Numerical Analysis (MATH 395/372)
Singular Value Decomposition

Singular Value Decomposition

- The singular value decomposition (SVD) of a matrix $\mathbf{A} \in M_{m,n}(\mathbb{R})$ is given by

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$$

- Where $\mathbf{\Sigma} \in M_{m,n}(\mathbf{R})$ is a diagonal matrix with elements

$$\sigma_{ij} = \begin{cases} \sigma_i & \text{for } i = j \\ 0, & \text{for } i \neq j \end{cases}$$

  The σ_i's are the *singular values*, and are ordered so that $\sigma_i \geq \sigma_{i+1}$
- And $\mathbf{U} \in M_m(\mathbb{R})$ and $\mathbf{V} \in M_n(R)$ are unitary matrices. The columns of $\mathbf{U}$ and $\mathbf{V}$ are the left and right *singular vectors*

Reduced SVD

- The reduced SVD of a matrix $\mathbf{A} \in M_{m,n}(\mathbb{R})$, $m \geq n$ is given by

$$\mathbf{A} = \hat{\mathbf{U}}\hat{\mathbf{\Sigma}}\mathbf{V}^T$$

- Where $\hat{\mathbf{\Sigma}} \in M_n(\mathbf{R})$ is a square diagonal matrix (the upper $n \times n$ block of $\mathbf{\Sigma}$
- And $\hat{\mathbf{U}} \in M_{m,n}(\mathbb{R})$ has orthonormal columns (the left $m \times n$ block of $\mathbf{U}$)
- $\mathbf{V} \in M_n(\mathbb{R})$ is the same as in the full SVD.

Properties

Theorem 1:

If r is the number of non-zero singular values of $\mathbf{A}$, then $\text{rank}(\mathbf{A}) = r$.

Proof: Since $\mathbf{U}$ and $\mathbf{V}$ are unitary, they both have full rank. Thus, $\text{rank}(\mathbf{U}^T\mathbf{A}\mathbf{V}) = \text{rank}(\mathbf{A})$. But, $\mathbf{\Sigma} = \mathbf{U}^T\mathbf{A}\mathbf{V}$, so $\text{rank}(\mathbf{\Sigma}) = \text{rank}(\mathbf{A})$. The rank of a diagonal matrix is the number of nonzero diagonal elements.

Properties

Theorem 2:

If $\mathbf{A} \in M_{m,n}(\mathbb{R})$, $m \geq n$, has full rank, then $\text{range}(\mathbf{A}) = \langle \mathbf{u}_1, \ldots, \mathbf{u}_n \rangle = \text{range}(\hat{\mathbf{U}})$. Here $\mathbf{u}_k$ denotes the kth column of $\mathbf{U}$.

Proof: If $\mathbf{A}$ has full rank, then $\hat{\mathbf{\Sigma}}$ is invertible, and $\hat{\mathbf{U}} = \mathbf{A}\mathbf{V}\hat{\mathbf{\Sigma}}^{-1}$. If $\mathbf{y} \in \text{range}(\hat{\mathbf{U}})$, then $\mathbf{y} = \hat{\mathbf{U}}\mathbf{x} = \mathbf{A}\left(\mathbf{V}\hat{\mathbf{\Sigma}}^{-1}\mathbf{x}\right)$ for some $\mathbf{x} \in \mathbb{R}^n$, so $\mathbf{y} \in \text{range}(\mathbf{A})$. If $\mathbf{y} \in \text{range}(\mathbf{A})$, then $\mathbf{y} = \mathbf{A}\mathbf{x} = \hat{\mathbf{U}}\left(\hat{\mathbf{\Sigma}}\mathbf{V}^T\mathbf{x}\right)$ for some $\mathbf{x} \in \mathbb{R}^n$, so $\mathbf{y} \in \text{range}(\hat{\mathbf{U}})$. Thus, $\text{range}(\mathbf{A}) = \text{range}(\hat{\mathbf{U}})$.

