Machine Learning, STOR 565

The Singular Value Decomposition

Andrew Nobel

January, 2021

Preliminaries

Fact: If $\mathbf{A} \in \mathbb{R}^{m \times n}$ is any matrix then

- $\mathbf{A}\mathbf{A}^t \in \mathbb{R}^{m \times m}$ is symmetric, non-negative definite
- $\mathbf{A}^t\mathbf{A} \in \mathbb{R}^{n \times n}$ is symmetric, non-negative definite
- $r :=$ rank$(\mathbf{A}) =$ rank$(\mathbf{A}^t\,\mathbf{A}) =$ rank$(\mathbf{A}\mathbf{A}^t) \leq \min(m, n)$
- $\mathbf{A}\mathbf{A}^t$ and $\mathbf{A}^t\mathbf{A}$ have same non-zero eigenvalues $\lambda_1, \ldots, \lambda_r$

The Singular Value Decomposition (SVD)

Theorem (SVD): *Any* matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ can be written in the form

$$\mathbf{A} \;=\; \mathbf{U}\mathbf{D}\mathbf{V}^t$$

- $\mathbf{U} = [\mathbf{u}_1, \cdots, \mathbf{u}_m] \in \mathbb{R}^{m \times m}$ is orthogonal. Its columns $\mathbf{u}_i$ are the (orthonormal) eigenvectors of $\mathbf{A}\mathbf{A}^t$, called *left singular vectors*

- $\mathbf{V} = [\mathbf{v}_1, \cdots, \mathbf{v}_n] \in \mathbb{R}^{n \times n}$ is orthogonal. Its columns $\mathbf{v}_i$ are the (orthonormal) eigenvectors of $\mathbf{A}^t\mathbf{A}$, called *right singular vectors*

- $\mathbf{D} = \text{diag}(\sigma_1(\mathbf{A}), \ldots, \sigma_r(\mathbf{A}))$ is an $m \times n$ diagonal matrix. Here

$$\sigma_1(\mathbf{A}) \geq \ldots \geq \sigma_r(\mathbf{A}) \geq 0 \;\; \text{with} \;\; \sigma_i(\mathbf{A}) = \sqrt{\lambda_i(\mathbf{A}\mathbf{A}^t)}$$

are called the *singular values* of $\mathbf{A}$, and $r = \text{rank}(\mathbf{A})$

The SVD, cont.

By expanding the expression $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^t$ we find that SVD can be written in the equivalent form

$$\mathbf{A} = \sum_{i=1}^{r} \sigma_i(\mathbf{A})\, \mathbf{u}_i\, \mathbf{v}_i^t$$

- $\mathbf{A}$ is a weighted sum of r rank-1 matrices $\mathbf{u}_i \mathbf{v}_i^t$ with $||\mathbf{u}_i \mathbf{v}_i^t|| = 1$
- $||\mathbf{A}||_F^2 = \sum_{i=1}^{r} ||\sigma_i(\mathbf{A})\, \mathbf{u}_i\, \mathbf{v}_i^t||^2 = \sum_{i=1}^{r} \sigma_i(\mathbf{A})^2$
- $\sigma_i(\mathbf{A}) \geq \sigma_{i+1}(\mathbf{A}) \geq 0$ so terms in sum are ordered by weight

SVD and Low-Rank Matrix Approximation

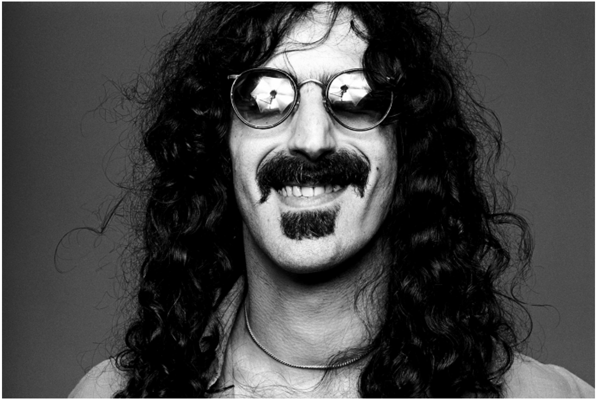
Idea: First d terms in the SVD give a rank d approximation of $\mathbf{A}$

$$\hat{\mathbf{A}}_d \; := \; \sum_{i=1}^{d} \sigma_i(\mathbf{A}) \, \mathbf{u}_i \, \mathbf{v}_i^t, \quad d = 1, \ldots, r$$

