CSCI B609:
"Foundations of Data Science"

Lecture 6: Best-Fit Subspaces and SVD

Slides at http://grigory.us/data-science-class.html

Grigory Yaroslavtsev

http://grigory.us

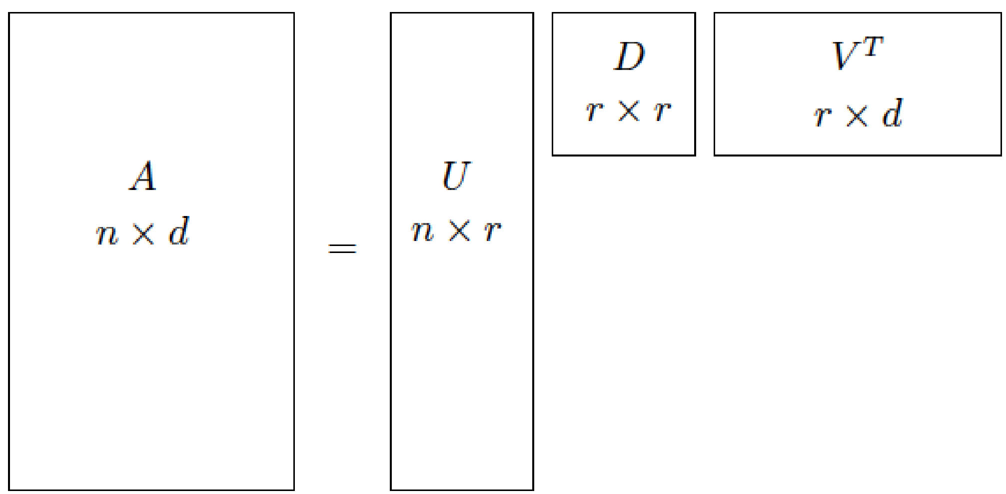
Singular Value Decomposition: Intro

- $\boldsymbol{n} \times \boldsymbol{d}$ data matrix A ($\boldsymbol{n}$ rows and $\boldsymbol{d}$ columns)
- Each row is a $\boldsymbol{d}$-dimensional vector
- Find best-fit k-dim. subspace S_k for rows of A?
- Minimize sum of squared distances from A_i to S_k

SVD: Greedy Strategy

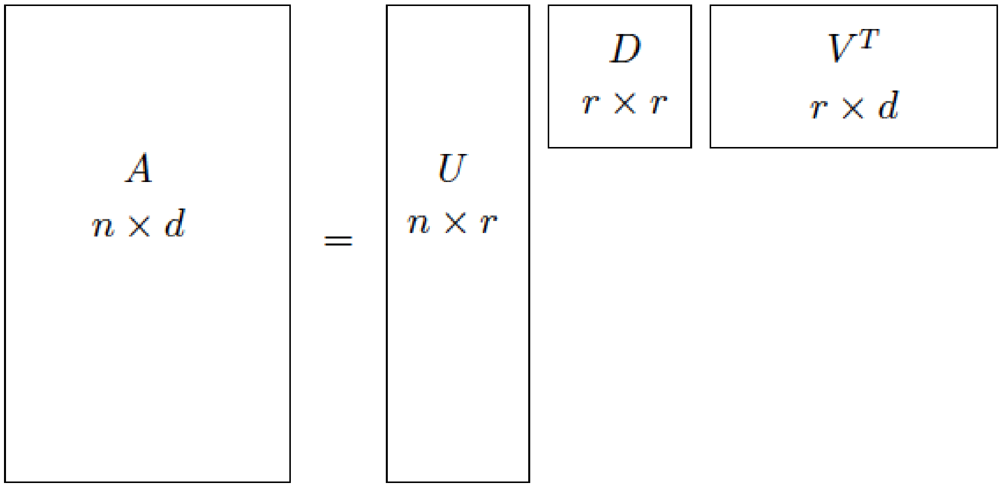
- Find best fit 1-dimensional line
- Repeat k times
- When $k = r = rank(A)$ we get the SVD:

$$A = UDV^T$$



$A = UDV^T$: Basic Properties

- D = Diagonal matrix (positive real entries d_{ii})
- U, V: orthonormal columns:
  - $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r \in \mathbb{R}^d$ (best fitting lines)
  - $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r \in \mathbb{R}^n$ (~projections of rows of A on $\boldsymbol{v}_i'$s)
  - $\langle \boldsymbol{u}_i, \boldsymbol{u}_j \rangle = \delta_{ij}, \langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle = \delta_{ij}$
- $A = \sum_i d_{ii} \boldsymbol{u}_i \boldsymbol{v}_i^T$