Application of SVD to $\mathbf{Ax} \cong \mathbf{b}$

- Suppose $\mathbf{A} \in M_{m,n}(\mathbb{R})$, $m \geq n$, has full rank
- Since range$(\mathbf{A}) = \langle \mathbf{u}_1, \ldots, \mathbf{u}_n \rangle$, and $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is orthonormal,
$$\mathbf{P} = \hat{\mathbf{U}}\hat{\mathbf{U}}^T$$
is an orthogonal projector onto range$(\mathbf{A})$.
- Recall that the least squares solution satisfies
$$\mathbf{Ax} = \mathbf{Pb}$$
- Thus,
$$\hat{\mathbf{U}}\hat{\mathbf{\Sigma}}\mathbf{V}^T\mathbf{x} = \hat{\mathbf{U}}\hat{\mathbf{U}}^T\mathbf{b}$$
- Consequently
$$\mathbf{x} = \mathbf{V}\hat{\mathbf{\Sigma}}^{-1}\hat{\mathbf{U}}^T\mathbf{b}$$
- This also gives us another formula for the pseudoinverse
$$\mathbf{A}^+ = \mathbf{V}\hat{\mathbf{\Sigma}}^{-1}\hat{\mathbf{U}}^T$$

Some other SVD properties/applications

- $\|\mathbf{A}\|_2 = \sigma_1$.
- For $\mathbf{A} \in M_m(\mathbb{R})$, $|\det(\mathbf{A})| = \prod_{i=1}^{m} \sigma_i$
- The nonzero singular values of $\mathbf{A}$ are the square roots of the eigenvalues of $\mathbf{A}^T\mathbf{A}$ or $\mathbf{A}\mathbf{A}^T$. The right singular vectors of $\mathbf{A}$ are the corresponding eigenvectors of $\mathbf{A}^T\mathbf{A}$. The *left* singular vectors are the eigenvectors of $\mathbf{A}\mathbf{A}^T$.
- If $\mathbf{A} = \mathbf{A}^T$, then the singular values of $\mathbf{A}$ are the absolute values of the eigenvalues of $\mathbf{A}$.

Low-Rank Approximations

Theorem 3

A is the sum of r rank-one matrices,

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

Theorem 4

For any ν such that $0 \leq \nu \leq r$, let

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

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

$$\|\mathbf{A} - \mathbf{A}_\nu\|_2 = \inf_{\mathbf{B} \in M_{m,n}(\mathbb{R}),\ \mathrm{rank}(\mathbf{B}) \leq \nu} \|\mathbf{A} - \mathbf{B}\|_2 = \sigma_{\nu+1}.$$

Example: Low rank approximations to three-dimensional data

mtcars dataset for R from 1974 *Motor Trend* magazine. Fuel consumption and 10 other design/performance variables for 32 automobiles from 1973-1974 model years:

Variable	Description
mpg	Miles/(US) gallon
cyl	Number of cylinders
disp	Displacement (cu.in.)
hp	Gross horsepower
drat	Rear axle ratio
wt	Weight (lb/1000)
qsec	1/4 mile time
vs	V/S
am	Transmission (0 = automatic, 1 = manual)
gear	Number of forward gears
carb	Number of carburetors

(Source: *mtcars* documentation.)

Example: Low rank approximations to three-dimensional data

Data:

Weight	2.620	2.875	2.320	3.215	3.440	3.460	3.570	3.190	...	2.780
Displacement	160.0	160.0	108.0	258.0	360.0	225.0	360.0	146.7	...	121.0
MPG	21.0	21.0	22.8	21.4	18.7	18.1	14.3	24.4	...	21.4

Summary Statistics:

	mean	standard deviation
Weight	3.217	0.978
Displacement	230.722	123.939
MPG	20.091	6.027

We will standardize the data by subtracting the means and dividing by the standard deviations. New variables have mean $= 0$, standard deviation $= 1$, and are dimensionless.

