More on SVD

Monday, October 10, 2005
11:04 AM

The basic idea of SVD is that, given any rectangular matrix A with dimensions $m \times n$,

$A = U \Sigma V^T$;
A: $m \times n$, $m \geq n$
Σ: $m \times n$
V: $n \times n$
$U^T U = I_m$, $V^T V = I_n$

$$\Sigma = \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_r \end{pmatrix}_{m \times n}$$

Rank(A) = r

Last time we looked at a matrix called $\tilde{A}$

$$\tilde{A} = \begin{pmatrix} 0 & A \\ A^T & 0 \end{pmatrix}$$ where $\lambda(\tilde{A}) = \pm\, \sigma(A)$

Observe that $\tilde{A}$ is symmetric so its eigenvectors $\boldsymbol{z}$ can be normalized

$$\|z_i\|_2 = 1$$

$\|z_i\|_2 = 1$

$\vec{z} = \begin{pmatrix} \vec{x} \\ \vec{y} \end{pmatrix}, \quad \vec{z} = \begin{pmatrix} \vec{x} \\ -\vec{y} \end{pmatrix}$ is also an eigenvector

Thus the mass is distributed over both parts of the eigenvector, i.e. $x^T x = 1/2$, $y^T y = 1/2$

Define a new matrix,

$$\tilde{Z} = \begin{pmatrix} x & x \\ y & -y \end{pmatrix} \frac{1}{\sqrt{2}} \quad \text{for} \quad \sigma \neq 0$$

Then we claim that $\tilde{A}$ can be written as

$$\tilde{A} = \tilde{Z} \Lambda \tilde{Z}^T = \begin{pmatrix} x & x \\ y & -y \end{pmatrix} \begin{pmatrix} \Sigma_r & 0 \\ 0 & -\Sigma_r \end{pmatrix} \tilde{Z}^T$$

$$\tilde{A} = \begin{pmatrix} 0 & A \\ A^T & 0 \end{pmatrix} = \begin{pmatrix} 0 & X\Sigma_r Y^T \\ Y\Sigma_r X^T & 0 \end{pmatrix}$$

$$\Rightarrow A = X\Sigma_r Y^T$$

Jordan Canonical Form

Monday, October 10, 2005
11:17 AM

JCF is more useful as an analytical tool than numerical

Every matrix A can be written as $A = QJQ^{-1}$ where

$$J = \begin{pmatrix} J_1 & & 0 \\ & \ddots & \\ 0 & & J_k \end{pmatrix}$$

Where

$$J_r = \begin{pmatrix} \lambda_r & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_r \end{pmatrix}$$

The rank of $J_r - \lambda_r I$?

$$(J_r - \lambda_r I)\vec{x} = 0 = 0$$

$$\begin{pmatrix} 0 & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & 0 \end{pmatrix} \vec{x} = 0$$

$$(J_r - \lambda_r I)e_1 = 0$$

e1 is an eigenvector of this matrix

JCF is a useful for computing powers of matrices:
$A^2 = QJQ^{-1}QJQ^{-1} = QJ^2Q^{-1}$ and $A^n = QJ^nQ^{-1}$

$$J_r^k = \begin{pmatrix} \lambda_r & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_r \end{pmatrix}^k = (\lambda_r I + K)^k$$

$$K = \begin{pmatrix} 0 & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & 0 \end{pmatrix}$$

So now, we want to look at

$$(\lambda_r I + K)^k = \sum_{j=0}^{k} \binom{k}{j} \lambda_r^{k-j} K^j$$

$$= \begin{pmatrix} \lambda_r^k & \binom{k}{1}\lambda_r^{k-1} & \text{—} & \binom{k}{n_r-1}\lambda_r^{k-(n_r-1)} \\ & \ddots & \ddots & \\ 0 & & & \lambda^k \end{pmatrix}$$

$$\begin{pmatrix} & & \ddots & \\ 0 & & & \lambda_r^k \end{pmatrix}$$

Jordon Canonical form

Example

Monday, October 10, 2005
11:25 AM

$$\begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix}^3 = \begin{pmatrix} \lambda^3 & 3\lambda^2 & 3\lambda \\ & \lambda^3 & 3\lambda^2 \\ 0 & & \lambda^3 \end{pmatrix}$$