- $||\hat{\mathbf{A}}_d||^2 = \sum_{i=1}^{d} \sigma_i(\mathbf{A})^2$
- $||\mathbf{A} - \hat{\mathbf{A}}_d||^2 = \sum_{i=d+1}^{r} \sigma_i(\mathbf{A})^2$
- $\hat{\mathbf{A}}_d$ minimizes $||\mathbf{A} - \mathbf{B}||$ over all rank d matrices $\mathbf{B}$
- Proportion of variation explained by rank d approximation $\mathbf{A}_d$ is

$$\text{PVE}(\hat{\mathbf{A}}_d) \; = \; \frac{||\hat{\mathbf{A}}_d||^2}{||\mathbf{A}||^2} \; = \; \frac{\sum_{i=1}^{d} \sigma_i(\mathbf{A})^2}{\sum_{i=1}^{r} \sigma_i(\mathbf{A})^2}$$

Example of SVD on Image Data



- **Matrix:** $\mathbf{A} = 458 \times 685$ matrix of pixel intensities
- **Question:** Will low rank approximation $\hat{\mathbf{A}}_d$ *look* good?

Proportion of Variation Explained

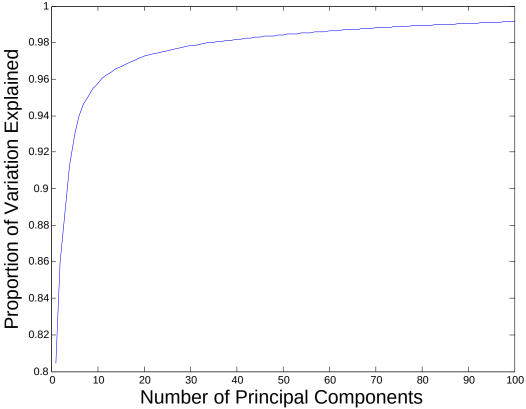
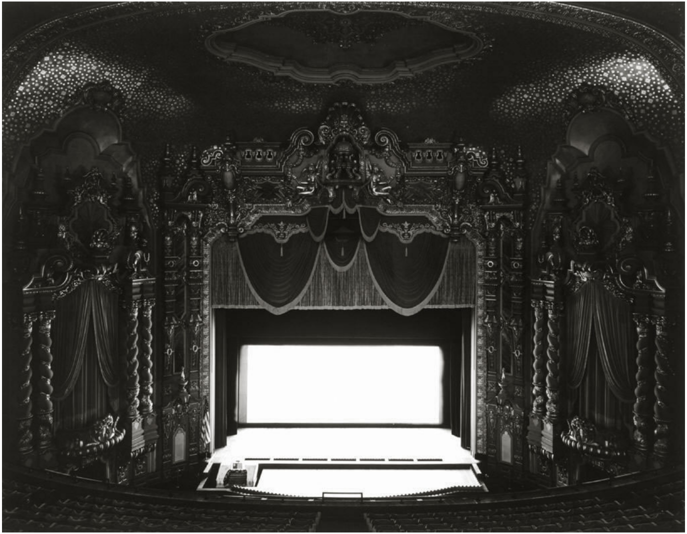


Image Reconstruction

Image Reconstruction: Ohio Theater Photo (H. Sugimoto)



SVD of Ohio Theater Photo (H. Sugimoto)