j$, *the left singular vectors* $\{\mathbf{u}_j\}$ *and the right singular vectors* $\{\mathbf{v}_j\}$ *are uniquely determined to within a factor of* ± 1.

SVD in terms of eigenvalues

$(A^TA)^T = A^TA = Q\Lambda Q^T$

Theorem 4 *The nonzero singular values of* $\mathbf{A}$ *are the (positive) square roots of the nonzero eigenvalues of* $\mathbf{A}^T\mathbf{A}$ *or* $\mathbf{A}\mathbf{A}^T$ *(these matrices have the same nonzero eigenvalues).*

① $A^TA = (U\Sigma V^T)^T U\Sigma V^T$

$= V\Sigma U^T(U\Sigma V^T)$

$\boxed{A^TA = V\Sigma^2V^T}$

$Q = V$

$\Sigma^2 = \Lambda$

② $AA^T = U\Sigma V^T V \Sigma U^T$
$= U\Sigma^2 U^T$

③ $A = U\Sigma V^T \Rightarrow AV = U\Sigma$
$AV\Sigma^{-1} = U$

① $\Sigma^2 = \Lambda$, $A^TA = V\Sigma^2 V^T$, $Q = V$

Let $A = [1 \quad 2 \quad 3]$, what is U, Σ, V

Matrix norms

Theorem 5 $\|\mathbf{A}\|_2 = \sigma_1$, *where* $\|\mathbf{A}\|_2 = \max_{\mathbf{x}\neq 0} \frac{\|\mathbf{A}\mathbf{x}\|}{\|\mathbf{x}\|} = \max_{\|\mathbf{x}\|\leq 1} \|\mathbf{A}\mathbf{x}\|$.

$\|x\| = 1$ is a circle in 2d
when $x = [x_1 \quad x_2]$.

$$\|A\| = \max_{\|x\|\neq 0} \frac{\|Ax\|}{\|x\|} = \max_{\|x\|=1} \|Ax\| \quad (*)$$

$$= \max_{\|x\|\neq 0} \frac{\|U\Sigma V^T x\|}{\|x\|} = \max_{\|x\|\neq 0} \frac{\|\Sigma V^T x\|}{\|x\|}$$

$$= \max_{\|x\|\neq 0} \frac{\|\Sigma V^T x\|}{\|V^T x\|} = \max_{\|y\|\neq 0} \frac{\|\Sigma y\|}{\|y\|}$$

$$= \max_{\|y\|=1} \|\Sigma y\| = \max_{\|y\|=1} \sqrt{(\sigma_1 y_1)^2 + (\sigma_2 y_2)^2 + \cdots}$$

Optimal $y = [1\ 0\ 0\ 0]^T = \sigma_1$

SVD in terms of rank 1 matrices

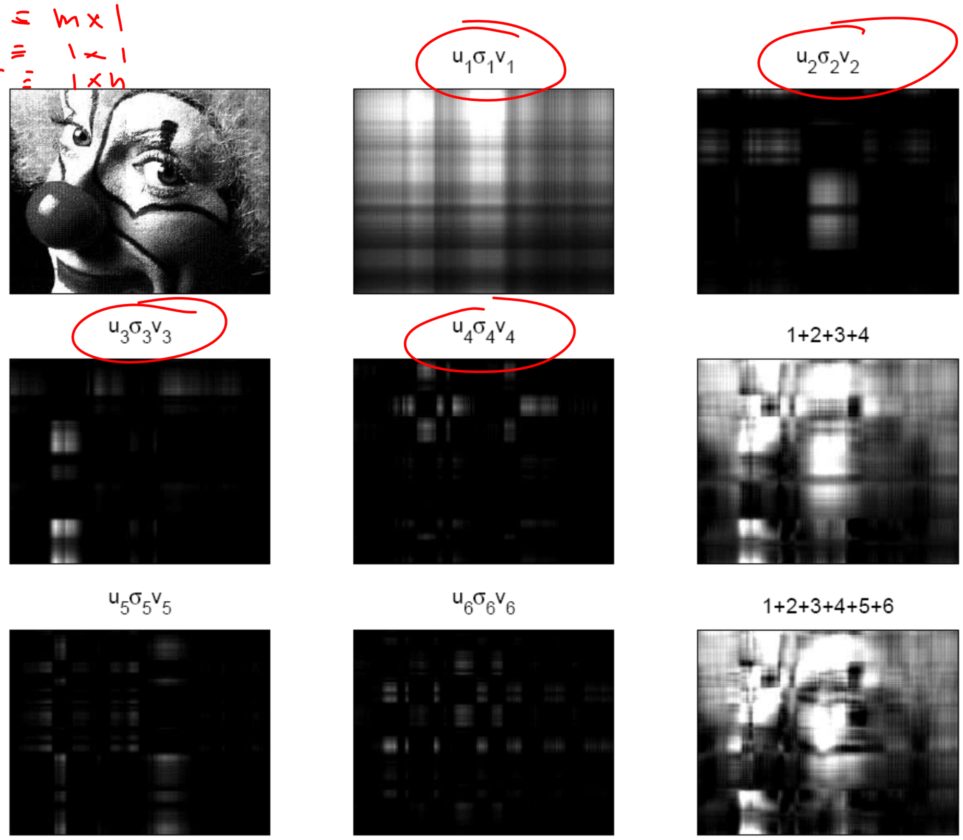
Theorem 6

$$\mathbf{A} = \sum_{j=1}^{r} \sigma_j \mathbf{u}_j \mathbf{v}_j^T$$

where r is the rank of $\mathbf{A}$.

$u = m \times 1$
$\sigma = 1 \times 1$
$v = 1 \times n$

$m \times n$
$A =$



Energy interpretation and approximation error

What is so useful about this expansion is that the ν^{th} *partial sum captures as much of the "energy" of* $\mathbf{A}$ *as possible by a matrix of at most rank-*ν. In this case, "energy" is defined by the 2-norm.

Theorem 7 *For any* ν *with* $0 \leq \nu \leq r$ *define*

$$\mathbf{A}_\nu = \sum_{j=1}^{\nu} \sigma_j \mathbf{u}_j \mathbf{v}_j^T$$

If $\nu = p = \min(m, n)$*, define* $\sigma_{\nu+1} = 0$.

Then,

$$\|\mathbf{A} - \mathbf{A}_\nu\|_2 = \sigma_{\nu+1}$$