This is important in doing powers of matrices; if we have

$$|\lambda| < 1$$

And raise that matrix to the *k*th power, we'd get

$$\begin{pmatrix} \lambda^k & k\lambda^{k-1} & \binom{k}{2}\lambda^{k-2} \\ & \ddots & \ddots & \ddots \\ 0 & & & \end{pmatrix}$$

Which approaches 0, but not in a regular way -- for example the terms $k\lambda^{-1}$ go up, then down

Differential Equation Example

Monday, October 10, 2005
11:28 AM

Functions of matrices:

$$\tilde{y}' = A\tilde{y}, \quad \tilde{y}(t_0) = y_0$$

$$= QJQ^{-1}\tilde{y}$$

$$Q^{-1}\tilde{y}' = JQ^{-1}\tilde{y}$$

$$\tilde{z} = Q^{-1}\tilde{y}$$ substitution

$$\tilde{z}' = J\tilde{z}$$

Say J is a diagonal matrix

$$J = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$

$$z_i' = \lambda_i z_i, \quad z_i(t) = e^{\lambda_i t} z_i(0)$$

$$\tilde{y}(t) = Q \begin{pmatrix} e^{\lambda_1 t} & & & 0 \\ & e^{\lambda_2 t} & & \\ & & \ddots & \\ 0 & & & e^{\lambda_n t} \end{pmatrix} Q^{-1} y(0)$$

$$= e^{At} y(0)$$

$$F(A) = Q F(J) Q^{-1}$$

$$\varphi(\lambda) = \alpha_0 + \alpha_1 \lambda + \cdots + \alpha_m \lambda^m$$

is called the characteristic equation; the Cayley-Hamilton theorem says that

$$\varphi(A) = \alpha_0 I + - + \alpha_m A^m = 0$$

Every matrix satisfies this!

Alternatively,

$$\alpha_0 A^{-1} + \alpha_1 I + \alpha_2 A + - + \alpha_m A^{n-1} = 0$$

$$A^{-1} = \beta_1 I + \beta_2 (A) + - + \beta_{n-1} A^{n-1}$$

$A^{-1} =$ Polynomial of degree *n-1* in *A*

The amazing thing is that this says that any analytic function *F(A)* can be expressed by polynomials (require "analytic" so that a power series construction exists)! That is, if

$$\alpha_1 I + \alpha_1 A + - + \alpha_m A^n = 0$$

Then A^n = poly(A) of degree n-1

$$A^{n+1} = \gamma_0 I + \gamma_1 A + \dots + \gamma_{n-1} A^{n-1}$$

$$= A \cdot A^n$$

$$= A \cdot (\text{degree } n\text{-}1)$$

For example if we have a 2×2 matrix, any power of it can be written down as a polynomial of degree 2

More on SVD

Monday, October 10, 2005
11:40 AM

The SVD is called "robust" by engineers, but JCF is called "dynamic"

Say $A = U\Sigma V^T$

1) Given A, find $\tilde{A}$ so that $\tilde{A}$ is rank r and $||A-\tilde{A}||_2$ is minimal; in other words, what is the best rank r approximation?

Answer: if $A = U\Sigma V^T$, then

$$\tilde{A} = U\Omega V^T, \quad \Omega = \begin{pmatrix} \sigma_1 & & & 0 \\ & \ddots & & \\ & & \sigma_r & \\ 0 & & & 0 \end{pmatrix}$$

i.e. compute SVD, take the first r values, and throw the rest away

$A = U\Sigma V^T \quad \tilde{A} = U\Omega V^T$

$\|A - \tilde{A}\| = \|U\Sigma V^T - U\Omega V^T\|_F$

$= \|U(\Sigma - \Omega)V^T\|_F$

from last time

$= \|(\Sigma - \Omega)\|_F$

$\Omega = \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots \ \sigma_r & \\ 0 & & 0 \end{pmatrix}$