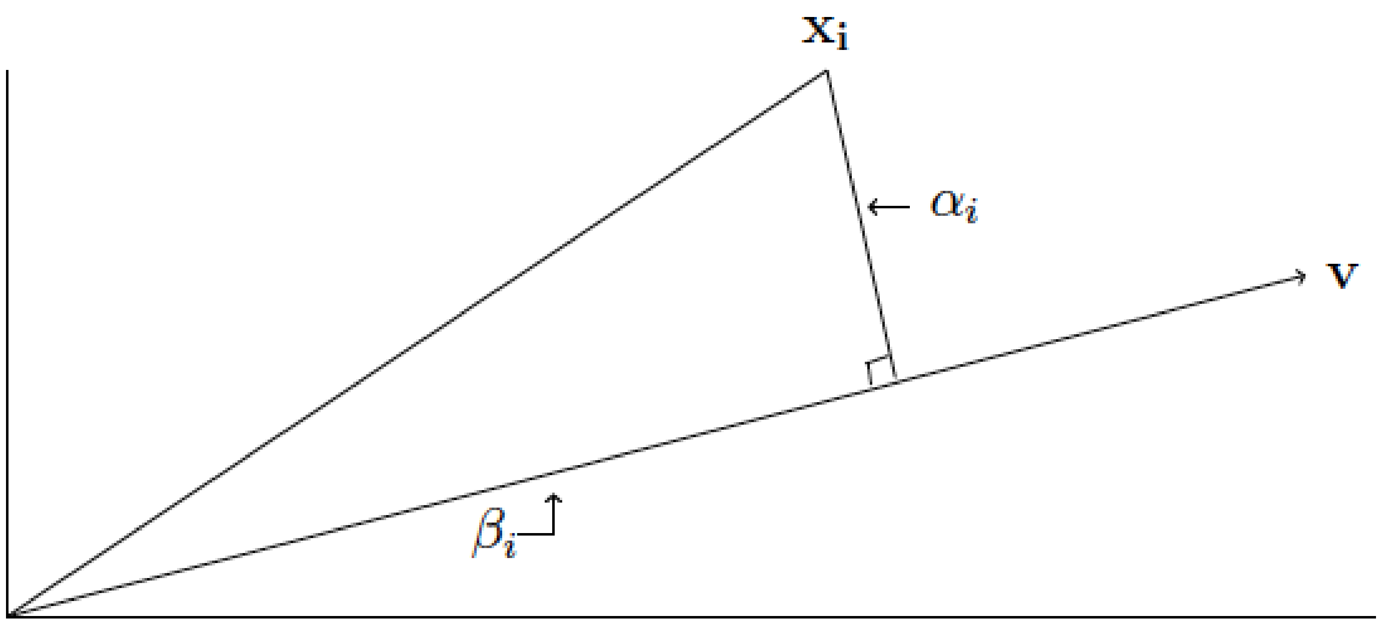
Singular Values vs. Eigenvalues

- If A is a square matrix:
  - Vector $\boldsymbol{v}$ such that $A\boldsymbol{v} = \lambda\boldsymbol{v}$ is an eigenvector
  - λ = eigenvalue
  - For symmetric real matrices $\boldsymbol{v}$'s are orthonormal
    $$A = VDV^T$$
  - $V's$ columns are eigenvectors of A
  - Diagonal entries of D are eigenvalues $\lambda_1, \dots, \lambda_n$
- SVD is defined for all matrices (not just square)
  - Orthogonality of singular vectors is automatic
    $$A\boldsymbol{v}_i = d_{ii}\boldsymbol{u}_i \text{ and } A^T\boldsymbol{u}_i = d_{ii}\boldsymbol{v}_i \text{ (will show)}$$
    $$A^TA\boldsymbol{v}_i = d_{ii}^2\boldsymbol{v}_i \Rightarrow \boldsymbol{v}_i's \text{ are eigenvectors of } A^TA$$

Projections and Distances

- Minimizing distance = maximizing projection

$$||\boldsymbol{x}||_2^2 = (projection)^2 + (distance\ to\ line)^2$$



SVD: First Singular Vector

- Find best fit 1-dimensional line
- $\boldsymbol{v}$ = unit vector along the best fit line
- $\boldsymbol{a}_i$= i-th row of A, length of its projection: $|\langle \boldsymbol{a}_i, \boldsymbol{v} \rangle|$
- Sum of squared projection lengths: $||A\boldsymbol{v}||_2^2$
- **First singular vector**:

$$\boldsymbol{v}_1 = \arg \max_{||\boldsymbol{v}||_2 = 1} ||A\boldsymbol{v}||_2$$

- If there are ties, break arbitrarily
- $\sigma_1(A) = ||A\boldsymbol{v}_1||_2$ is the **first singular value**