Weight	-0.610	-0.350	-0.917	-0.002	0.228	0.248	0.361	-0.028	...	-0.447
Displacement	-0.571	-0.571	-0.990	0.220	1.043	-0.046	1.043	-0.678	...	-0.885
MPG	0.151	0.151	0.450	0.217	-0.231	-0.330	-0.961	0.715	...	0.217

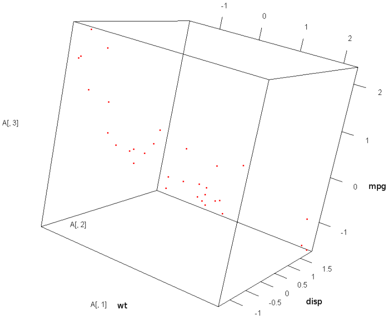
Example: Low rank approximations to three-dimensional data

Let

$$\mathbf{A} = \begin{bmatrix} -0.610 & -0.350 & -0.917 & -0.002 & 0.228 & 0.248 & 0.361 & -0.028 & \dots & -0.447 \\ -0.571 & -0.571 & -0.990 & 0.220 & 1.043 & -0.046 & 1.043 & -0.678 & \dots & -0.885 \\ 0.151 & 0.151 & 0.450 & 0.217 & -0.231 & -0.330 & -0.961 & 0.715 & \dots & 0.217 \end{bmatrix}^*$$

Note that $\mathbf{A} \in M_{32,3}(\mathbb{R})$ has column for each variable and row for each vehicle.

Plot of **A**.



Example: Low rank approximations to three-dimensional data

- Reduced SVD of $\mathbf{A}$:
$$\mathbf{A} = \hat{\mathbf{U}}\hat{\mathbf{\Sigma}}\mathbf{V}^T,$$
where,
  - $\hat{\mathbf{U}} \in M_{32,3}(\mathbb{R})$
  - $\hat{\mathbf{\Sigma}} \in M_3(\mathbb{R})$
  - $\mathbf{V} \in M_3(\mathbb{R})$.
-
$$\hat{\mathbf{\Sigma}} = \begin{bmatrix} 9.21 & 0 & 0 \\ 0 & 2.20 & 0 \\ 0 & 0 & 1.84 \end{bmatrix}$$
Note: $\mathbf{A}$ is full-rank $r = p = 3$.

Example: Low rank approximations to three-dimensional data

- $$\mathbf{V} = \begin{bmatrix} 0.582 & -0.209 & 0.786 \\ 0.577 & -0.575 & -0.580 \\ -0.573 & -0.791 & 0.214 \end{bmatrix}$$

- $$\hat{\mathbf{U}} = \begin{bmatrix} -0.084 & 0.153 & -0.063 \\ -0.067 & 0.128 & 0.048 \\ -0.148 & 0.184 & -0.027 \\ 0.000 & -0.136 & -0.045 \\ 0.094 & -0.211 & -0.259 \\ 0.033 & 0.107 & 0.082 \\ \vdots & \vdots & \vdots \\ -0.097 & 0.196 & 0.114 \end{bmatrix}$$

Example: Low rank approximations to three-dimensional data

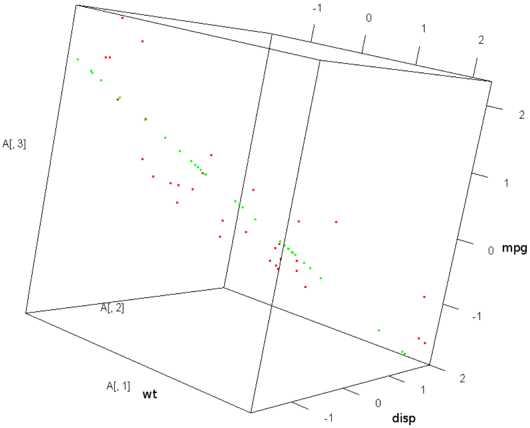
Rank-1 approximation of **A**:

$$\mathbf{A}_1 = \sigma_1 u_1 v_1^T = 9.21 \begin{bmatrix} -0.084 \\ -0.067 \\ -0.148 \\ 0.000 \\ 0.094 \\ 0.033 \\ \vdots \\ -0.097 \end{bmatrix} \begin{bmatrix} 0.582 & 0.577 & -0.573 \end{bmatrix}$$

$$= \begin{bmatrix} -0.449 & -0.445 & 0.442 \\ -0.360 & -0.358 & 0.355 \\ -0.793 & -0.787 & 0.781 \\ 0.001 & 0.001 & -0.001 \\ 0.504 & 0.501 & -0.497 \\ 0.179 & 0.177 & -0.176 \\ \vdots & \vdots & \vdots \\ -1.545 & -1.533 & 1.521 \end{bmatrix}.$$

Example: Low rank approximations to three-dimensional data

Plot of $\mathbf{A}$ and $\mathbf{A}_1$.



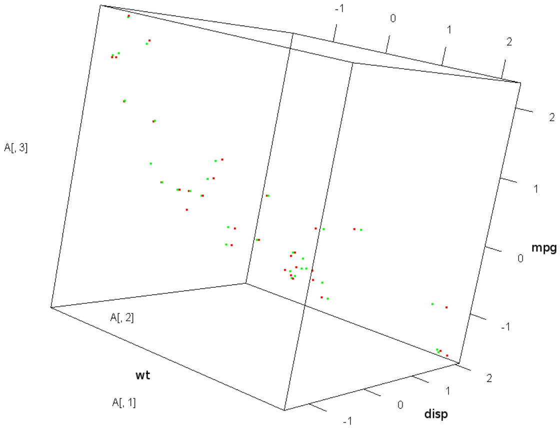
Example: Low rank approximations to three-dimensional data

- What is the equation of this line?
- Note that rows of the rank one approximation are just scalar multiples of $\mathbf{v}_1^T$.
- That is, all of the (green) points lie along the vector, $\left[\begin{array}{ccc} 0.582 & 0.577 & -0.573 \end{array}\right]$.
- So, for each point, there is a scalar t, so that $\left[\begin{array}{ccc} \mathrm{wt} & \mathrm{disp} & \mathrm{mpg} \end{array}\right] = t\left[\begin{array}{ccc} 0.582 & 0.577 & -0.573 \end{array}\right]$.
- This is just the parametric equation (with parameter t) for a line in $\mathbb{R}^3$ that passes through the origin.

Example: Low rank approximations to three-dimensional data

- Rank-2 approximation of $\mathbf{A}$:

$$\mathbf{A}_2 = \mathbf{A}_1 + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T.$$



Example: Low rank approximations to three-dimensional data

- What is the equation of this plane?
- Note that rows of the rank two approximation are just linear combinations of $\mathbf{v}_1^T$ and $\mathbf{v}_2^T$.
- Thus, all of the points are orthogonal to $\mathbf{v}_3^T$.
- That is, they satisfy

$$\begin{bmatrix} 0.786 & -0.580 & 0.214 \end{bmatrix} \begin{bmatrix} \mathrm{wt} \\ \mathrm{disp} \\ \mathrm{mpg} \end{bmatrix} = 0$$

- Thus, the equation for the plane can be written as

$$0.786 \cdot \mathrm{wt} - 0.580 \cdot \mathrm{disp} + 0.214 \cdot \mathrm{mpg} = 0$$

Computing the SVD

- More discussion on this later
- It is more costly than QR factorization for least squares problems
- Algorithms similar to eigenvalue computations

Before next time...

- Prove the statement $\text{rank}(\mathbf{U}^T\mathbf{A}\mathbf{V}) = \text{rank}(\mathbf{A})$ from the proof of theorem 1.
- In the `mtcars` example, we standardized the data before taking the SVD. Hence, the equation for the line and the plane that we derived described the standardized data. Starting with the same line and the same plane, find corresponding equations (models) for the non-standardized data.
- Read Heath Sections 4.1-4.2