$= (\sigma_{r+1}^2 + \dots + \sigma_n^2)^{1/2}$

We talked before about

$A = \begin{pmatrix} 1 & 0 \\ 0 & 10^{-6} \end{pmatrix}$, the best rank 1 approx is $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$

Aside: Say we have $A = \tilde{A}+E$ with $||E||_F = \varepsilon$

Find $\tilde{A}$ so that $\|A - \tilde{A}\| \leq \varepsilon$ and rank $\tilde{A}$ is minimum

Say A has singular values $\sigma_1, \dots, \sigma_n$

and find lowest r so that $\sigma_1 + \dots + \sigma_r \leq \varepsilon$

Say we have a matrix *A*, and we want it
to be orthogonal, but it isn't. The singular
values of an orthogonal matrix are ± 1.
How to get the best orthogonal matrix approximation to
A?

$$A = U \Sigma V^T \qquad \tilde{Q} = U I V^T$$

$$\|A - \tilde{Q}\|_F = \| U \Sigma V^T - U I V^T \|_F$$

$$= \| U (\Sigma - I) \|_F = \left[(\sigma_1 - 1) + (\sigma_2 - 1)^2 + (\sigma_n - 1)^2 \right]^{1/2}$$

Say we have two matrices containing data

$$A, B \quad m \times n$$

They contain the same data, maybe perturbed by a rotation

$$\min_{Q^T Q = I} \| A - B Q \|_I = \min$$

What's the orthogonal *Q* best for this?

$$B^T A = U \Sigma V^T$$

$$Q = U I V^T$$

This is only known for the Frobenius norm; for other norms
nobody knows what *Q* is

Polar decomposition

Monday, October 10, 2005
11:56 AM

$x = a + ib$

$= \rho e^{i\theta}$ $\qquad \rho > 0$

A is an $n \times n$ matrix with an SVD $U\Sigma V^T$

rewrite this as $\underline{UV^T}\ \underline{V\Sigma V^T}$

$= QH$, H is hermitian, Q orthogonal

Claim that this is the polar decomposition,
an orthogonal matrix times a hermitian

we already saw that $\|b - Ax\|_2 = \min$ is
solved by $\vec{x} = A^+ b$

But maybe we "don't know" A

$\hat{A} = A + E, \quad \|E\|_2 \leq \varepsilon$

Say $\hat{A}$ is our approximation, then

$\hat{x} = \hat{A}^+ b$

Perturbation Theory

Monday, October 10, 2005
12:10 PM

If for some natural norm, $||A|| < 1$, we have

1) $I-A$ is non-singular

2) $\| (I-A)^{-1} \| \leq \frac{1}{1-\|A\|}$

$\geq \frac{1}{1+\|A\|}$

1) Suppose A is Singular, then there is a vector $\boldsymbol{z}$ so that $A\boldsymbol{z} = 0$

If $(I-A)$ is singular, there exists nonzero z with $(I-A)\boldsymbol{z} = 0$

so $z = Az \Rightarrow \|A\| \geq 1$, a contradiction

Proof of (2)

Know inverse exists, so write $I = (I-A)(I-A)^{-1}$

$\|I\| = \|(I-A)(I-A)^{-1}\| \leq \|(I-A)\| \cdot \|(I-A)^{-1}\|$

$\leq (1 + \|A\|)($

So $\frac{1}{1+\|A\|} \leq \|(I-A)^{-1}\|$

we had $(I-A)^{-1}(I-A) = I$

$(I-A)^{-1} - (I-A)^{-1}A = I$

$(I-A)^{-1} = I + (I-A)^{-1}A$

$$\|(I-A)^{-1}\| \leqslant 1 + \|(I-A)^{-1}\| \cdot \|A\|$$

$$(-\|A\| \cdot \|(I-A)^{-1}\| + \|(I-A)^{-1}\| \leqslant 1$$

$$\|(I-A)^{-1}\| \leqslant \frac{1}{1-\|A\|}$$