SVD: Greedy Construction

- Find best fit 1-dimensional line, repeat r times (until projection is 0)
- **Second singular vector and value:**

$$\boldsymbol{v}_2 = \arg \max_{\boldsymbol{v} \perp \boldsymbol{v}_1, ||\boldsymbol{v}||_2 = 1} ||A\boldsymbol{v}||_2$$

$$\sigma_2(A) = ||A\boldsymbol{v}_2||_2$$

- **k-th singular vector and value:**

$$\boldsymbol{v}_k = \arg \max_{\boldsymbol{v} \perp \boldsymbol{v}_1, \ldots \boldsymbol{v}_{k-1}, ||\boldsymbol{v}||_2 = 1} ||A\boldsymbol{v}||_2$$

$$\sigma_k(A) = ||A\boldsymbol{v}_k||_2$$

- Will show:$(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ is best-fit subspace

Best-Fit Subspace Proof: $k = 2$

- W = best-fit 2-dimensional subspace
- Orthonormal basis $(\boldsymbol{w}_1, \boldsymbol{w}_2)$: $||A\boldsymbol{w}_1||_2^2 + ||A\boldsymbol{w}_2||_2^2$
- Key observation: choose $\boldsymbol{w}_2 \perp \boldsymbol{v}_1$
  - If $W \perp \boldsymbol{v}_1$ then any vector in W works
  - Otherwise $\boldsymbol{v}_1 = \boldsymbol{v}_1^{||} + \boldsymbol{v}_1^{\perp}$ for $\boldsymbol{v}_1^{||}$ = projection on W
  - Choose $\boldsymbol{w}_2 \perp \boldsymbol{v}_1^{||}$:

  $\langle \boldsymbol{w}_2, \boldsymbol{v}_1 \rangle = \langle \boldsymbol{w}_2, \boldsymbol{v}_1^{||} + \boldsymbol{v}_1^{\perp} \rangle = \langle \boldsymbol{w}_2, \boldsymbol{v}_1^{||} \rangle + \langle \boldsymbol{w}_2, \boldsymbol{v}_1^{\perp} \rangle = 0$
- $||A\boldsymbol{w}_1||_2^2 \le ||A\boldsymbol{v}_1||_2^2$ and $||A\boldsymbol{w}_2||_2^2 \le ||A\boldsymbol{v}_2||_2^2$

$$||A\boldsymbol{w}_1||_2^2 + ||A\boldsymbol{w}_2||_2^2 \le ||A\boldsymbol{v}_1||_2^2 + ||A\boldsymbol{v}_2||_2^2$$

Best-Fit Subspace Proof: General k

- W = best-fit k-dimensional subspace
- $V_{k-1} = span(\boldsymbol{v}_1, \dots, \boldsymbol{v}_{k-1})$ best fit (k-1)-dimensional subspace
- Orthonormal basis $\boldsymbol{w}_1, \dots, \boldsymbol{w}_k$, where $\boldsymbol{w}_k \perp V_{k-1}$

$$\sum_{i=1}^{k-1} ||A\boldsymbol{w}_i||_2^2 \leq \sum_{i=1}^{k-1} ||A\boldsymbol{v}_i||_2^2$$

- $\boldsymbol{w}_k \perp V_{k-1} \Rightarrow$ by def. of $\boldsymbol{v}_k$ $||A\boldsymbol{w}_k||_2^2 \leq ||A\boldsymbol{v}_k||_2^2$

$$\sum_{i=1}^{k} ||A\boldsymbol{w}_i||_2^2 \leq \sum_{i=1}^{k} ||A\boldsymbol{v}_i||_2^2$$

Singular Values and Frobenius Norm

- $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r$ span the space of all rows of A
- $\langle \boldsymbol{a}_j, \boldsymbol{v} \rangle = 0$ for all $\boldsymbol{v} \perp \boldsymbol{v}_1, \ldots, \boldsymbol{v}_r \Rightarrow$

$$\left|\left|\boldsymbol{a}_j\right|\right|_2^2 = \sum_{i=1}^{r} \langle \boldsymbol{a}_j, \boldsymbol{v}_i \rangle^2$$

$$\sum_{j=1}^{n} \sum_{k=1}^{d} a_{jk}^2 = \sum_{j=1}^{n} \left|\left|\boldsymbol{a}_j\right|\right|_2^2 = \sum_{j=1}^{n} \sum_{i=1}^{r} \langle \boldsymbol{a}_j, \boldsymbol{v}_i \rangle^2 =$$

$$\sum_{i=1}^{r} \sum_{j=1}^{n} \langle \boldsymbol{a}_j, \boldsymbol{v}_i \rangle^2 = \sum_{i=1}^{r} \left|\left|A\boldsymbol{v}_i\right|\right|_2^2 = \sum_{i=1}^{r} \sigma_i^2(A)$$

- $\sqrt{\sum_{j=1}^{n} \sum_{k=1}^{d} a_{jk}^2} = \left|\left|\mathrm{A}\right|\right|_{\mathrm{F}}$ (Frobenius norm) $= \sqrt{\sum_{i=1}^{r} \sigma_i^2(A)}$